The edit distance function: Forbidding induced powers of cycles and other questions.

by

Zhanar Berikkyzy

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Mathematics

Program of Study Committee:

Ryan R. Martin, Major Professor

Steve Butler

Leslie Hogben

Justin Peters

Sung-Yell Song

Iowa State University

Ames, Iowa

2016

DEDICATION

To my parents, Berik and Beibitgul.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

This dissertation would not have been written without the fervent support and careful guidance of numerous people.

I would like to express my sincere gratitude to my advisor Dr. Ryan R. Martin for his continuous support and guidance during my PhD studies and research. He provided much needed encouragement and patience in times of difficulties. His enthusiasm, vast knowledge and expert advice have been invaluable throughout all stages of my work. I consider myself very fortunate to have Dr. Martin as my advisor.

I am grateful to my committee members Dr. Steve Butler, Dr. Leslie Hogben, Dr. Justin Peters, and Dr. Sung-Yell Song for their helpful suggestions, insightful comments and professional assistance in making my defense enjoyable. I would also like to thank the Mathematics Department at Iowa State, including all faculty and staff, for offering me ample opportunities which enriched my experience, and providing a stimulating and supportive environment. I am very thankful for the travel support to numerous conferences, and for the Wolfe Research Fellowship which allowed me to focus solely on my research and improve its quality.

I owe my gratitude to Melanie for the tremendous help she provided on countless occasions, to Dr. Jennifer Newman for her understanding and support with teaching assignments when I needed it the most, and to Dr. Song for the friendly and encouraging conversations. I express my sincere gratitude to Dr. Butler, Dr. Hogben, Dr. Stolee, and Dr. Young for mentoring me on various projects which broadened my research area and collaborations.

I would also like to thank my undergraduate professors in my home country, Dr. Askar Dzhumadildaev and Dr. Stanislav Kharin, for giving me the opportunity to start a research in mathematics early. They were first collaborators and role models for me.

I would like to thank all my friends, especially Zhuldyz P. and Cholpon O., for their encouragement and emotional support. I am grateful to my dear friend Aliya K., a caring, generous and righteous person, whom I had always looked up to, and who helped me become a better person. She had passed away briefly before I started my PhD program. With her by my side, this long journey wouldn't have been so lonely at times; she would have been very proud of me.

I take this opportunity to express my deep gratitude to my parents for their unconditional love, prayers, and support in all aspects of my life; to my brothers and sweet little sister for their love and seeing me as a role model (although I am quite far from being one). This kind of support has given me a much needed resolve to go further.

Finally, I wish to thank my husband for his love, encouragement, patience and understanding. He always sees more potential in me than I do myself. I am very thankful to our wonderful children for being my strongest motivation and cheering me up all these years. To have such a family throughout all the critical periods of one's life is truly a blessing.

ABSTRACT

The edit distance between two graphs on the same labeled vertex set is defined to be the size of the symmetric difference of the edge sets. The edit distance between a graph, G, and a graph property, $\mathcal{H}$, is the minimum edit distance between G and a graph in $\mathcal{H}$. The edit distance function of a graph property $\mathcal{H}$ is a function of $p \in [0,1]$ that measures, in the limit, the maximum normalized edit distance between a graph of density p and $\mathcal{H}$.

In this thesis, we address the edit distance function for the property of having no induced copy of C_h^t, the t^{th} power of the cycle of length h. For $h \geq 2t(t+1)+1$ and h not divisible by $t+1$, we determine the function for all values of p. For $h \geq 2t(t+1)+1$ and h divisible by $t+1$, the function is obtained for all but small values of p. We also obtain some results for smaller values of h, present alternative proofs of some important previous results using simple optimization techniques and discuss possible extension of the theory to hypergraphs.

CHAPTER 1. INTRODUCTION

The edit distance in graphs was introduced independently by Alon and Stav [1] and by Axenovich, Kézdy, and Martin [3]. The question considered is "Given a class of graphs $\mathcal{H}$ what is the minimum number $m = m(n)$ such that for every graph on n vertices, there is a set of m edge-additions and edge-deletions that ensure the resultant graph is a member of $\mathcal{H}$?"

The edit distance between two graphs on the same labeled vertex set is defined to be the size of the symmetric difference of the edge sets. The edit distance function of a graph property $\mathcal{H}$ is a function of $p \in [0,1]$ that measures, in the limit, the maximum normalized edit distance between a graph of density p and $\mathcal{H}$. This thesis studies the edit distance function for graph properties of the form $\mathrm{Forb}(H)$, where $H = C_h^t$, the t^{th} power of the cycle of length h.

In this chapter we provide most of the terminology necessary to introduce the problem, discuss the related work and previous results, and state the main theorems of the thesis. Proofs of the theorems are given in Chapter 2.

1.1 Definitions and Notation

The standard graph theory definitions have been adapted from the book by West [27] and the edit distance definitions primarily come from work by Martin [13, 15]. We only define the basic definitions and background necessary to introduce the problem in this chapter. Other terminology will be defined as needed within the text of this thesis. The reader familiar with basic graph theory terminology may wish to begin with Section 1.1.2.

1.1.1 Basic graph theory terminology

Definitions in this section mostly come from the book by West [27]. A *graph* G is a triple consisting of a vertex set $V(G)$, an edge set $E(G)$ (possibly empty), and a relation that associates with each edge two vertices (not necessarily distinct) called its *endpoints*. An edge whose endpoints are equal is called a *loop*. If edges have the same endpoints they are called *multiple edges*. A graph is called *simple* if it has no loops or multiple edges.

When two vertices are the endpoints of an edge we say that they are *adjacent*; otherwise, they are said to be *nonadjacent*. A *neighborhood* of a vertex v, denoted $N(v)$, is the set of vertices adjacent to v. A vertex v is *incident* to edge e if it is an endpoint of e. The *degree* of a vertex v is the number of edges incident to v. A *path* is a simple graph whose vertices can be ordered so that two vertices are adjacent if and only if they are consecutive in the list. A *length* of a path is its number of edges. A graph is *connected* if each pair of vertices belong to a path; otherwise, the graph is *disconnected*. The *distance* between two vertices in a graph is the minimum length of paths connecting them; if no such path exists then the distance is set equal to ∞. The *density* of a graph is the number of edges divided by the maximum possible number of edges in the graph.

A graph H is *subgraph* of G, written $H \subseteq G$, if $V(H) \subseteq V(G)$, $E(H) \subseteq E(G)$, and the assignment of endpoints to edges in H is the same as in G. An *induced subgraph* is a subgraph obtained by deleting a set of vertices and incident edges. An *isomorphism* from graph G to graph G' is a bijection $f : V(G) \to V(G')$ such that $uv \in E(G)$ if and only if $f(u)f(v) \in E(G')$. If there is an isomorphism from G to G' we say that they are *isomorphic*, denoted $G \cong G'$. Given a graph G, the *complement* $\overline{G}$ of G is a graph with vertex set $V(G)$ defined by $e \in E(\overline{G})$ if and only if $e \notin E(G)$.

A *clique* in a graph is a set of pairwise adjacent vertices and the *clique number* of a graph G, denoted $\omega(G)$, is the maximum size of a clique in G. An *independent set* in a graph is a set of pairwise nonadjacent vertices and the *independence number* of a graph G, denoted $\alpha(G)$, is the maximum size of an independent set in G. The *chromatic number* of a graph G, denoted $\chi(G)$, is the minimum number of colors needed to label the vertices so that the adjacent vertices have different colors.

There are certain families of graphs that are often used in this thesis. The Erdős-Rényi *random graph* $G(n, p)$ is a graph on n vertices in which every pair of vertices is joined by an edge with probability p, independently [11]. A path on n vertices is denoted P_n. A *complete* *graph* on n vertices, denoted K_n, is a graph whose vertices are pairwise adjacent. A graph is *bipartite* if its vertex set is the union of two disjoint independent sets called *partite sets*. A *complete bipartite graph* $K_{s,t}$ is a bipartite graph such that partite sets have sizes s and t, and two vertices are adjacent if and only if they are in different partite sets. A *cycle* on n vertices, denoted C_n, is a graph with an equal number of vertices and edges whose vertices can be places around a circle so that two vertices are adjacent if and only if they appear consecutively along the circle.

The t^{th} *power* of a graph G, denoted G^t, is a graph with vertex set $V(G)$ and edge set $\{uv : \text{ the distance between } u \text{ and } v \text{ is at most } t\}$. In particular, in this thesis we are interested in powers of cycles C_h^t.

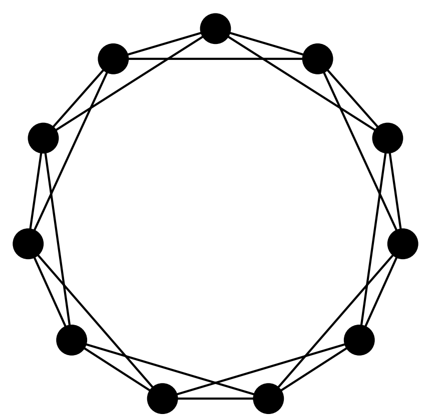

Figure 1.1 The graph C_{11}^2.

1.1.2 Problem specific terminology

All graphs considered in this thesis are simple. A *hereditary property* is a family of graphs that is closed under isomorphism and deletion of vertices. Many interesting graph properties are hereditary, such as being planar, perfect, k-colorable, etc. The property of having no induced subgraph H (forbidden induced subgraph) is called a *principal hereditary property*, denoted $\text{Forb}(H)$. Every hereditary property can be defined by a (possibly infinite) family of forbidden induced subgraphs, i.e. there is a family of graphs $\mathcal{F}(\mathcal{H})$ such that $\mathcal{H} = \bigcap_{H \in \mathcal{F}(\mathcal{H})} \text{Forb}(H)$. A hereditary property is *nontrivial* if it contains an infinite sequence of graphs. All hereditary properties considered in this thesis are nontrivial.

The *edit distance* between two graphs G and G' on the same labeled vertex set, denoted $\mathrm{dist}(G, G')$, is the symmetric difference of the edge sets.

Definition 1. *The edit distance between a graph G and a hereditary property $\mathcal{H}$ is defined as*

$$\mathrm{dist}(G, \mathcal{H}) = \min\{\mathrm{dist}(G, G') : V(G) = V(G'), G' \in \mathcal{H}\}.$$

Definition 2. *The edit distance from the set of all n-vertex graphs to the hereditary property $\mathcal{H}$ is defined as*

$$\mathrm{dist}(n, \mathcal{H}) = \max\{\mathrm{dist}(G, \mathcal{H}) : |G| = n\}.$$

Definition 3. *The edit distance function of a hereditary property $\mathcal{H}$ is a function of $p \in [0, 1]$ that measures, in the limit, the maximum normalized edit distance between a graph of density p and $\mathcal{H}$, i.e.*

$$\mathrm{ed}_{\mathcal{H}}(p) = \lim_{n\to\infty} \max\{\mathrm{dist}(G, \mathcal{H}) : |V(G)| = n, |E(G)| = \lfloor p\tbinom{n}{2}\rfloor\}/\tbinom{n}{2}. \tag{1.1}$$

The existence of above limit is shown by Balogh and Martin in [6], where authors further prove that $\mathrm{ed}_{\mathcal{H}}(p)$ is continuous and concave down for all $p \in [0, 1]$.

Definition 4. *A colored regularity graph (CRG), K, is a complete graph with a partition of the vertices into white* $\mathrm{VW}(K)$ *and black* $\mathrm{VB}(K)$*, and a partition of the edges into white* $\mathrm{EW}(K)$*, gray* $\mathrm{EG}(K)$*, and black* $\mathrm{EB}(K)$*.*

A CRG K' is said to be a *sub-CRG* of K if K' can be obtained by deleting vertices of K. We say that a graph H *embeds in* K, denoted $H \mapsto K$, if there is a function $\varphi : V(H) \to V(K)$ so that if $h_1h_2 \in E(H)$, then either $\varphi(h_1) = \varphi(h_2) \in \mathrm{VB}(K)$ or $\varphi(h_1)\varphi(h_2) \in \mathrm{EB}(K) \cup \mathrm{EG}(K)$, and if $h_1h_2 \notin E(H)$, then either $\varphi(h_1) = \varphi(h_2) \in \mathrm{VW}(K)$ or $\varphi(h_1)\varphi(h_2) \in \mathrm{EW}(K) \cup \mathrm{EG}(K)$. We denote $\mathcal{K}(\mathcal{H})$ to be the subset of CRGs such that no forbidden graph embeds into them, i.e. $\mathcal{K}(\mathcal{H}) = \{K : H \not\mapsto K, \forall H \in \mathcal{F}(\mathcal{H})\}$. In our case, $\mathcal{K}(\mathcal{H}) = \{K : H \not\mapsto K\}$ for $\mathcal{H} = \mathrm{Forb}(H)$ with $H = C_h^t$.

For every CRG K we associate functions f_K and g_K on $[0, 1]$ that can be used to compute the edit distance function. In the definition below $\mathbf{0}$ is the all-zeros vector, $\mathbf{1}$ is the all-ones vector, and the vector inequality is used entrywise.

Definition 5. *Let K be a CRG with a vertex set $\{v_1, \ldots, v_k\}$ and let $p \in [0,1]$. Then the functions f_K and g_K are defined as follows:*

$$f_K(p) = [p\,(|\mathrm{VW}(K)| + 2|\mathrm{EW}(K)|) + (1-p)\,(|\mathrm{VB}(K)| + 2|\mathrm{EB}(K)|)]/k^2 \tag{1.2}$$

$$g_K(p) = \min\{\mathbf{x}^T \mathbf{M}_K(p)\mathbf{x} : \mathbf{x}^T\mathbf{1} = 1, \mathbf{x} \geq \mathbf{0}\}, \tag{1.3}$$

where

$$[\mathbf{M}_K(p)]_{ij} = \begin{cases} p, & \text{if } v_iv_j \in \mathrm{EW}(K) \text{ or } v_i = v_j \in \mathrm{VW}(K); \\ 1-p, & \text{if } v_iv_j \in \mathrm{EB}(K) \text{ or } v_i = v_j \in \mathrm{VB}(K); \\ 0, & \text{if } v_iv_j \in \mathrm{EG}(K). \end{cases} \tag{1.4}$$

Balogh and Martin showed in [6] that $\mathrm{ed}_{\mathcal{H}}(p) = \inf_{K \in \mathcal{K}(\mathcal{H})} g_K(p) = \inf_{K \in \mathcal{K}(\mathcal{H})} f_K(p)$ and Marchant and Thomason further showed in [12] that for every $p \in [0,1]$, there is a CRG $K \in \mathcal{K}(\mathcal{H})$ such that $\mathrm{ed}_{\mathcal{H}}(p) = g_K(p)$, and in order to find such CRG we only need to look at p-core CRGs. A CRG K is called *p-core* if $g_K(p) < g_{K'}(p)$ for every sub-CRG K' of K. Certain CRGs are important for constructions. The CRG with r white vertices, s black vertices and all edges gray is denoted $K(r,s)$.

Definition 6. *The clique spectrum of the hereditary property $\mathcal{H} = \mathrm{Forb}(H)$, denoted $\Gamma(\mathcal{H})$, is the set of all pairs (r,s) such that $H \not\mapsto K(r,s)$.*

It is easy to see that, for any hereditary property $\mathcal{H}$ its clique spectrum $\Gamma = \Gamma(\mathcal{H})$ can be expressed as a Ferrers diagram. That is, given $(r,s) \in \Gamma$, if $r \geq 1$ then $(r-1,s) \in \Gamma$ and if $s \geq 1$ then $(r, s-1) \in \Gamma$.

Definition 7. *An extreme point of a clique spectrum Γ is a pair $(r,s) \in \Gamma$ such that $(r+1,s)$ and $(r,s+1)$ do not belong to Γ. The set of all extreme points of Γ is denoted by Γ^*.*

Given a clique spectrum of a hereditary property, we define the function $\gamma_{\mathcal{H}}$ which gives us an upper bound for the function $\mathrm{ed}_{\mathcal{H}}$ for all $p \in [0,1]$.

Definition 8. *Let $p \in [0,1]$ and $\mathcal{H}$ be a hereditary property. Then the function $\gamma_{\mathcal{H}}$ is defined as*

$$\gamma_{\mathcal{H}}(p) = \min\{g_{K(r,s)}(p) : (r,s) \in \Gamma(\mathcal{H})\}.$$

Notice that in order to compute the $\gamma_{\mathcal{H}}$ function we need to consider only the extreme points rather than the whole clique spectrum, that is, $\gamma_{\mathcal{H}}(p) = \min\{g_{K(r,s)}(p) : (r,s) \in \Gamma^*(\mathcal{H})\}$. The following example illustrates how the $\gamma_{\mathcal{H}}$ function can be computed from the clique spectrum of a hereditary property.

Example 9. Let $\mathcal{H} = \mathrm{Forb}(C_{15}^3)$ and $\Gamma = \Gamma(\mathcal{H})$. We first compute the clique spectrum of $\mathcal{H}$. Since the chromatic number of the graph C_{15}^3 is five, $(5,0) \notin \Gamma$ and $(4,0) \in \Gamma$.

Notice that $C_{15}^3 \not\mapsto K(3,2)$. This is because the largest size of a clique is four and after placing vertices of C_{15}^3 into two cliques, the remaining (at least) seven vertices can not be partitioned into three independent sets as they will contain K_4. Therefore, $(r,s) \in \Gamma$ for $0 \leq s \leq 2$ and $0 \leq r \leq 3$. On the other hand, $C_{15}^3 \mapsto K(1,3)$, therefore $(r,3) \notin \Gamma$ for $r \geq 1$.

Last, since the largest clique in C_{15}^3 is K_4, placing vertices into three cliques will leave out at least three vertices, and so $C_{15}^3 \not\mapsto K(0,3)$. However, grouping vertices consecutively, the vertices can be partitioned into four cliques. Therefore, $(0,4) \notin \Gamma$ and $(0,s) \in \Gamma$ for $0 \leq s \leq 3$.

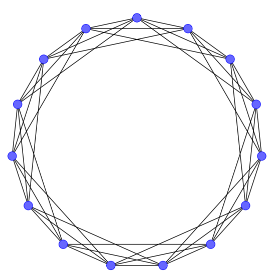

Figure 1.2 The graph C_{15}^3.

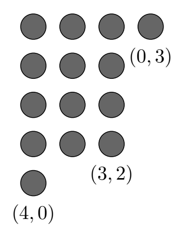

Figure 1.3 Ferrers diagram of $\Gamma(\mathrm{Forb}(C_{15}^3))$.

The graph C_{15}^3 and the clique spectrum of $\mathcal{H}$ expressed as a Ferrers diagram are shown in Figure 9 above. Notice that the extreme points of Γ are $(0,3), (3,2)$ and $(4,0)$, therefore

$$\gamma_{\mathcal{H}}(p) = \min\{g_{K(r,s)}(p) : (r,s) \in \{(0,3),(3,2),(4,0)\}\}.$$

Let K be a CRG with a vertex $v \in V(K)$, and let $\mathbf{x}$ be an optimal solution to the quadratic program (1.3), we often call $\mathbf{x}$ an *optimal weight vector*. The *weight* of v, denoted $\mathbf{x}(v)$, is the entry corresponding to v of the vector $\mathbf{x}$. We say that $w \in V(K)$ is a *gray neighbor* of $v \in V(K)$ if w is adjacent to v via a gray edge. White and black neighbors are defined analogously. The set of all gray neighbors of v is denoted by $N_G(v)$ and the number of vertices adjacent to v via gray edges is denoted by $\deg_G(v)$, i.e. $\deg_G(v) = |N_G(v)|$. Similarly, $\deg_W(v) = |N_W(v)|$, where $N_W(v)$ is the set of all white neighbors of v and $\deg_B(v) = |N_B(v)|$, where $N_B(v)$ is the set of all black neighbors of v.

In contrast, the *weighted gray degree* of v, denoted $d_G(v)$, is the sum of the weights of gray neighbors of v, i.e. $d_G(v) = \sum\{\mathbf{x}(w) : w \in N_G(v)\}$. The *weighted white degree* of v, denoted $d_W(v)$, is the sum of the weights of the white neighbors of v plus the weight of v if it is a white vertex. Similarly, the *weighted black degree* of v, denoted $d_B(v)$, is the sum of the weights of the black neighbors of v plus the weight of v if it is a black vertex. So, $d_G(v) + d_W(v) + d_B(v) = 1$ for all $v \in V(K)$.

The number of common gray neighbors of vertices v and w is denoted by $\deg_G(v, w)$. The *weighted gray codegree* of vertices v and w, denoted $d_G(v, w)$, is the sum of the weights of the common gray neighbors of v and w. For a set of vertices $\{v_1, v_2, \ldots, v_\ell\}$, we say $v_1 v_2 \cdots v_\ell$ is a *gray path* if $v_i v_{i+1} \in \mathrm{EG}(K)$ for $i = 1, \ldots, \ell - 1$. Analogously, we say $v_1 v_2 \cdots v_\ell v_1$ is a *gray cycle* if $v_1 v_\ell \in \mathrm{EG}(K)$ and $v_i v_{i+1} \in \mathrm{EG}(K)$ for $i = 1, \ldots, \ell - 1$.

1.2 Literature review and known results

The edit distance in graphs was introduced independently by Alon and Stav [1] and by Axenovich, Kézdy, and Martin [3]. Applications of edit distance problems to biology and computer science are discussed in [1, 2, 3, 15]. The question considered is "Given a class of graphs $\mathcal{H}$ what is the minimum number $m = m(n)$ such that for every graph on n vertices, there is a set of m edge-additions and edge-deletions that ensure the resultant graph is a member of $\mathcal{H}$?" The work of Axenovich et al. gives general bounds for $\mathrm{dist}(n, \mathrm{Forb}(H))$ in terms of the so-called binary chromatic number.

Definition 10. *The binary chromatic number of a graph G, denoted $\chi_B(G)$, is the least integer $k+1$ such that, for all $c \in \{0, \ldots, k+1\}$, there exists a partition of $V(G)$ into c cliques and $k + 1 - c$ independent sets.*

The binary chromatic number was first introduced by Prömel and Steger as a parameter

τ in [20], and then was generalized as a so-called colouring number of a hereditary property in [8, 9].

Theorem 11 ([3])**.** *Given a graph H with binary chromatic number $\chi_B(H) = k+1$,*

$$\left(\frac{1}{2k} - o(1)\right)\binom{n}{2} \le \mathrm{dist}(n, \mathrm{Forb}(H)) \le \frac{1}{k}\binom{n}{2}.$$

It follows from the theorem above that $\mathrm{dist}(n, \mathrm{Forb}(H)) = (1+o(1))\frac{n^2}{4k}$ for a self-complementary graph H. The authors also gave exact results for certain graphs.

Theorem 12 ([3])**.** *If $H \in \{K_3, \overline{K_3}, K_{1,2}, \overline{K_{1,2}}\}$, then* $\mathrm{dist}(n, \mathrm{Forb}(H)) = \binom{\lceil n/2 \rceil}{2} + \binom{\lfloor n/2 \rfloor}{2}$.

Alon and Stav prove in [1] that for every hereditary property $\mathcal{H}$, there is a $p^* = p^*_{\mathcal{H}}$ such that the Erdős-Rényi random graph $G(n, p^*)$ is asymptotically extremal.

Theorem 13 ([1])**.** *Let $\mathcal{H}$ be a hereditary property. Then there exists $p^* = p^*_{\mathcal{H}} \in [0, 1]$ such that, with high probability,*

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

Using this fact it is shown that $\lim_{n\to\infty} \mathrm{dist}(n, \mathcal{H})/\binom{n}{2}$ exists [6], this limit is denoted by $d^*_{\mathcal{H}}$. A graph property is called *complement invariant* if it is closed under taking the complement of a graph. In [2] it is shown that $p^*_{\mathcal{H}} = 1/2$ for a hereditary complement invariant graph property $\mathcal{H}$, as well as $d^*_{\mathrm{Forb}(K_{1,3})} = p^*_{\mathrm{Forb}(K_{1,3})} = 1/3$. This result is generalized in [6] to determine the values $p^*_{\mathcal{H}}$ and $d^*_{\mathcal{H}}$ for hereditary properties of the form $\mathcal{H} = \mathrm{Forb}(K_a + E_b)$, where $K_a + E_b$ is a disjoint union of a complete graph K_a and an empty graph E_b.

The value of $d^*_{\mathcal{H}}$ is often determined using the edit distance function of a hereditary property defined in (1.1).

Theorem 14 ([6])**.** *Let $\mathcal{H}$ be a hereditary property. Then*

$$\mathrm{ed}_{\mathcal{H}}(p) = \lim_{n\to\infty} \mathbb{E}[\mathrm{dist}(G(n, p), \mathcal{H})]/\binom{n}{2}.$$

The edit distance function has useful properties such as being continuous and concave down [6], so we sometimes compute $d^*_{\mathcal{H}}$ without determining the entire edit distance function.

Furthermore, we can compute $\mathrm{ed}_{\mathcal{H}}(1/2)$ in terms of $\chi_B(\mathcal{H})$ because by Theorem 15 below every 1/2-core CRG has only gray edges. In particular, $\mathrm{ed}_{\mathcal{H}}(1/2) = \frac{1}{2(\chi_B(\mathcal{H})-1)}$ [3].

Edit distance function and CRGs are closely related to the study of 2-coloring of the edges of the complete graph by Marchant and Thomason [12]. The connection between two problems studied is discussed in [12] and in a survey by Thomason [26]. Below are the essential results for solving the edit distance problem that come from this study stated in edit distance terminology. Theorem 15 gives a structural classification of p-core CRGs.

Theorem 15 ([12], Theorem 3.23)**.** *Let K be a p-core CRG. Then all edges of K are gray, apart from*

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

Theorem 16 ([12], Theorem 3.25)**.** *Let $\mathcal{H}$ be a hereditary property and $0 \le p \le 1$. Then there is a p-core CRG, $K \in \mathcal{K}(\mathcal{H})$, such that $\mathrm{ed}_{\mathcal{H}}(p) = g_K(p)$. That is, $\mathrm{ed}_{\mathcal{H}}(p) = \min\{g_K(p) : K \in \mathcal{K}(\mathcal{H})\}$.*

A sub-CRG K' of a CRG K is called a *component* of K if for all $v \in V(K')$ and all $w \in V(K) - V(K')$ the edge vw is gray. The following theorem is useful for determining g_K from the g functions of components of K.

Theorem 17 ([13])**.** *Let K be a CRG with components $K^{(1)}, \ldots, K^{(\ell)}$. Then*

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

Many results were obtained using a powerful tool called symmetrization. This term was used by Pikhurko [19] for the method observed by Sidorenko [24]. The version of symmetrization we work with uses the matrix $\mathbf{M}_K(p)$ defined in (1.4).

Theorem 18 ([13])**.** *Let $p \in [0,1]$ and let K be a p-core CRG with associated matrix $\mathbf{M}_K(p)$. If $\mathbf{x}^*$ is an optimal solution to the quadratic program (1.3), then*

$$\mathbf{M}_K(p) \cdot \mathbf{x}^* = g_K(p)\mathbf{1}.$$

It follows from Theorem 18 that for any white vertex v of K,

$$pd_W(v) + (1-p)d_B(v) = g_K(p). \tag{1.5}$$

Using the characterization of p-core CRGs given in Theorem 15 and (1.5), the weighted gray degree of each vertex of a p-core CRG can be computed.

Theorem 19 ([13]). *Let $p \in (0,1)$ and K be a p-core CRG with optimal weight vector $\mathbf{x}$.*

(i) If $p \le 1/2$ then $\mathbf{x}(v) = g_K(p)/p$ for all $v \in VW(K)$ and

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

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

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

Theorem 19 also gives an upper bound for weights of individual vertices.

Theorem 20 ([13]). *Let $p \in (0,1)$ and K be a p-core CRG with optimal weight vector $\mathbf{x}$.*

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

(ii) If $p \ge 1/2$ then $\mathbf{x}(v) \le g_K(p)/p$ for all $v \in VW(K)$.

We finish this section with some known results for certain hereditary properties.

Theorem 21 ([12]). *Let $p \in [0,1]$ and $\mathcal{H} = \mathrm{Forb}(C_6^*)$, where C_6^* is the 6-cycle with a diagonal. Then $\mathrm{ed}_{\mathcal{H}}(p) = \min\{p/(1+2p), (1-p)/2\}$ and $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (1/2, 1/4)$.*

The following theorem is a result on the graph H_9 shown in Figure 1.4 introduced in [6] whose $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}})$ values cannot be determined by the clique spectrum.

Theorem 22 ([14]). *Let $p \in [0,1]$ and H_9 be the graph shown in Figure 1.4. Let $\mathcal{H} = \mathrm{Forb}(H_9)$. Then $\mathrm{ed}_{\mathcal{H}}(p) = \min\{\frac{p}{3}, \frac{p}{1+4p}, \frac{1-p}{2}\}$ and $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (1/8(1+\sqrt{17}), 1/8(7-\sqrt{17}))$.*

Theorem 23 ([6]). *Let $\mathcal{H} = \mathrm{Forb}(K_{3,3})$. Then $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (\sqrt{2}-1, 3-2\sqrt{2})$.*

Theorem 24 ([16]). *Let $\mathcal{H} = \mathrm{Forb}(K_{2,t})$ and $p \in [0,1]$.*

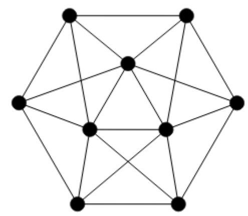

Figure 1.4 The graph H_9.

(i) If $t = 3$, *then* $\mathrm{ed}_{\mathcal{H}}(p) = \min\{p(1-p), \frac{1-p}{2}\}$ *and* $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (1/2, 1/4)$.

(ii) If $t = 4$, *then* $\mathrm{ed}_{\mathcal{H}}(p) = \min\{p(1-p), \frac{7p+1}{15}, \frac{1-p}{3}\}$ *and* $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (1/3, 2/9)$.

(iii) If $t \geq 5$ *and odd, then* $p^*_{\mathcal{H}} \supseteq \left[\frac{2t-1}{t(t+1)}, \frac{2}{t+1}\right]$ *and* $d^*_{\mathcal{H}} = \frac{1}{t+1}$.

Theorem 25 ([14])**.** *Let H be a split graph with independence number $\alpha \geq 2$ and clique number $\omega \geq 2$. Let $p \in [0,1]$ and $\mathcal{H} = \mathrm{Forb}(H)$. Then $\mathrm{ed}_{\mathcal{H}}(p) = \min\{\frac{p}{\omega-1}, \frac{1-p}{\alpha-1}\}$ and hence $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (\frac{\omega-1}{\alpha+\omega-2}, \frac{1}{\alpha+\omega-2})$.*

Theorem 26 ([12, 13])**.** *Let $\mathcal{H} = \mathrm{Forb}(C_h)$. Then*

- *for* $h = 3$, $\mathrm{ed}_{\mathcal{H}}(p) = \frac{p}{2}$ *and* $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (1, 1/2)$,
- *for* $h = 4$, $\mathrm{ed}_{\mathcal{H}}(p) = p(1-p)$ *and* $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (1/2, 1/4)$,
- *for* $h = 5$, $\mathrm{ed}_{\mathcal{H}}(p) = \min\{\frac{p}{2}, \frac{1-p}{2}\}$ *and* $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (1/2, 1/4)$,
- *for* $h = 6$, $\mathrm{ed}_{\mathcal{H}}(p) = \min\{p(1-p), \frac{1-p}{2}\}$ *and* $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (1/2, 1/4)$
- *for* $h = 7$, $\mathrm{ed}_{\mathcal{H}}(p) = \min\{\frac{p}{2}, \frac{p(1-p)}{1+p}, \frac{1-p}{3}\}$ *and* $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (\sqrt{2}-1, 3-2\sqrt{2})$,
- *for* $h = 8$, $\mathrm{ed}_{\mathcal{H}}(p) = \min\{\frac{p(1-p)}{1+p}, \frac{1-p}{3}\}$ *and* $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (\sqrt{2}-1, 3-2\sqrt{2})$,
- *for* $h = 9$, $\mathrm{ed}_{\mathcal{H}}(p) = \min\{\frac{p}{2}, \frac{1-p}{4}\}$ *and* $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (1/3, 1/6)$,
- *for* $h = 10$ *and* $p \in [1/7, 1]$, $\mathrm{ed}_{\mathcal{H}}(p) = \min\{\frac{p(1-p)}{1+2p}, \frac{1-p}{4}\}$ *and* $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = ((\sqrt{3}-1)/2, (2-\sqrt{3})/2$.

The following theorems are of particular interest to this thesis. Some cases of the squared cycles were investigated and the above results for cycles were further generalized.

Theorem 27 ([18])**.** *Let* $\mathcal{H} = \mathrm{Forb}(C_h)$ *with* $h \geq 4$*.*

(i) If h *is odd, then* $\mathrm{ed}_{\mathcal{H}}(p) = \min\left\{\frac{p}{2}, \frac{p(1-p)}{1-p+(\lceil h/3\rceil-1)p}, \frac{1-p}{\lceil h/2\rceil-1}\right\}$ *for all* $p \in [0,1]$*.*

(ii) If h *is even, then* $\mathrm{ed}_{\mathcal{H}}(p) = \min\left\{\frac{p(1-p)}{1-p+(\lceil h/3\rceil-1)p}, \frac{1-p}{\lceil h/2\rceil-1}\right\}$ *for all* $p \in [\lceil h/3\rceil^{-1}, 1]$*.*

Theorem 28 ([18])**.** *Let* $\mathcal{H} = \mathrm{Forb}(C_h^2)$ *and* $p \in [0,1]$*. Then*

(i) for $h = 8$*,* $\mathrm{ed}_{\mathcal{H}}(p) = \min\{\frac{p}{3}, \frac{p(1-p)}{2-p}, \frac{1-p}{2}\}$ *and* $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (2-\sqrt{2}, 3-\sqrt{2})$*,*

(ii) for $h = 9$*,* $\mathrm{ed}_{\mathcal{H}}(p) = \min\{\frac{p(1-p)}{2-p}, \frac{p(1-p)}{1+p}\}$ *and* $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (1/2, 1/6)$*,*

(iii) for $h = 10$*,* $\mathrm{ed}_{\mathcal{H}}(p) = \min\{\frac{1-p}{3}, \frac{p}{3}\}$ *and* $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (1/2, 1/6)$*,*

(iv) for $h = 11$ *and* $p \leq 1/2$*,* $\mathrm{ed}_{\mathcal{H}}(p) = \min\{\frac{1-p}{3}, \frac{p}{3}, \frac{p(1-p)}{2}\}$ *and* $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (1/2, 1/8)$*,*

(v) for $h = 12$ *and* $p \leq 1/2$*,* $\mathrm{ed}_{\mathcal{H}}(p) = \min\{\frac{1-p}{3}, \frac{p(1-p)}{2}\}$ *and* $(p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) = (1/2, 1/8)$*.*

There were some errors in the proof of Theorem 27 presented in [18]. We have corrected them and generalized the results to powers of cycles in [7].

1.3 Main Results

The main results of this thesis are from the submitted paper [7]. Proofs will be given in Chapters 2. We first establish the $\gamma_{\mathcal{H}}$ function in Theorem 29 below, which gives an upper bound for the edit distance function for all $p \in [0,1]$. Then we show in Theorem 30 that these two functions agree when $(t+1) \nmid h$ for all $0 \leq p \leq 1$ and when $(t+1)|h$ for all but small values of p. Other related results will be given in Chapter 3 which are based on work in progress [6].

Theorem 29. *Let* $t \geq 1$ *and* $h \geq \max\{t(t+1), 4\}$ *be integers, and let* $p_0 = {\ell_t}^{-1}$*, where* $\ell_a = \left\lceil \frac{h}{t+a+1} \right\rceil$ *for* $a \in \{0, \ldots, t\}$*. Then for all* $p \in [0,1]$ *and* $\mathcal{H} = \mathrm{Forb}(C_h^t)$*,*

$$\gamma_{\mathcal{H}}(p) = \min_{a \in \{0,1,\ldots,t\}} \left\{ \frac{p(1-p)}{a(1-p) + (\ell_a - 1)\,p} \right\}, \qquad \text{if } (t+1) \mid h;$$

$$\gamma_{\mathcal{H}}(p) = \min_{a \in \{0,1,\ldots,t\}} \left\{ \frac{p}{t+1}, \frac{p(1-p)}{a(1-p) + (\ell_a - 1)\,p} \right\}, \qquad \text{if } (t+1) \nmid h.$$

Note: If $a = 0$, then $\frac{p(1-p)}{a(1-p)+(\ell_a-1)p} = \frac{p(1-p)}{(\ell_0-1)p}$, which we define to be $\frac{1-p}{\ell_0-1}$ at $p = 0$.

Theorem 30. *Let $t \geq 1$ and $h \geq 2t(t+1)+1$ be positive integers and let $\mathcal{H} = \mathrm{Forb}(C_h^t)$. If either $(t+1) \not| h$ with $0 \leq p \leq 1$ or $(t+1)|h$ with $p_0 \leq p \leq 1$, then*

$$\mathrm{ed}_{\mathcal{H}}(p) = \gamma_{\mathcal{H}}(p). \tag{1.6}$$

The known result for cycles is the special case of the theorems above with $t = 1$.

Corollary 31. *Let $h \geq 5$ be a positive integer and $\mathcal{H} = \mathrm{Forb}(C_h)$.*

- *If h is even, then for $\lceil h/3 \rceil^{-1} \leq p \leq 1$,*

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

- *If h is odd, then for $0 \leq p \leq 1$,*

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

Notice that when $t = 1$, the furthest graph from $\mathrm{Forb}(C_h)$ is a graph which has density $p^* = 1/(\lceil h/2 \rceil - \lceil h/3 \rceil + 1)$ when $h \geq 5$ and $h \notin \{7, 8, 10, 16\}$, and has density $p^* = 1/(1+\sqrt{\lceil h/3 \rceil - 1})$ when $h \in \{4, 7, 8, 10, 16\}$. Also, observe that the maximum value of the edit distance function can be an irrational number.

Our proof techniques often require us to compare the g_K function of a CRG to one of the individual functions that are given in Theorem 29. However, when h is large enough at most three of these functions are necessary to define $\gamma_{\mathcal{H}}$.

Corollary 32. *Let $t \geq 2$ and $h \geq 4t^2 + 10t + 24$ be positive integers and let $\mathcal{H} = \mathrm{Forb}(C_h^t)$. Recall that $\ell_t = \left\lceil \frac{h}{2t+1} \right\rceil$ and $p_0 = \ell_t^{-1}$. Then*

- *If $(t+1) \mid h$, then for $p_0 \leq p \leq 1$,*

$$\mathrm{ed}_{\mathcal{H}}(p) = \min\left\{ \frac{p(1-p)}{t(1-p)+(\ell_t - 1)p}, \frac{1-p}{\ell_0 - 1} \right\}.$$

- *If $(t+1) \not| h$, then for $0 \leq p \leq 1$,*

$$\mathrm{ed}_{\mathcal{H}}(p) = \min\left\{ \frac{p}{t+1}, \frac{p(1-p)}{t(1-p)+(\ell_t - 1)p}, \frac{1-p}{\ell_0 - 1} \right\}.$$

Theorem 30 establishes the edit distance function for $\mathrm{Forb}(C_h^t)$ over all $p \in [0,1]$ when $t+1$ does not divide h. When $t+1$ divides h the function is not known for $p \in (0, p_0)$. While for the cycles it was sufficient to consider black-vertex CRGs, the powers of cycles require various constructions with both white and black vertices. In particular, the most complicated case for C_h^t is the one when the CRG under question has $t-1$ white vertices. We show that if $K \in \mathcal{K}(\mathrm{Forb}(C_h^t))$ is a p-core CRG with $p < 1/2$ which has $a \neq t-1$ white vertices, then $g_K(p) = \gamma_{\mathrm{Forb}(C_h^t)}(p)$. Therefore, to solve the problem for the remaining case when $t+1$ divides h, and p is small, we only need to consider CRGs with exactly $t-1$ white vertices. A particular barrier to this is Lemma 40 which requires $p \geq p_0$ to ensure that the graph induced by the black vertices and gray edges of the CRG has the property that any two vertices have at least one common neighbor. Such a condition need not hold for small p.

CHAPTER 2. EDIT DISTANCE FUNCTION FROM FORBIDDEN INDUCED POWERS OF CYCLES

This chapter contains proofs of the main results from Section 1.3 as well as other results.

2.1 Proof of Theorem 29: Computation of the $\gamma_{\mathcal{H}}$ function

We compute the $\gamma_{\mathcal{H}}$ function, which gives an upper bound for the edit distance function. For any $t \geq 1$, $h \geq 2t+2$ and $a \in \{0, \dots, t\}$, we denote $\ell_a = \left\lceil \frac{h}{t+a+1} \right\rceil$. The first observation is the value of the chromatic number of C_h^t, denoted $\chi(C_h^t)$.

Proposition 33 ([21])**.** *Let $t \geq 1$ and $h \geq \max\{t+1, 3\}$ be positive integers. Let $h = q(t+1)+r$, where $r \in \{0, \dots, t\}$. Then, $\chi(C_h^t) = t + \lceil r/q \rceil + 1$. In particular, if $h \geq \max\{t(t+1), 3\}$, then*

$$\chi(C_h^t) = \begin{cases} t+1, & \text{if } (t+1) \mid h, \\ t+2, & \text{if } (t+1) \nmid h. \end{cases}$$

Let $h \geq \max\{t(t+1), 2t+2\}$ and $\chi = \chi(C_h^t)$. Denote the vertices of C_h^t by $\{1, \dots, h\}$ such that distinct i and j are adjacent if and only if $|i-j| \leq t \pmod h$. For each $a \in \{0, \dots, t\}$, we first show that $(a, \ell_a - 1) \in \Gamma = \Gamma(\mathrm{Forb}(C_h^t))$ and $(a, \ell_a) \notin \Gamma$. We then further show that if $\chi > t+1$ then $\{(t+1, 0), \dots, (\chi-1, 0)\} \subset \Gamma$ and $(t+1, 1) \notin \Gamma$. This will imply that $\Gamma^* \subseteq \{(a, \ell_a - 1) : a = 0, 1, \dots, t\} \cup \{(\chi - 1, 0)\}$, which is a stronger result than we need.

Case 1: $a \in \{0, \dots, t\}$**.**

First, we show that $(a, \ell_a - 1) \in \Gamma$. By contradiction, assume there is a partition of $V(C_h^t)$ into a independent sets and $\ell_a - 1$ cliques. Let $k = \ell_a - 1$, and let $C_1, \dots, C_k$ be the cliques. We may assume that the vertices in each C_i are consecutive. This is because if j_1 and j_2 are in the same clique, then by the nature of adjacency in the power of a cycle, every vertex between j_1

and j_2 is adjacent to every member of the clique, and hence can be added to the clique. Thus, $|C_i| \leq t+1$ for $i = 1, \ldots, k$.

For $i = 1, \ldots, k-1$, let B_i be the set of vertices between C_i and C_{i+1}, and let B_k be the set of vertices between C_k and C_1. The sets B_i might or might not be empty. If some $|B_i| \geq a+1$, then the first $a+1 \leq t+1$ vertices form a clique and so must be in different independent sets, which is not possible since there are only a independent sets. Therefore, $|B_i| \leq a$ for $i = 1, \ldots, k$.

Consequently, we need $k(t+a+1) \geq h$ in order to cover C_h^t with a independent sets and k cliques. Hence, $k \geq \ell_a$, a contradiction to our choice of k. Thus $(a, \ell_a - 1) \in \Gamma$ for $a = 0, \ldots, t$.

Next, we show that $(a, \ell_a) \notin \Gamma$. Again, let $k = \ell_a - 1$. For $i = 1, \ldots, k$, let $S_i = \{(i-1)(t+a+1)+1, \ldots, i(t+a+1)\}$ and let $S_{k+1} = \{1, \ldots, h\} - \cup_{i=1}^{k} S_i$. For $i = 1, \ldots, k$, let C_i be the first $t+1$ vertices of S_i and let C_{k+1} be the first $\min\{t+1, |S_{k+1}|\}$ vertices of S_{k+1}. For $j = 1, \ldots, a$, let A_j consist of the $(t+1+j)^{\text{th}}$ vertex of $S_1, \ldots, S_k$ and the $(t+1+j)^{\text{th}}$ vertex of S_{k+1} if $|S_{k+1}| \geq t+1+j$.

The sets $(A_1, \ldots, A_a, C_1, \ldots, C_{k+1})$ form a partition of $V(C_h^t)$. Clearly each C_i, $i = 1, \ldots, k$, is a clique of size $t+1$ and since there is a clique of size $t+1$ between pairs of vertices in each A_j, each A_j is an independent set. Thus $(a, \ell_a) \notin \Gamma$ for $a = 0, \ldots, t$.

Case 2: $a \geq t+1$**.**

If $(t+1) \mid h$, then Proposition 33 gives that C_h^t can be partitioned into $t+1$ independent sets and so $(t+1, 0) \notin \Gamma$. If $(t+1) \nmid h$, then Proposition 33 gives that $\chi \geq t+2$ and since C_h^t cannot be partitioned into fewer than χ independent sets, we have $(t+1, 0), \ldots, (\chi - 1, 0) \in \Gamma$. Since C_h^t can be partitioned into χ independent sets, $(\chi, 0) \notin \Gamma$.

Finally, let $k = \lceil h/(t+1) \rceil - 1$. For $j = 1, \ldots, t+1$, let $A_j = \{(i-1)(t+1)+j : i = 1, \ldots, k\}$. Let $C_0 = \{k(t+1)+1, \ldots, h\}$. The sets $(A_1, \ldots, A_{t+1}, C_0)$ form a partition of $V(C_h^t)$. Clearly, C_0 is a clique of size at most $t+1$ and since there are at least t vertices between pairs of vertices in each A_j, each A_j is an independent set. Thus $(t+1, 1) \notin \Gamma$.

Using Theorem 17, if $h = q(t+1) + r$ where $r \in \{0, \ldots, t\}$, then

$$\gamma_{\mathcal{H}}(p) = \min_{a \in \{0,1,\ldots,t\}} \left\{ \frac{p(1-p)}{a(1-p) + (\ell_a - 1)\, p} \right\}, \qquad \text{if } r = 0;$$
$$\gamma_{\mathcal{H}}(p) = \min_{a \in \{0,1,\ldots,t\}} \left\{ \frac{p}{t + \lceil r/q \rceil}, \frac{p(1-p)}{a(1-p) + (\ell_a - 1)\, p} \right\}, \qquad \text{if } r \neq 0.$$

Restricting ourselves to $h \geq \min\{t(t+1), 4\}$, we have the result in the statement of the theorem. □

2.2 Forbidden cycles

Before we can prove Theorem 30, we need to study the properties of the CRGs into which C_h^t does not embed. Recall that we may assume $h \geq 2t + 2$. An important property of such CRGs is that the set of lengths of gray cycles on black vertices is restricted, as is shown in Lemma 36. Its proof needs inequalities in Facts 34 and 35. For completeness, we give their proofs in Section 2.4.

Fact 34. *Let h, x, y be positive integers. Then*

(a) $\lfloor h/x \rfloor \geq y$ if and only if $\lfloor h/y \rfloor \geq x$.

(b) $\lceil h/x \rceil \leq y$ if and only if $\lceil h/y \rceil \leq x$.

Fact 35. *Let $t \geq 1$, $h \geq \max\{t(t-1), 2t+2\}$, and $a \in \{0, \ldots, t-1\}$ be positive integers. Then $\left\lceil \frac{h}{t+a+1} \right\rceil \leq \left\lfloor \frac{h}{t} \right\rfloor$.*

Lemma 36 is a key lemma in proving the main result, Theorem 30.

Lemma 36. *Let $p \in (0, 1/2]$ and let $t \geq 1$ and $h \geq 2t + 2$ be integers. Let $\tilde{K}$ be a p-core CRG with exactly a white vertices such that $C_h^t \not\mapsto \tilde{K}$. Let K be the sub-CRG of $\tilde{K}$ induced by the set of all black vertices of $\tilde{K}$. Then, the following occurs:*

(a) If $a \in \{0, \ldots, t-1\}$ and $h \geq t^2 - t$, then K has no gray cycle which has length in $\left\{ \left\lceil \frac{h}{t+a+1} \right\rceil, \ldots, \left\lfloor \frac{h}{t} \right\rfloor \right\}$.

(b) If $a = t$, then $|V(K)| \leq \ell_t - 1$.

(c) If $a \geq t+1$, then $(t+1) \nmid h$ and $V(K) = \emptyset$.

Note: We interpret a gray cycle of length 2 to be a gray edge.

Proof of Lemma 36. Denote the vertices of C_h^t by $\{1, \ldots, h\}$ such that distinct i and j are adjacent if and only if $|i-j| \leq t \pmod{h}$.

Partition: Let K have a gray cycle on vertex set $\{v_1, \ldots, v_k\}$ such that $v_i v_{i+1}$ is a gray edge, where the indices are taken modulo k. We describe a partition of $V(C_h^t)$, which gives an interval of forbidden gray cycle lengths. We will construct at most a independent sets and k cliques $C_1, \ldots, C_k$ such that there is no edge between nonconsecutive cliques.

Partition $V(C_h^t)$ into k sets of consecutive vertices $S_1, \ldots, S_k$, with each set S_i of size either $\lceil h/k \rceil$ or $\lfloor h/k \rfloor$. We will construct at most a independent sets and k cliques $C_1, \ldots, C_k$ with $C_i \subseteq S_i$ such that there is no edge between C_i and $C_{i'}$ unless $|i - i'| \equiv 1 \pmod{k}$.

If $a = 0$, then simply let $C_i = S_i$ for $i = 1, \ldots, k$. Using Fact 37, each C_i has size at least t and so nonconsecutive sets have no edge between them. Fact 37 is a simple observation of number theory, related to Frobenius numbers [25].

Fact 37. *A set of size h can be partitioned into sets of size t or $t+1$ if and only if $h \geq t(t-1)$. Moreover, for any $k \in \{\lceil h/(t+1) \rceil, \ldots, \lfloor h/t \rfloor\}$, such a partition exists with exactly k parts.*

So, we assume $a \geq 1$ and choose $a' \in \{\lfloor h/k \rfloor - t, \lceil h/k \rceil - (t+1)\}$ such that $0 \leq a' \leq a$. This is possible as long as both (a) $0 \leq \lfloor h/k \rfloor - t$ and (b) $\lceil h/k \rceil - (t+1) \leq a$. (This is only nontrivial if $k \mid h$, in which case at least one of the two choices of a' will be in $\{0, \ldots, a\}$.)

If $a' = 0$, again let $C_i = S_i$ for $i = 1, \ldots, k$. If $a' \geq 1$, let A_j consist of the j^{th} vertex of each of $S_1, \ldots, S_k$ and let $C_i = S_i - \cup_{j=1}^{a'} A_j$. Observe that if $a' \geq 1$, then $|S_i| \geq t+1$ and so there are at least t vertices between each pair of vertices in every A_j. Therefore, A_j is an independent set for $j = 1, \ldots, a'$. We have $|C_i| \leq t+1$, so C_i is a clique for $i = 1, \ldots, k$. In addition, $|C_i| \geq t$ and so there are no edges between C_i and $C_{i'}$ unless $|i - i'| \pmod{k}$.

The mapping, for all $a \geq 0$, is as follows: each A_j is mapped to a different white vertex and C_i to v_i for $i = 1, \ldots, k$. If $a = 0$, Fact 37 gives that K has no cycle with length in

$\{\lceil h/(t+1)\rceil, \ldots, \lfloor h/t \rfloor\}$. If $a \geq 1$, Fact 34, gives that K has no cycle with length in

$$\left\{\left\lceil \frac{h}{t+a+1} \right\rceil, \ldots, \left\lfloor \frac{h}{t} \right\rfloor\right\}, \tag{2.1}$$

and (2.1) is valid in the case of $a = 0$ also.

Case *(a)*: $a \in \{0, \ldots, t-1\}$ and $h \geq t^2 - t$.

The result is given by (2.1). It suffices to show that $\left\lceil \frac{h}{t+a+1} \right\rceil \leq \left\lfloor \frac{h}{t} \right\rfloor$. Fact 35 gives that this holds if $h \geq t^2 - t$.

Case *(b)*: $a = t$.

In this case, we use another construction. Partition $V(C_h^t)$ into $k+1$ consecutive parts, $S_1, \ldots, S_{k+1}$, where $k = \lceil h/(2t+1) \rceil - 1$ and $r = h - (k-1)(2t+1)$. Since $h \geq 2t+2$, $k \geq 1$. Let $|S_1| = \cdots = |S_{k-1}| = 2t+1$, $|S_k| = \lceil r/2 \rceil$ and $|S_{k+1}| = \lfloor r/2 \rfloor$. Note that $t+1 \leq |S_{k+1}| \leq |S_k| \leq 2t+1$.

For $j = 1, \ldots, t$, let A_j consist of the j^{th} vertex in each part and let $C_i = S_i - \bigcup_{j=1}^{t} A_j$.

Each of $A_1, \ldots, A_t$ is an independent set. Furthermore, there are no edges between C_i and $C_{i'}$ if $i \neq i'$. Therefore, K has at most $k = \lceil h/(2t+1) \rceil - 1$ vertices; otherwise, $A_1, \ldots, A_t$ can be mapped arbitrarily to each of the t white vertices and $C_1, \ldots, C_{k+1}$ can be mapped arbitrarily to $k+1$ different black vertices in K.

Case *(c)*: $a \geq t+1$.

If $(t+1) \mid h$, then $\chi(C_h^t) = t+1$ and $\tilde{K}$ having at least $t+1$ white vertices means that C_h^t embeds in $\tilde{K}$, a contradiction. If $(t+1) \not\mid h$, then partition $V(C_h^t)$ into $k = \lfloor h/(t+1) \rfloor + 1$ parts $S_1, \ldots, S_k$ of consecutive vertices, each of $S_1, \ldots, S_{k-1}$ of size $t+1$. For $j = 1, \ldots, t+1$, let A_j consist of the j^{th} vertex in each S_i for $i = 1, \ldots, k-1$. The graph induced by $V(C_h^t) - \bigcup_{j=1}^{t+1} A_j$ forms a clique of size at most t in S_k. Since all vertices in K are black, this clique will embed into any vertex of $V(K)$. Thus $V(K) = \emptyset$. $\square$

2.3 Proof of Theorem 30: $\mathrm{ed}_{\mathcal{H}} = \gamma_{\mathcal{H}}$

We use Lemma 36 to prove Theorem 30. Recall that $h \geq 2t(t+1)+1 \geq t(t+1)$. By Proposition 33, this means that $\chi(C_h^t) = t+1$ if $(t+1) \mid h$ and $\chi(C_h^t) = t+2$ if $(t+1) \nmid h$. The lemmas and facts uses in this section are stated without proofs, we defer their proofs to Section 2.4.

Proof of Theorem 30. By definition, $\mathrm{ed}_{\mathcal{H}}(p) \leq \gamma_{\mathcal{H}}(p)$ for all $p \in [0,1]$, so we need to show the inequality in other direction.

Case 1: $p \in [1/2, 1]$.

Fact 38 below establishes that $\gamma_{\mathcal{H}}(p) = \frac{1-p}{\ell_0 - 1}$ for $p \in [1/2, 1]$.

Fact 38. *Let h and t be positive integers. If $h \geq (t+1)^2 + 1$, then*

$$\frac{1-p}{\ell_0 - 1} \leq \frac{p}{t+1}.$$

For $a \in \{1, \ldots, t\}$ if $h \geq (t+1)(t+a)+1$, then for all $p \in [1/2, 1]$,

$$\frac{1-p}{\ell_0 - 1} \leq \frac{p(1-p)}{a(1-p) + (\ell_a - 1)p}.$$

Note: The condition $h \geq 2t(t+1)+1$ suffices to achieve all of the conclusions in Fact 38.

By Proposition 39 below, $\mathrm{ed}_{\mathcal{H}}(p) = \gamma_{\mathcal{H}}(p)$ for the two values of $p \in \{1/2, 1\}$.

Proposition 39 (Balogh-Martin [6])**.** *If $\mathcal{H}$ is a hereditary property, then $\mathrm{ed}_{\mathcal{H}}(1/2) = \gamma_{\mathcal{H}}(1/2)$. Moreover, if $K_\ell \in \mathcal{H}$ for all positive integers ℓ, then $\mathrm{ed}_{\mathcal{H}}(1) = \gamma_{\mathcal{H}}(1) = 0$ and if $\overline{K_\ell} \in \mathcal{H}$ for all positive integers ℓ, then $\mathrm{ed}_{\mathcal{H}}(0) = \gamma_{\mathcal{H}}(0) = 0$.*

We have $\mathrm{ed}_{\mathcal{H}}(p) \leq \gamma_{\mathcal{H}}(p)$ and the two functions are equal at $p = 1/2$ and at $p = 1$. The function $\gamma_{\mathcal{H}}(p)$ is linear over $p \in [1/2, 1]$ for $h \geq 2t(t+1)+1$. Since $\mathrm{ed}_{\mathcal{H}}(p)$ is continuous and concave down, we may conclude that $\mathrm{ed}_{\mathcal{H}}(p) = \gamma_{\mathcal{H}}(p) = \frac{1-p}{\ell_0 - 1}$ for $p \in [1/2, 1]$. This concludes Case 1.

Note that Proposition 39 gives $\mathrm{ed}_{\mathcal{H}}(0) = \gamma_{\mathcal{H}}(0) = 0$. Let $p \in (0, 1/2)$ and $\mathrm{ed}_{\mathcal{H}}(p) = g_{\tilde{K}}(p)$ for some p-core CRG $\tilde{K}$. Assume by contradiction that $g_{\tilde{K}}(p) < \gamma_{\mathcal{H}}(p)$. Suppose $\tilde{K}$ has a white vertices. Recall that for any $t \geq 1$, $h \geq 2t+2$ and $a \in \{0, \dots, t\}$, we denote $\ell_a = \left\lceil \frac{h}{t+a+1} \right\rceil$. We consider several cases and show that we arrive at a contradiction in each case.

Case 2: $a \geq t$ and $p \in (0, 1/2)$.

If $a \geq t+1$, then by Lemma 36*(c)*, $V(K) = \emptyset$. As long as $h \geq \max\{t(t+1), 3\}$, Proposition 33 gives that $\chi(C_h^t) \leq t+2$ with equality only if $(t+1) \not| h$. Thus, $a = t+1$ and Theorem 17 gives that $g_{\tilde{K}}(p) = p/(t+1)$, a contradiction to the assumption that $g_{\tilde{K}}(p) < \gamma_{\tilde{K}}(p)$.

If $a = t$, then Case *(b)* of Lemma 36 gives that $|V(K)| \leq \ell_t - 1$. Consequently, $g_K(p) \geq \frac{1-p}{\ell_t - 1}$. We can partition $\tilde{K}$ into $t+1$ sub-CRGs, K and t white vertices, and use Theorem 17 to conclude that

$$\begin{aligned}\left(g_{\tilde{K}}(p)\right)^{-1} &\leq tp^{-1} + \left(\frac{1-p}{\ell_t - 1}\right)^{-1} \\ g_{\tilde{K}}(p) &\geq \frac{p(1-p)}{t(1-p) + (\ell_t - 1)p}.\end{aligned}$$

Hence, $\mathrm{ed}_{\mathcal{H}}(p) \geq \gamma_{\mathcal{H}}(p)$, again a contradiction. This concludes Case 2.

Case 3: $a \leq t-2$ and $p \in (0, 1/2)$.

Recall that $\tilde{K}$ is a CRG with a white vertices, with $0 \leq a \leq t-2$. By Theorem 17, $g_{\tilde{K}}(p)^{-1} = ap^{-1} + g_K^{-1}(p)$. Therefore,

$$g_K(p) < \left(\max_{a' \in \{0,1,\dots,t\}} \left\{ \frac{a'-a}{p} + \frac{\ell_{a'} - 1}{1-p} \right\} \right)^{-1} =: g_0(a, t; p). \tag{2.2}$$

Given our assumptions on $g_K(p)$, Lemma 40 gives lower bounds on the gray degree of vertices and the codegree of pairs of vertices. Recall that $\deg_G(v)$ denotes the *number* of gray neighbors of $v \in V(K)$.

Lemma 40. *Let $p \in (0, 1/2)$, $t \geq 1$ be an integer and $a \in \{0, \dots, t-1\}$. Let $p_0 = \ell_t^{-1} = \left\lceil \frac{h}{2t+1} \right\rceil^{-1}$. Let K be a p-core CRG with all black vertices such that $g_K(p) < g_0(a, t; p)$. Then*

(a) for every $v \in V(K)$, $\deg_G(v) \geq \ell_{a+1}$, and

(b) for every $v, w \in V(K)$*,*

$$\deg_G(v,w) \geq \begin{cases} \ell_{a+2}, & \text{if } a \leq t-2; \\ 1, & \text{if } a = t-1 \text{ and } p \geq p_0. \end{cases}$$

Note: Since $h \geq 2t+2$, it is the case that $\ell_{a+1} \geq 2$ for $a \leq t-1$ and $\ell_{a+2} \geq 2$ for $a \leq t-2$.

Now we consider the derived graph F with vertex set $V(K)$ and edge set $\mathrm{EG}(K)$. Using Lemma 40, the lower bound on the number of common gray neighbors of v and w gives a structural restriction on this graph. Note that the length of a path is defined to be the number of vertices in said path.

Lemma 41. *Fix integers* $t \geq 1$*,* $h \geq \max\{t(t-1), 2t+2\}$ *and* $a \in \{0, \ldots, t-1\}$*. Recall that* $\ell_a = \lceil h/(t+a+1) \rceil$ *and let* $L = \lfloor h/t \rfloor$*.*

Let F *be a graph with no cycle with length in* $\{\ell_a, \ldots, L\}$ *and every pair of vertices either has at least* $\ell_{a+2} \geq 2$ *common neighbors if* $a \leq t-2$ *or has at least* 1 *common neighbor if* $a = t-1$*. Then* F *has no cycle of length more than* $\ell_a - 1$*.*

Now we consider a maximum-length path in the graph F. If such a path can be made into a cycle, then Proposition 42 gives that F must be Hamiltonian. By Lemma 41, this means that $|V(K)| \leq \ell_a - 1$ and, as such, $g_K(p) \geq \frac{1-p}{\ell_a - 1}$, which is the g function for the CRG on $\ell_a - 1$ black vertices with all edges gray. This is a contradiction to our assumption in (2.2) by setting $a' = a$. Proposition 42 is a common argument in proofs of Hamiltonian cycle results, including the classical theorems of Dirac [10] and Ore [17].

Proposition 42. *Let* F *be a connected graph. If some path of maximum length forms a cycle, then* F *is Hamiltonian.*

So we may assume that every maximum-length path in F is not a cycle. Let $v_1 \cdots v_\ell$ be such a maximum length path. The common neighbors of v_1 and v_ℓ in F must be on this path, otherwise F has a longer path. From Lemma 40, it follows that v_1 and v_ℓ have at least $\ell_{a+2} \geq 2$ common neighbors on this path. However, Lemma 43 gives that there can only be one such neighbor, a contradiction.

Lemma 43. *Fix integers $t \geq 1$, $h \geq 2t+2$ and $a \in \{0, \dots, t-1\}$. Recall that $\ell_a = \lceil h/(t+a+1) \rceil$. Let F be a graph with no cycle of length longer than $\ell_a - 1$, with every vertex having degree at least $\ell_{a+1} \geq 2$ and with every pair of vertices having at least one common neighbor. Furthermore, let F have the property that no maximum length path forms a cycle.*

Let $v_1 \cdots v_\ell$ be a path of maximum length in F. Then v_1 and v_ℓ have exactly one common neighbor v_c on this path. Furthermore, $N(v_1) \subseteq \{v_2, \dots, v_c\}$ and $N(v_\ell) \subseteq \{v_c, \dots, v_\ell\}$.

This concludes Case 3.

Case 4: $a = t-1$ and $p \in [p_0, 1/2)$.

Recall that $\tilde{K}$ is a CRG with $a = t-1$ white vertices. By Proposition 17, $g_{\tilde{K}}^{-1}(p) = (t-1)p^{-1} + g_K^{-1}(p)$. Therefore,

$$g_K(p) < g_0(t-1, t; p) = \left(\max_{a' \in \{0,1,\dots,t\}} \left\{ \frac{a' - (t-1)}{p} + \frac{\ell_{a'} - 1}{1-p} \right\} \right)^{-1} \leq \frac{1-p}{\ell_{t-1} - 1}.$$

Again, we consider the graph F with vertex set $V(K)$ and edge set $\mathrm{EG}(K)$. By Lemma 40, every vertex in F has degree at least ℓ_t and every pair of vertices has at least one common neighbor. By Lemma 41, F has no cycle of length more than $\ell_{t-1} - 1$. If there is a maximum-length path that is a cycle, then Proposition 42 gives that F is Hamiltonian, which means $|V(K)| \leq \ell_{t-1} - 1$. As a result, $g_K(p) \geq \frac{1-p}{\ell_{t-1}-1}$, a contradiction.

So we may assume that every maximum-length path in F is not a cycle. Let $v_1 \dots v_\ell$ be such a maximum-length path such that, in K, the sum $\mathbf{x}(v_1) + \mathbf{x}(v_\ell)$ is the largest among such paths. Let v_c be the unique common neighbor of v_1 and v_ℓ.

Let v_1 have d neighbors in F. Since v_1 cannot have neighbors outside of this path, the sum of the weights, in K, of the neighbors of v_1 satisfy $d_G(v_1) \leq \mathbf{x}(v_2) + \cdots + \mathbf{x}(x_c)$. Notice that if $v_i \in \{v_1, \dots, v_{c-1}\}$ is a predecessor of a neighbor of v_1, then it is an endpoint of a path containing the same ℓ vertices, namely $v_i v_{i-1} \cdots v_1 v_{i+1} v_{i+2} \cdots v_c \cdots v_\ell$. Hence all d predecessors of gray neighbors of v_1 (including v_1 itself) have weight at most $\mathbf{x}(v_1)$. All other vertices have weight at most $\frac{g_K(p)}{1-p}$. Theorem 19 gives

$$\frac{p - g_K(p)}{p} + \frac{1-p}{p}\mathbf{x}(v_1) = \mathbf{x}(v_1) + d_G(v_1) \leq \mathbf{x}(v_1) + \cdots + \mathbf{x}(v_c) \leq d\mathbf{x}(v_1) + (c-d)\frac{g_K(p)}{1-p}.$$

Rearranging the terms, we obtain

$$g_K(p)\left(\frac{c-d}{1-p}+\frac{1}{p}\right) \geq 1-\mathbf{x}(v_1)\left(d-\frac{1-p}{p}\right).$$

Since $p^{-1} \leq p_0^{-1} = \ell_t$ and $\ell_t < d+1$, we may, by Lemma 40, lower bound the right-hand side by using $\mathbf{x}(v_1) \leq \frac{g_K(p)}{1-p}$ from Theorem 20,

$$\begin{aligned} g_K(p)\left(\frac{c-d}{1-p}+\frac{1}{p}\right) &\geq 1-\frac{g_K(p)}{1-p}\left(d-\frac{1-p}{p}\right) \\ g_K(p)\left(\frac{c}{1-p}\right) &\geq 1. \end{aligned}$$

Lemma 41 bounds the size of the longest cycle, so $c \leq \ell_{t-1}-1$. Thus, $g_K(p) \geq \frac{1-p}{c} \geq \frac{1-p}{\ell_{t-1}-1} \geq g_0(t-1,t;p)$, a contradiction. This concludes Case 4.

Case 5: $a = t-1$ and $p \in (0, p_0)$.

It remains to prove the theorem for $0 < p < p_0 = \ell_t^{-1}$ in the case where $(t+1) \not| \; h$ and $a = t-1$.

Fact 44. *Let h and t be positive integers such that $h \geq 2t+2$. Let $p_0 = \ell_t^{-1} = \left\lceil \frac{h}{2t+1} \right\rceil^{-1}$ and recall that*

$$\gamma_{\mathcal{H}}(p) = \min_{a \in \{0,\ldots,t\}} \left\{ \frac{p}{t+1}, \frac{p(1-p)}{a(1-p)+(\ell_a-1)p} \right\}.$$

Then $\gamma_{\mathcal{H}}(p) = p/(t+1)$ for $p \in [0, p_0]$.

We have $\mathrm{ed}_{\mathcal{H}}(p) \leq \gamma_{\mathcal{H}}(p)$ and the previous case gives that the two functions are equal at $p = p_0$. They are also equal at $p = 0$. By Fact 44, the function $\gamma_{\mathcal{H}}(p)$ is linear over $p \in [0, p_0]$ for $h \geq 2t+2$. Since $\mathrm{ed}_{\mathcal{H}}(p)$ is continuous and concave down, we may conclude that $\mathrm{ed}_{\mathcal{H}}(p) = \gamma_{\mathcal{H}}(p) = \frac{p}{t+1}$ for $p \in [0, p_0]$.

This concludes Case 5 and completes the proof of Theorem 30. □

2.4 Proofs of Lemmas and Facts

Corollary 32 gives that when h is large enough at most three functions are necessary to define the gamma function, in particular, the ones corresponding to $a = 0$ and $a = t$, and when

h is not divisible by $t + 1$ the function $\frac{p}{t+1}$.

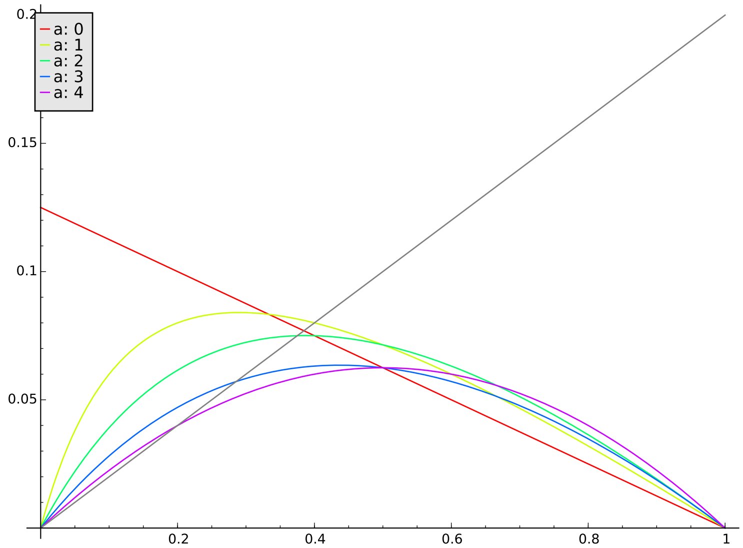

Figure 2.1 Functions that define $\gamma_{\mathrm{Forb}(C^4_{41})}$.

The general behavior of the functions with respect to a is interesting and is shown for C^4_{41} in Figure 2.1 above. The gray line corresponds to the function $\frac{p}{5}$, and this function together with the ones corresponding to $a = 0$ and $a = 4$ give the edit distance function and its maximum value, Figure 2.2 below.

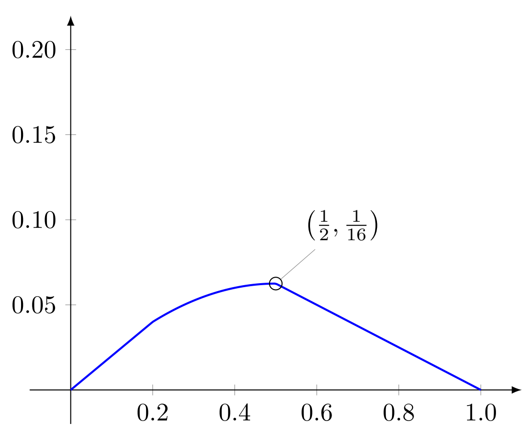

Figure 2.2 Plot of the edit distance function for $\mathrm{Forb}(C^4_{41})$.

Proof of Corollary 32. The case of $t = 1$ is covered by Corollary 31.

Let $a \in \{1, \dots, t-1\}$.

$$\text{If } p \geq \frac{a}{a + \ell_0 - \ell_a}, \qquad \text{then } \frac{p(1-p)}{a(1-p) + (\ell_a - 1)p} \geq \frac{1-p}{\ell_0 - 1}.$$
$$\text{If } p \leq \frac{t-a}{t - a + \ell_a - \ell_t}, \qquad \text{then } \frac{p(1-p)}{a(1-p) + (\ell_a - 1)p} \geq \frac{p(1-p)}{t(1-p) + (\ell_t - 1)p}.$$

Therefore, it suffices to show

$$\frac{t-a}{t - a + \ell_a - \ell_t} \geq \frac{a}{a + \ell_0 - \ell_a}$$
$$(\ell_0 - \ell_a)(t-a) \geq (\ell_a - \ell_t)a. \tag{2.3}$$

To that end,

$$\begin{aligned}
(\ell_0 - \ell_a)(t-a) - (\ell_a - \ell_t)a &= (t-a)\ell_0 + a\ell_t - t\ell_a \\
&> \frac{(t-a)h}{t+1} + \frac{ah}{2t+1} - \frac{th}{t+a+1} - t \\
&= \frac{at(t-a)h}{(t+1)(t+a+1)(2t+1)} - t \\
&\geq \frac{t(t-1)h}{(t+1)(2t)(2t+1)} - t.
\end{aligned}$$

If $h \geq 4t^2 + 10t + 12 + \frac{12}{t-1}$, then (2.3) is satisfied and the corollary follows. □

Proof of Fact 34. We only need to prove one direction because x and y are arbitrary. In both cases, we will prove the forward implication.

(a) Let $\lfloor h/x \rfloor \geq y$ and $h = qx + r$, where $r \in \{0, \dots, x-1\}$. Then $y \leq \lfloor h/x \rfloor = q$, so $h \geq xy + r$. Thus $\lfloor h/y \rfloor \geq x + \lfloor r/y \rfloor \geq x$.

(b) Let $\lceil h/x \rceil \leq y$ and $h = qx - r$, where $r \in \{0, \dots, x-1\}$. Then $y \geq \lceil h/x \rceil = q$, so $h \leq yx - r$. Thus $\lceil h/y \rceil \leq x - \lfloor r/y \rfloor \leq x$.

□

Proof of Fact 35. Clearly, if $a \in \{0, \dots, t-1\}$, then $\left\lceil \frac{h}{t+a+1} \right\rceil \leq \left\lceil \frac{h}{t+1} \right\rceil$ so it suffices to prove this fact for $a = 0$. Let $h = qt + r$ with $r \in \{0, \dots, t-1\}$. Since $h \geq t(t-1)$, we have $q \geq t - 1 \geq r$.

Then

$$\left\lceil \frac{h}{t+1} \right\rceil = q + \left\lceil \frac{r-q}{t+1} \right\rceil \leq q = \left\lfloor \frac{h}{t} \right\rfloor.$$

□

Proof of Fact 38. If $h \geq (t+1)^2 + 1$, then $t + 2 \leq \lceil h/(t+1) \rceil = \ell_0$. Consequently,

$$t + 1 \leq \frac{1}{2}(\ell_0 + t) \leq p(\ell_0 + t)$$

and so $\frac{1-p}{\ell_0 - 1} \leq \frac{p}{t+1}$.

For $a \in \{1, \ldots, t\}$, let $h = q(t+1)+r$, where $r \in \{1, \ldots, t+1\}$. The bound $h \geq (t+1)(t+a)+1$ ensures $q \geq t + a$. Then,

$$\begin{aligned} a + \left\lceil \frac{h}{t+a+1} \right\rceil &= a + \left\lceil \frac{q(t+a+1) + r - qa}{t+a+1} \right\rceil \\ &= q + \left\lceil \frac{a(t+a+1) + r - qa}{t+a+1} \right\rceil \\ &\leq q + \left\lceil \frac{a(t+a+1) + t + 1 - (t+a)a}{t+a+1} \right\rceil \\ &\leq q + 1 = \left\lceil \frac{h}{t+1} \right\rceil \end{aligned}$$

and so $\frac{1-p}{\ell_0 - 1} \leq \frac{p(1-p)}{a(1-p)+(\ell_a - 1)p}$. □

Proof of Lemma 40.

(a) Let $v \in V(K)$. Using Theorem 19,

$$\begin{aligned} \deg_G(v) &\geq \left\lceil \frac{d_G(v)}{\max\{\mathbf{x}(u)\}} \right\rceil \geq \left\lceil \frac{\frac{p - g_K(p)}{p} + \frac{1-2p}{p}\mathbf{x}(v)}{\frac{g_K(p)}{1-p}} \right\rceil \\ &\geq \frac{(p - g_K(p))(1-p)}{p g_K(p)} = \frac{1-p}{g_K(p)} - \frac{1-p}{p} \\ &> \max_{a' \in \{0,1,\ldots,t\}} \left\{ \frac{(a'-a)(1-p) + (\ell_{a'} - 1)\, p}{p} - \frac{1-p}{p} \right\} \\ &= \max_{a' \in \{0,1,\ldots,t\}} \left\{ \frac{(a'-a-1)(1-p)}{p} + \ell_{a'} - 1 \right\} \\ &\geq \ell_{a+1} - 1. \end{aligned}$$

The last inequality is obtained by choosing $a' = a + 1$.

(b) By inclusion-exclusion, $1 \geq d_G(v)+d_G(w)-d_G(v,w)$, we have that $d_G(v,w) \geq 2\frac{p-g_K(p)}{p} + \frac{1-2p}{p}(\mathbf{x}(v)+\mathbf{x}(w)) - 1 > \frac{p-2g_K(p)}{p}$. Therefore,

$$\begin{aligned}
\deg_G(v,w) &\geq \left\lceil \frac{d_G(v,w)}{\max\{\mathbf{x}(u)\}} \right\rceil \geq \left\lceil \frac{\frac{p-2g_K(p)}{p}}{\frac{g_K(p)}{1-p}} \right\rceil = \frac{1-p}{g_K(p)} - \frac{2(1-p)}{p} \\
&> \max_{a' \in \{0,1,\dots,t\}} \left\{ \frac{(a'-a)(1-p) + (\ell_{a'}-1)\,p}{p} - \frac{2(1-p)}{p} \right\} \\
&= \max_{a' \in \{0,1,\dots,t\}} \left\{ \frac{(a'-a-2)(1-p)}{p} + \ell_{a'} - 1 \right\}.
\end{aligned}$$

If $a \leq t-2$, then we choose $a' = a+2$. Then $\deg_G(v,w) > \ell_{a+2} - 1$, and because $\deg_G(v,w)$ is an integer, $\deg_G(v,w) \geq \ell_{a+2}$.

If $a = t-1$, then we choose $a' = t$. Then $\deg_G(v,w) > -\frac{1-p}{p} + \ell_t - 1 = \ell_t - p^{-1} \geq 0$, since $p \geq p_0 = \ell_t^{-1}$. Because $\deg_G(v,w)$ is an integer, $\deg_G(v,w) \geq 1$.

□

Proof of Lemma 41.

We say that a long cycle is a cycle of length at least $L+1$ and will show that there are no long cycles. Let $v_1 \cdots v_\ell$ be a smallest cycle in G among all those length greater than L.

Case 1: $0 \leq a \leq t-2$.

Observe that this case requires $t \geq 2$. Consider the path $v_1 \cdots v_{\ell_a - 1}$ on the cycle $v_1 \cdots v_\ell v_1$. There is no cycle of length ℓ_a and so the common neighbors of v_1 and v_{ℓ_a-1} are all in $\{v_2, \dots, v_{\ell_a-2}\}$. Note that Lemma 40 establishes that v_1 and v_{ℓ_a-1} have at least $\ell_{a+2} \geq 2$ common neighbors.

Since all common neighbors of v_1 and v_{ℓ_a-1} are in $\{v_2, \dots, v_{\ell_a-2}\}$, we have $\ell_a - 3 \geq \ell_{a+2}$. Hence,

$$\frac{h}{t+a+3} \leq \left\lceil \frac{h}{t+a+3} \right\rceil \leq \left\lceil \frac{h}{t+a+1} \right\rceil - 3 < \frac{h}{t+a+1} - 2$$

and so $h > (t+a+1)(t+a+3)$.

This gives that the number of common neighbors of v_1 and v_{ℓ_a-1} is at least $\ell_{a+2} = \left\lceil \frac{h}{t+a+3} \right\rceil \geq t+a+2 \geq 4$.

Therefore, v_1 and v_{ℓ_a-1} has at least two common neighbors in $\{v_3, \ldots, v_{\ell_a-3}\}$. Let $i > 2$ and $j < \ell_a - 2$ be, respectively, the smallest and largest indices of vertices in $\{v_3, \ldots, v_{\ell_a-3}\}$ that are common neighbors of v_1 and v_{ℓ_a-1}. That is, $3 \leq i \leq j \leq \ell_a - 3$. The cycle $v_1 v_i v_{i+1} \cdots v_{\ell-1} v_\ell$ has length $\ell - i + 2$. The cycle $v_1 v_2 \cdots v_{j-1} v_j v_{\ell_a-1} v_{\ell_a} \cdots v_{\ell-1} v_\ell$ has length $j + \ell - \ell_a + 2$.

Since these two cycles have length less than ℓ, they cannot be long cycles. Hence, their length is at most $\ell_a - 1$, giving us

$$\ell - i + 2 \leq \ell_a - 1$$
$$\ell + j - \ell_a + 2 \leq \ell_a - 1.$$

We can add these inequalities and use the fact that $\ell \geq L + 1$. Rearranging the terms, we conclude the following:

$$3\ell_a - 2L - 7 \geq 3\ell_a - 2\ell - 5 \geq j - i + 1 \geq \ell_{a+2} - 2. \tag{2.4}$$

To verify there are no long cycles, we must show that (2.4) produces a contradiction. Since $0 \leq a \leq t - 2$,

$$\begin{aligned} 3\ell_a - 2L - 7 &= 3\left\lceil \frac{h}{t+a+1} \right\rceil - 2\left\lfloor \frac{h}{t} \right\rfloor - 7 \\ &< 3\left(\frac{h}{t+a+1} + 1\right) - 2\left(\frac{h}{t} - 1\right) - 7 \\ &= \frac{h}{t+a+3} - 2 - \frac{2h(at + a^2 + 4a + 3)}{t(t+a+1)(t+a+3)} \\ &< \left\lceil \frac{h}{t+a+3} \right\rceil - 2 = \ell_{a+2} - 2, \end{aligned}$$

a contradiction for all $t \geq 2$ and $h \geq 2t + 2$. Therefore, for $0 \leq a \leq t - 2$, G has no cycle of length longer than $\ell_a - 1$.

Case 2: $a = t - 1$.

Since all common neighbors of v_1 and $v_{\ell_{t-1}-1}$ are in $\{v_2, \ldots, v_{\ell_{t-1}-2}\}$, we have $\ell_{t-1} - 3 \geq 1$.

Hence,

$$1 \leq \ell_{t-1} - 3 = \left\lceil \frac{h}{2t} \right\rceil - 3 < \frac{h}{2t} - 2$$

and so $h > 6t$, which means $\ell_{t-1} \geq 4$.

Consider the path $v_1 \cdots v_{\ell_{t-1}}$ on the cycle $v_1 \cdots v_\ell v_1$. To see there is no cycle of length $\ell_{t-1} + 1$, we set $h = q(2t) - r$ with $q \geq 2$ and $r \in \{0, \ldots, 2t-1\}$ and have

$$\ell_{t-1} + 1 = \left\lceil \frac{h}{2t} \right\rceil + 1 = q + 1 \leq 2q - 2 \leq 2q + \left\lfloor \frac{-r}{t} \right\rfloor \leq \left\lfloor \frac{h}{t} \right\rfloor = L.$$

Since v_1 and $v_{\ell_{t-1}}$ have a common neighbor v_i, either $v_1 v_i v_{i+1} \cdots v_\ell v_1$ or $v_1 \cdots v_i v_{\ell_{t-1}} v_{\ell_{t-1}+1} \cdots v_\ell v_1$ has length less than ℓ. Without loss of generality, we will assume that it is the former. This gives

$$\ell_{t-1} - 1 \geq \ell - i + 2 \geq \ell - (\ell_{t-1} - 1) + 2 \geq L + 1 - (\ell_{t-1} - 1) + 2.$$

Consequently,

$$2\ell_{t-1} - L - 5 \geq 0. \tag{2.5}$$

To see that (2.5) is contradicted,

$$\begin{aligned} 2\ell_{t-1} - L - 5 &= 2 \left\lceil \frac{h}{2t} \right\rceil - \left\lfloor \frac{h}{t} \right\rfloor - 5 \\ &< 2\left(\frac{h}{2t} + 1\right) - \left(\frac{h}{t} - 1\right) - 5 = -2 < 0. \end{aligned}$$

Therefore, for $a = t - 1$, G has no cycle of length longer than $\ell_{t-1} - 1$.

□

Proof of Proposition 42. Let $v_1 \cdots v_\ell$ be a longest path in G such that $v_1 v_\ell \in E(G)$. If G is not Hamiltonian, there exists a $w \in V(G) - \{v_1, \ldots, v_\ell\}$. Because G is connected, there exists $i \in \{1, \ldots, \ell\}$ and $w' \in V(G) - \{v_1, \ldots, v_\ell\}$ such that v_i is adjacent to w'. Then there is a longer path: $v_{i+1} \cdots v_\ell v_1 \cdots v_i w'$, a contradiction. □

Proof of Lemma 43.

Because $v_1 \cdots v_\ell$ is a longest path in F, neither v_1 nor v_k can have neighbors off this path, as that would yield a longer path. Thus $N(v_1) \cup N(v_\ell) \subseteq \{v_1, \ldots, v_k\}$ in F.

Case 1: $\ell \leq \ell_a$.

If v_i is adjacent to v_1, then v_{i-1} cannot be adjacent to v_ℓ. Thus, the predecessors of $N(v_1)$ and the neighbors of v_ℓ are disjoint subsets in $\{v_1, \ldots, v_{\ell-1}\}$. Since both v_1 and v_ℓ have degree at least ℓ_{a+1}, hence

$$2\ell_{a+1} \leq \ell - 1 \leq \ell_a - 1.$$

However,

$$\begin{aligned} \ell_a - 2\ell_{a+1} - 1 &= \left\lceil \frac{h}{t+a+1} \right\rceil - 2\left\lceil \frac{h}{t+a+2} \right\rceil - 1 \\ &< \frac{h}{t+a+1} - \frac{2h}{t+a+2} = -\frac{h(t+a)}{(t+a+1)(t+a+2)} < 0. \end{aligned} \tag{2.6}$$

Case 2: $\ell \geq \ell_a + 1$.

Partition the vertices of this path into $2s+1$ consecutive sets $A_0, B_1, A_1, \ldots, A_s, B_s$ with $s \geq 0$, constructed so that, in each set A_i, neighbors of v_1 appear before neighbors of v_ℓ as follows:

We let neighbors of v_1 be denoted with v_{p_i} and neighbors of v_ℓ be denoted with v_{q_i} in this construction. Let A_0 contain v_1 and add consecutive vertices of this path until we arrive at a neighbor of v_ℓ. From this point forward we do not allow another neighbor of v_1 to be in A_0, i.e. we continue adding consecutive vertices until we reach the last neighbor v_{q_0} of v_ℓ before another neighbor v_{p_1} of v_1. Then $A_0 = \{v_1, \ldots, v_{q_0}\}$, and we define $B_1 = \{v_{q_0+1}, \ldots, v_{p_1-1}\}$. Note that this definition does not preclude B_1 being an empty set. Continuing with this algorithm, we define sets $A_1 = \{v_{p_1}, \ldots, v_{q_1}\}$ and $B_2 = \{v_{q_1+1}, \ldots, v_{p_2-1}\}$, where v_{p_1} is a neighbor of v_1 on this path, v_{q_1} is the last neighbor of v_ℓ in A_1 before another neighbor v_{p_2} of v_1 as shown in Figure 2.3. We continue in this way and define sets $A_i = \{v_{p_i}, \ldots, v_{q_i}\}$ and $B_i = \{v_{q_{i-1}+1}, \ldots, v_{p_i-1}\}$ for $i \in \{1, \ldots, s\}$, adding the last vertex v_ℓ into the set A_s.

Now we analyze this partition:

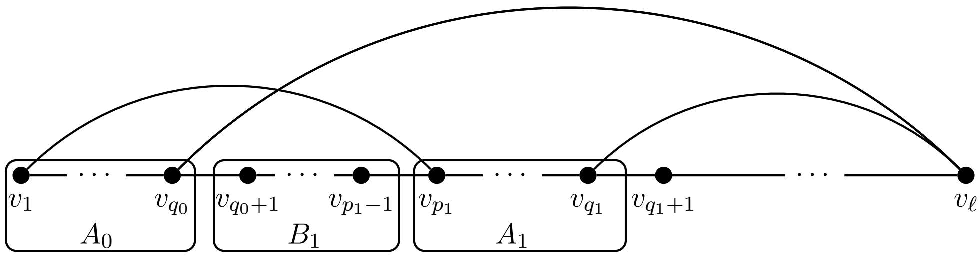

Figure 2.3 Partition of vertices of the path into sets A_i and B_i.

- We call the sets B_i, $i \in \{1, \dots, s\}$, *gaps* as they do not contain any neighbors of either v_1 or v_ℓ, but only contain vertices that succeed a given neighbor of v_ℓ and precede a given neighbor of v_1. According to the definition, gaps may be empty, but we will see below that this is not possible in this case.

- Each set A_i, $i \in \{0, \dots, s\}$, contains at most one common neighbor of v_1 and v_ℓ.

- By construction, neighbors of v_1 (other than a common neighbor, if it exists) precede neighbors of v_ℓ in each A_i, $i \in \{0, \dots, s\}$.

It will suffice to show that $s = 0$. This will imply that no neighbor of v_1 follows the first neighbor of v_ℓ on this path, which further implies that $N(v_1)$ entirely precedes $N(v_\ell)$, except possibly for a single common vertex. Since v_1 and v_ℓ have at least one common neighbor, the lemma will follow.

Notice that $v_1 \cdots v_{q_0} v_\ell v_{\ell-1} \cdots v_{p_1} v_1$ is a cycle as seen in Figure 2.3. In fact, for any $i \geq 1$, removing the gap B_i from vertices $\{v_1, \dots, v_\ell\}$ forms a cycle, so by assumption, $\ell - |B_i| \leq \ell_a - 1$ and none of the gaps can be empty. Therefore, $\sum_{i=1}^{s} |B_i| \geq s(\ell - \ell_a + 1)$.

On the other hand, by the degree assumption and since each set A_i contains at most one common neighbor of v_1 and v_ℓ, we obtain $2\ell_{a+1} \leq |N(v_1)| + |N(v_\ell)| \leq (\sum_{i=0}^{s} |A_i|) + (s+1) - 2$. Combining these two inequalities we have

$$\begin{aligned}\ell = \sum_{i=0}^{s} |A_i| + \sum_{i=1}^{s} |B_i| &\geq 2\ell_{a+1} - (s+1) + 2 + s(\ell - \ell_a + 1) \\ &= s(\ell - \ell_a) + 2\ell_{a+1} + 1.\end{aligned}$$

If $s \geq 1$, then we have $\ell \geq \ell - \ell_a + 2\ell_{a+1} + 1$ which simplifies to $\ell_a - 2\ell_{a+1} - 1 \geq 0$, which is contradicted by (2.6). Therefore $s = 0$ and the lemma follows. □

Proof of Fact 44. We need to show that $\gamma_{\mathcal{H}}(p_0) = p_0/(t+1)$. Since

$$\gamma_{\mathcal{H}}(p_0) = p_0 \cdot \min_{a \in \{0,\dots,t\}} \left\{ \frac{1}{t+1}, \frac{1-p_0}{a(1-p_0) + (\ell_a - 1)\, p_0} \right\},$$

we need to show that $\frac{\ell_a - 1}{\ell_t - 1} \leq t - a + 1$ for all $a \in \{0, \dots, t-1\}$.

To do this, let $h = q(2t+1) - r$ where $r \in \{0, \dots, 2t\}$ and $q \geq 2$ (because $h \geq 2t+2$). Then,

$$\begin{aligned}
\frac{\ell_a - 1}{\ell_t - 1} &= \frac{1}{q-1}\left(q - 1 + \left\lceil \frac{q(t-a) - r}{t+a+1} \right\rceil\right) \\
&\leq \frac{1}{q-1}\left(q - 1 + \frac{q(t-a) + t + a}{t+a+1}\right) \\
&= t - a + 1 + \frac{t^2 - a^2 + 2t - q(t^2 - a^2)}{(q-1)(t+a+1)},
\end{aligned}$$

which is at most $t - a + 1$ if $q \geq 3$ or if $a \leq t - 2$ and $q = 2$. In the case where $a = t - 1$ and $q = 2$, then $\frac{\ell_a - 1}{\ell_t - 1} = 1 + \left\lceil \frac{2-r}{2t} \right\rceil \leq 2 = t - a + 1$. □

CHAPTER 3. OTHER RESULTS

Generalizations of the edit distance problem to multicolorings of complete graphs and directed graphs were investigated by Axenovich and Martin in [4]. The motivation for work in this chapter is the possible extension of the theory to r-uniform hypergraphs. This chapter is based on joint work in progress with Ryan Martin [6]. Many definitions from Section 1.1.2 can be generalized, however many essential tools used for the graph case seem to be very difficult to generalize to hypergraphs. In particular, we need an extension of the result of Balogh and Martin which says that the edit distance function is the infimum of g functions, as well as extension of the results similar to the symmetrization theorem and the characterization of p-cores. We, therefore, begin with taking a closer look at the p-core CRGs as they are the key objects of the study.

3.1 Sidorenko's symmetrization technique

In this section we study the symmetric matrices that have a certain number of nonnegative values and use the results to give an alternative proof to the very useful tool of Marchant and Thomason [12] (Theorem 15) which gives structural characterization of p-core CRGs. Recall that a CRG K is p-core if $g_K(p) < g_{K'}(p)$ for all nontrivial sub-CRGs K' of K. We reprove Theorem 15 using Sidorenko's symmetrization method [24].

Let $\mathbf{M}$ be a nonnegative symmetric matrix of order at least two whose diagonal entries are positive. Define the quadratic program

$$g = g(\mathbf{M}) = \min\{\mathbf{x}^T\mathbf{M}\mathbf{x} : \mathbf{x}^T\mathbf{1} = 1, \mathbf{x} \geq \mathbf{0}\}. \tag{3.1}$$

We say that a symmetric matrix $\mathbf{M}$ is *core* if all optimal solutions to (3.1) have only nonzero entries. Recall that the symmetrization result from [13], Theorem 18, states that given a core matrix $\mathbf{M}$ and an optimal solution vector $\mathbf{x}^*$ to (3.1),

$$\mathbf{M} \cdot \mathbf{x}^* = g\,\mathbf{1}. \tag{3.2}$$

This theorem is the key tool in proving the following main result of this section.

Theorem 45. *Let $\mathbf{M} = (m_{ij})$ be a symmetric nonnegative core matrix, whose diagonal entries are positive.*

(i) Then $m_{ij} < \frac{m_{ii}+m_{jj}}{2}$ for all $i \neq j$.

(ii) If $m_{ij} \in \{a, b, c\}$, where a, b and c are nonnegative real numbers not all equal, then $m_{ij} < \min\{m_{ii}, m_{jj}\}$ for all $i \neq j$.

Note that the condition of the (at most) three nonnegative values in the statement of the second part of the theorem above is necessary. There are core matrices $\mathbf{M} = (m_{ij})$ with $m_{ij} \in \{a, b, c, d\}$, where a, b, c, d are distinct nonnegative real numbers, for which $m_{ij} \geq \min\{m_{ii}, m_{jj}\}$ for some $i \neq j$.

Example 46. Let $\mathbf{M} = (m_{ij})$ be the following matrix

$$\mathbf{M} = \begin{bmatrix} 3 & 1 & 0 & 0 \\ 1 & 3 & 1 & 0 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 2 & 3 \end{bmatrix}.$$

The matrix $\mathbf{M}$ is core since the solution to (3.1) is $\mathbf{x} = (1/3, 1/6, 1/3, 1/6)$. The entries of $\mathbf{M}$ have four distinct values $\{0, 1, 2, 3\}$ and $m_{43} = \min\{m_{33}, m_{44}\}$.

It seems that such example with a matrix of order three is rare (or doesn't exist), as well as an example of a matrix with four distinct values for which the strict inequality $m_{ij} > \min\{m_{ii}, m_{jj}\}$ holds. An example of a core matrix with five distinct values for which the strict inequality is achieved is the following

$$\mathbf{M} = \begin{bmatrix} 21 & 7 & 0 & 0 \\ 7 & 21 & 7 & 0 \\ 0 & 7 & 14 & 15 \\ 0 & 0 & 15 & 21 \end{bmatrix}.$$

We give the proof of Theorem 45 in the following subsection, and relate these results to the edit distance problem and derive Theorem 15 in Subsection 3.1.2.

3.1.1 Proof of Theorem 45

Proof (i) Without loss of generality, let $i < j$ and consider the principal submatrix $\begin{bmatrix} m_{ii} & m_{ij} \\ m_{ji} & m_{jj} \end{bmatrix}$ of $\mathbf{M}$. Define the vector $\mathbf{z} = (z_k)$ by

$$z_k = \begin{cases} -1, & \text{if } k = i, \\ 1, & \text{if } k = j, \\ 0, & \text{otherwise.} \end{cases}$$

Let $\mathbf{x}^*$ be an optimal solution to 3.1 and let x_i^* be the ith coordinate of $\mathbf{x}^*$. Choose $\varepsilon > 0$ such that $x_i^* + \varepsilon z_i = 0$. Then the vector $\mathbf{x}^* + \varepsilon \mathbf{z}$ has only nonnegative entries with exactly one zero entry and $(\mathbf{x}^* + \varepsilon \mathbf{z})^T \mathbf{1} = 1$. Furthermore, using (3.2)

$$\begin{aligned} (\mathbf{x}^* + \varepsilon \mathbf{z})^T \mathbf{M} (\mathbf{x}^* + \varepsilon \mathbf{z}) &= (\mathbf{x}^*)^T \mathbf{M} \mathbf{x}^* + 2\varepsilon \mathbf{z}^T \mathbf{M} \mathbf{x}^* + \varepsilon^2 \mathbf{z}^T \mathbf{M} \mathbf{z} \\ &= g + 2\varepsilon \mathbf{z}^T (g\mathbf{1}) + \varepsilon^2 \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} m_{ii} & m_{ij} \\ m_{ji} & m_{jj} \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\ &= g + \varepsilon^2 (m_{ii} + m_{jj} - 2m_{ij}). \end{aligned}$$

Now if $m_{ij} = \frac{m_{ii}+m_{jj}}{2}$, then from above computation $(\mathbf{x}^* + \varepsilon \mathbf{z})^T \mathbf{M} (\mathbf{x}^* + \varepsilon \mathbf{z}) = g$. Hence $\mathbf{x}^* + \varepsilon \mathbf{z}$ is an optimal solution to (3.1) which has a zero entry, contradicting to the fact that $\mathbf{M}$ is core.

On the other hand, if $m_{ij} > \frac{m_{ii}+m_{jj}}{2}$, then $(\mathbf{x}^* + \varepsilon\mathbf{z})^T \mathbf{M} (\mathbf{x}^* + \varepsilon\mathbf{z}) < g$, contradicting optimality of $\mathbf{x}^*$. Therefore, $m_{ij} < \frac{m_{ii}+m_{jj}}{2}$ for all $i \neq j$.

Proof (ii) For a symmetric matrix $\mathbf{M}$ with at most three nonnegative values (not all equal) in $\{a, b, c\}$ with $a \geq b \geq c$, define the matrix $\mathbf{M}' = \frac{1}{a+b-2c}(\mathbf{M} - c\mathbf{J})$ and let $p = \frac{b-c}{a+b-2c}$, where $\mathbf{J}$ is the all-ones matrix. Then the matrix $\mathbf{M}'$ is symmetric with at most three nonnegative values in $\{0, p, 1-p\}$, where $p \in [0, 1]$. Furthermore, $\mathbf{x}^*$ is an optimal solution to (3.1) if and only if it is an optimal solution to $\min\{\mathbf{x}^T\mathbf{M}'\mathbf{x} : \mathbf{x}^T\mathbf{1} = 1, \mathbf{x} \geq \mathbf{0}\}$. Therefore, without loss of generality, we may assume that the matrices we work with in this section have entries in $\{0, p, 1-p\}$ with $p \in [0, 1]$.

Let $\mathbf{M}$ be a core matrix with values in $\{0, p, 1-p\}$, where $p \in [0, 1]$. Notice that the diagonal entries of $\mathbf{M}$ can not be zero, otherwise the optimal solution to (3.1) will have a zero entry contradicting to the fact that $\mathbf{M}$ is core.

Let $p \leq \frac{1}{2}$. Since $\frac{m_{ii}+m_{jj}}{2} \in \{p, 1-p, \frac{1}{2}\}$, we have that $m_{ij} \neq 1-p$ for all $j > 1$ by Part *(i)*. Suppose $m_{11} = p$; the same argument works when $m_{ii} = p$. It is enough to show that $m_{1j} = 0$ for all $j > 1$. By way of contradiction, assume that there is an integer $l \geq 1$ such that

$$m_{1j} = \begin{cases} p, & \text{for } j = 1, \ldots, l \\ 0, & \text{for } j \geq l+1. \end{cases}$$

Let $\mathbf{x}^* = (x_i^*)$ be an optimal solution to (3.1) and define the vector $\mathbf{z} = (z_i)$ by

$$z_i = \begin{cases} -\frac{g}{p} + x_1^*, & \text{if } i = 1, \\ x_i^*, & \text{if } i = 2, \ldots, l, \\ 0, & \text{if } i \geq l+1. \end{cases}$$

Since $\mathbf{M}\mathbf{x}^* = g\mathbf{1}$, from the first row of the product we obtain

$$px_1^* + \cdots + px_l^* = g. \tag{3.3}$$

Now choose $\varepsilon > 0$ such that $\mathbf{x}^* + \varepsilon\mathbf{z} \geq 0$, i.e. so that $x_1^* + \varepsilon\left(-\frac{g}{p} + x_1^*\right) \geq 0$. Notice that using (3.3), $(\mathbf{x}^* + \varepsilon\mathbf{z})^T \cdot \mathbf{1} = 1 + \varepsilon\left(-\frac{g}{p} + x_1^* + \cdots + x_l^*\right) = 1$, so $\mathbf{x}^* + \varepsilon\mathbf{z}$ is a feasible solution to (3.1).

Furthermore, using (3.2)

$$\begin{aligned}(\mathbf{x}^* + \varepsilon\mathbf{z})^T\mathbf{M}(\mathbf{x}^* + \varepsilon\mathbf{z}) &= (\mathbf{x}^*)^T\mathbf{M}\mathbf{x}^* + 2\varepsilon\mathbf{z}^T\mathbf{M}\mathbf{x}^* + \varepsilon^2\mathbf{z}^T\mathbf{M}\mathbf{z} \\ &= g + 2\varepsilon g(\mathbf{z}^T \cdot \mathbf{1}) + \varepsilon^2\mathbf{z}^T\mathbf{M}\mathbf{z} \\ &= g + \varepsilon^2\mathbf{z}^T\mathbf{M}\mathbf{z}.\end{aligned}$$

Notice that by definition of $\mathbf{z}$ we have $\mathbf{z}^T\mathbf{M}\mathbf{z} \leq 0$. Now if $\mathbf{z}^T\mathbf{M}\mathbf{z} < 0$, then $(\mathbf{x}^*+\varepsilon\mathbf{z})^T\mathbf{M}(\mathbf{x}^*+\varepsilon\mathbf{z}) = g+\varepsilon^2\mathbf{z}^T\mathbf{M}\mathbf{z} < g$, contradicting optimality of g. If $\mathbf{z}^T\mathbf{M}\mathbf{z} = 0$, then $(\mathbf{x}^*+\varepsilon\mathbf{z})^T\mathbf{M}(\mathbf{x}^*+\varepsilon\mathbf{z}) = g + \varepsilon^2\mathbf{z}^T\mathbf{M}\mathbf{z} = g$, so $\mathbf{x}^* + \varepsilon\mathbf{z}$ is an optimal solution with a zero entry, contradicting to the fact that $\mathbf{M}$ is core. Therefore, it must be the case that $m_{1j} = 0$ for all $j > 1$.

The proof of the case when $p \geq \frac{1}{2}$ follows the same lines of the case above.

□

3.1.2 From matrices back to CRGs

Let $p \in [0, 1]$ and let K be a CRG with vertex set $V(K) = \{v_1, \ldots, v_k\}$ with corresponding matrix $\mathbf{M}_K(p)$. Notice that K is p-core if and only if the matrix $\mathbf{M}_K(p)$ is core. Then from Theorem45 observe that:

- If $p = \frac{1}{2}$, all off diagonal entries of $\mathbf{M}_K(p)$ are equal to 0, which means that all of the edges of K are gray.
- If $p < \frac{1}{2}$, then
  - $m_{ij} \neq 1 - p$ for all i, j with $i \neq j$, which means K has no black edges, and
  - if $m_{ii} = m_{jj} = p$ then $m_{ij} \neq p$, so white vertices of K cannot be incident to white edges.
- If $p > \frac{1}{2}$, then
  - $m_{ij} \neq p$ for all i, j with $i \neq j$, which means K has no white edges, and

– if $m_{ii} = m_{jj} = 1 - p$ then $m_{ij} \neq 1 - p$, so black vertices of K cannot be incident to black edges.

From this observation above, we can see that it agrees with the characterization of the CRGs presented in Theorem 15.

The CRGs for the multicolor edit distance problem studied in [4] have corresponding core matrices with more than three nonnegative values. Although there is a weak characterization of p-core CRGs for those matrices (Theorem 45, part *(i)*), the strong one does not hold (Theorem 45, part *(ii)*). It would be interesting to know if there is a similar structural characterization for those p-core CRGs.

3.2 Hypergraphs and other discussions

A *hypergraph* $\mathcal{G} \subseteq 2^V$ with a vertex set V is a collection of subsets from V. A hypergraph $\mathcal{G}$ is said to be an *r-uniform hypergraph* (*r-graph* for short) if every subset belonging to $\mathcal{G}$ has cardinality r.

The *edit distance between two r-graphs* on the same labeled vertex set is their symmetric difference, and this can be extended similarly to the graph case as follows.

Definition 47. *Let $\mathcal{G}$ be an r-graph and let $\mathcal{H}$ be a hereditary property of r-uniform hypergraphs. Then*

(i) $\mathrm{dist}(\mathcal{G}, \mathcal{H}) = \min\{\mathrm{dist}(\mathcal{G}, \mathcal{G}') : \mathcal{G}' \in \mathcal{H}, V(\mathcal{G}) = V(\mathcal{G}')\}$ *is the edit distance from $\mathcal{G}$ to $\mathcal{H}$.*

(ii) $\mathrm{dist}(n, \mathcal{H}) = \max\{\mathrm{dist}(\mathcal{G}, \mathcal{H}) : |V(\mathcal{G})| = n\}$ *is the maximum edit distance from the set of all n-vertex r-graphs to $\mathcal{H}$.*

The notion of colored regularity graphs can also be extended to hypergraphs as follows.

Definition 48 ([15])**.** *A colored regularity hypergraph of order r (r-CRH) is a pair (V, ϕ), where $V = \{v_1, \ldots, v_k\}$ and $\phi : V^r \longrightarrow \{W, G, B\}$ (corresponding to colors white, gray, and black, respectively) such that*

- $\phi(v, \ldots, v) \in \{W, B\}$ *for all $v \in V$ and*

- $\phi(v_1, \ldots, v_r) = \phi(v_{\sigma(1)}, \ldots, v_{\sigma(r)})$ *for every permutation* $\sigma \in S_r$ *and for every* $v_i \in V$*,* $i = 1, \ldots, r$*.*

Notice that a 2-CRH is a CRG. Analogously, we say that an r-graph $\mathcal{G}$ embeds in an r-CRH K, denoted $\mathcal{G} \to K$, if (hyper)edges of $\mathcal{G}$ are mapped to either black vertices of K or to the union of black and gray (hyper)edges of K, and nonedges of $\mathcal{G}$ are mapped to either white vertices of K or to union of white and gray (hyper)edges of K. We denote $\mathcal{K}(\mathcal{H})$ to be the set of r-CRHs such that no forbidden r-graph is mapped to K.

There is a natural multilinear form $\mathbf{M} = \mathbf{M}_K(p) = (m_{i_1,\ldots,i_r})$ associated with an r-CRH K on k vertices, where $m_{i_1,\ldots,i_r}$ is p, $1-p$ or 0 if $\phi(v_{i_1}, \ldots, v_{i_r})$ is W, B, or G, respectively. Let $R = \{\mathbf{x} \in \mathbb{R}^k : \mathbf{x} \geq \mathbf{0}, \mathbf{x}^T\mathbf{1} = 1\}$, then the g_K function can be defined similar to the graph case as follows

$$g_K(p) = \min\left\{\sum_{i_1,\ldots,i_r} m_{i_1,\ldots,i_r} x_{i_1} \cdots x_{i_r} : (x_1, \ldots, x_k) \in R\right\}.$$

Recall that Balogh and Martin in [6] showed that $\mathrm{ed}_{\mathcal{H}}(p) = \inf_{K \in \mathcal{K}(\mathcal{H})} g_K(p)$. In order to generalize the theory we need a similar result for hypergraphs. By choosing an r-CRH K from $\mathcal{K}(\mathcal{H})$ and editing an r-graph $\mathcal{G}$ according to K we can show that

$$\mathrm{ed}_{\mathcal{H}}(p) \leq \inf_{K \in \mathcal{K}(\mathcal{H})} g_K(p). \tag{3.4}$$

However, proving the lower bound is difficult, and possibly requires the hypergraph version of the Szemerédi's Regularity Lemma [22, 23].

The edit distance problem generalizes the Turán problem, asymptotically. Given a graph H and integer n, the *extremal function* $\mathrm{ex}(n, H)$ is defined by

$$\mathrm{ex}(n, H) = \max\{|E(G)| : G \text{ is } H\text{-free}\}.$$

The *Turán density* $\pi(H)$ of a graph H is defined by

$$\pi(H) = \lim_{n\to\infty} \mathrm{ex}(n, H)/\binom{n}{2}.$$

The result of Balogh and Martin mentioned above gives us

$$\pi(H) = 1 - \inf_{K: H' \not\to K, \forall H' \supset H} g_K. \tag{3.5}$$

This result can then be used to prove the Erdős-Stone Theorem.

Theorem 49. *Let H be a simple graph, and denote its chromatic number by $\chi = \chi(H)$. Then*

$$\pi(H) = 1 - \frac{1}{\chi - 1}.$$

The lower bound is achieved by the Turán graph $T(n, \chi - 1)$. The upper bound comes directly from (3.4), (3.5) and applying the edit distance function with $p = 1$.

CHAPTER 4. CONCLUSION AND OPEN QUESTIONS

We have obtained the edit distance function over all of its domain for C_h^t when $t+1$ does not divide h and $h \geq 2t(t+1)+1$. When $t+1$ divides h and $h \geq 2t(t+1)+1$, we have obtained the function for $p \in [p_0, 1]$, where $p_0 = \left\lceil \frac{h}{2t+1} \right\rceil^{-1}$. The function, however, is not known when $t+1$ divides h and $p \in [0, p_0)$ or when $h \leq 2t(t+1)$.

As to the case of $p < p_0$ (and h sufficiently large), we showed that if $K \in \mathcal{K}(\mathrm{Forb}(C_h^t))$ is a p-core CRG with $p < 1/2$ which has $a \neq t-1$ white vertices, then $g_K(p) = \gamma_{\mathrm{Forb}(C_h^t)}(p)$. Therefore, to solve the problem for the remaining case when $t+1$ divides h, and p is small, one only needs to consider CRGs with exactly $t-1$ white vertices. A particular barrier to this is Lemma 40 which requires $p \geq p_0$ to ensure that the graph induced by the black vertices and gray edges of the CRG has the property that any two vertices have at least one common neighbor. Such a condition need not hold for small p.

As to reducing the lower bound required of h, we note that in the proof of Theorem 30, we required $h \geq 2t(t+1)+1$ in Fact 38. This ensured that the $\gamma_{\mathcal{H}}$ function for $p \in [1/2, 1]$ is linear and by the concavity and continuity of the edit distance function, this ensures that $\mathrm{ed}_{\mathcal{H}}(p) = \gamma_{\mathcal{H}}(p)$ in that interval. So, more careful analysis of the case $p \geq 1/2$ may enable one to reduce the lower bound on h, but these arguments are very different from the case where $p < 1/2$. Elsewhere, we only require $h \geq \max\{t(t-1), 2t+2\}$ in order to complete the proof of Theorem 30. This bound is required in several places. See Fact 35, Lemma 36, Lemma 41 but especially the basic Fact 37 which says that a set of size h can be partitioned into sets of size t or $t+1$ if and only if $h \geq t(t-1)$. So we believe that it would be difficult to prove the theorem for values of h smaller than $\max\{t(t-1), 2t+2\}$ in general.

The edit distance problem for hypergraphs remains wide open as discussed in Section 3.2. In particular, is it possible to compute the edit distance function $\mathrm{ed}_{\mathcal{H}}$ from the g_K functions of $K \in \mathcal{K}(\mathcal{H})$ for all $p \in [0,1]$? The upper bound for $\mathrm{ed}_{\mathcal{H}}$ can be obtained using r-CRHs, so the main concern is the lower bound. Balogh and Martin [6] proved the lower bound for the graph case using the Regularity Lemma. Marchant and Thomason [12] gave a nice structural characterization of p-core CRGs. Is there a similiar characterization for p-core r-CRHs? Many other open problems on edit distance are given in the survey by Martin [15].

BIBLIOGRAPHY

[1] Alon, N. and Stav, A. (2008). What is the furthest graph from a hereditary property? *Random Structures Algorithms*, 33(1): 87–104.

[2] Alon, N. and Stav, A. (2008). The maximum edit distance from hereditary graph properties. *J. Combin. Theory Ser. B*, 98(4): 672–697.

[3] Axenovich, M., Kézdy, A., and Martin, R. (2008). On the editing distance of graphs. *J. Graph Theory*, 58(2): 123–138.

[4] Axenovich, M. and Martin, R. (2011). Multicolor and directed edit distance. *J. Combin.*, 2(4): 525–556.

[6] Balogh, J. and Martin, R. (2008). Edit distance and its computation. *Electronic J. Combin.*, 15(1): Research Paper 20, 27.

[6] Berikkyzy, Z. and Martin, R. On Sidorenko's Symmetrization Technique. In preparation.

[7] Berikkyzy, Z., Martin, R., and Peck, C. On the edit distance of powers of cycles. Submitted.

[8] Bollobás, B. (1998). Hereditary properties of graphs: asymptotic enumeration, global structure, and colouring. Proceedings of the international Congress of Mathematicians. *Doc. Math.*, Extra Vol. III: 333–342 (electronic).

[9] Bollobás, B. and Thomason, A. (1997). Hereditary and monotone properties of graphs. *The mathematics of Paul Erdős II*, volume 14 of *Algorithms Combin.*, pages 70-78. Springer, Berlin.

[10] Dirac, G.A. (1952). Some theorems on abstract graphs. *Proc. London Math. Soc.* 3(2): 69–81.

[11] Erdős, P. and Rényi, A. (1959). On random graphs, I. *Publicationes Mathematicae (Debrecen)*, 6: 290–297.

[12] Marchant, E. and Thomason, A. (2010). Extremal graphs and multigraphs with two weighted colours. *Fete of combinatorics and computer science*, volume 20 of *Bolyai Soc. Math. Stud.*: pages 239–286. János Bolyai Math. Soc., Budapest.

[13] Martin, R. (2013). The edit distance function and symmetrization. *Electronic J. Combin.*, 20(3): Research Paper 26, 25.

[14] Martin, R. (2015). On the computation of edit distance functions. *Discrete Mathematics*, 338(2): 291–305.

[15] Martin, R. (2016). The edit distance in graphs: methods, results and generalizations. *Recent Trends in Combinatorics*, 31–62, IMA Vol. Math. Appl., 159, Springer, Cham.

[16] Martin, R. and McKay, T. (2014). On the edit distance from $K_{2,t}$-free graphs. *J. Graph Theory*, 77(2): 117–143.

[17] Ore, O. (1960). Note on Hamilton circuits. *Amer. Math. Monthly* 67(1): 55.

[18] Peck, C. (2013). On the edit distance from a cycle- and squared cycle-free graph. Master's thesis, Iowa State University.

[19] Pikhurko, O. (2008). An exact Turán result for the generalized triangle. *Combinatorica*, 28(2): 187-208.

[20] Prömel, H.J. and Steger, A. (1992). Excluding induced subgraphs. III A general asymptotic. *Random Structures Algorithms*, 3(1): 19–31.

[21] Prowse, A. and Woodall, D. R. (2003). Choosability of powers of circuits. *Graphs and Combinatorics*, 19(1): 137-144.

[22] Rödl, V. and Skokan, J. (2004). Regularity lemma for k-uniform hypergraphs. *Random Structures Algorithms*, 25(1): 1–42.

[23] Rödl, V. and Skokan, J. (2006). Applications of the regularity lemma for uniform hypergraphs. *Random Structures Algorithms*, 28(2): 180–194.

[24] Sidorenko, A. (1993). Boundedness of optimal matrices in extremal multigraph and digraph problems. *Combinatorica*, 13(1): 109-120.

[25] Sylvester, J. J. (1884). Question 7382. *Mathematical Questions from the Educational Times*, 41: 21.

[26] Thomason, A. (2011). Graphs, colours, weights and hereditary properties. In Surveys in combinatorics, *LMS Lecture Note Series* 392: 333–364.

[27] West, D. B. (2001). *Introduction to Graph Theory.* Prentice Hall, 2nd edition.