ON THE EDIT DISTANCE FROM $K_{2,t}$-FREE GRAPHS I: CASES $t = 3, 4$

RYAN MARTIN AND TRACY MCKAY

Abstract. The edit distance between two graphs on the same vertex set is defined to be size of the symmetric difference of their edge sets. The edit distance function of a hereditary property, $\mathcal{H}$, is a function of p, and measures, asymptotically, the furthest graph of edge density p from $\mathcal{H}$ under this metric.

The edit distance function has proven to be difficult to compute for most hereditary properties. A number of surprising connections to classical extremal graph problems, such as the problem of Zarankiewicz, have been uncovered in attempts to compute various edit distance functions. In this paper, we address the hereditary property $\mathrm{Forb}(K_{2,t})$, the property of having no induced copy of the complete bipartite graph with 2 vertices in one class and t in the other, in the cases $t = 3$ and $t = 4$. We are able to, with a variety of techniques, determine the edit distance function over the entire domain $p \in [0, 1]$.

1. Introduction

The study of edit distance in graphs initially appeared in papers by Axenovich, Kézdy and the first author [2] and, independently, by Alon and Stav [1]. It has several potential applications, as described in [1], especially to property testing problems in theoretical computer science. More recently, interest has been shown in determining the value of the *edit distance function*, introduced in [3] by Balogh and the first author. Strategies for determining this function appear in [7], by Marchant and Thomason, as well as in [8].

Let $G(n, p)$ denote the Erdős-Rényi random graph on n vertices with edge probability p. Given a hereditary property (that is, a set of graphs closed under vertex deletion and isomorphism) how many edge additions or deletions are necessary to make $G(n, p)$ a member of the property? What is the behavior of this value as $n \to \infty$? In [2], the binary chromatic number of a graph H is used as a means of bounding this value when the hereditary property can be defined as all graphs that forbid H as an induced subgraph. These methods give an exact result in some cases, most notably when H is self-complementary.

In [1], a version of Szemerédi's regularity lemma is applied to show that as $n \to \infty$, the number of edge changes necessary to make $G(n, p)$ a member of a given hereditary property approaches the maximum possible number over all n-vertex graphs within $o(n^2)$ so long as p is chosen correctly with respect to the given hereditary property, $\mathcal{H}$. The edit distance function, $ed_{\mathcal{H}}(p)$, from [3], describes the expected normalized edit distance of $G(n, p)$ as $n \to \infty$ for all probabilities p. Not surprisingly, the maximum value of this function occurs at the same p value described in [1].

In this paper, we explore what can be said about the edit distance function for the hereditary property $\mathrm{Forb}(K_{2,t})$, the set of all graphs that do not contain a complete bipartite graph with cocliques of 2 and t vertices as an induced subgraph. In particular, we determine the entire edit distance functions for the hereditary properties of forbidding induced $K_{2,3}$ and $K_{2,4}$ subgraphs, denoted $\mathrm{Forb}(K_{2,3})$ and $\mathrm{Forb}(K_{2,4})$, respectively.

2010 *Mathematics Subject Classification.* Primary 05C35; Secondary 05C80.

Key words and phrases. edit distance, quadratic programming, strongly regular graphs.

The first author's research partially supported by NSF grant DMS-0901008 and by an Iowa State University Faculty Professional Development grant.

The second author's research partially supported from NSF grant DMS-0901008.

Prior results and notation come primarily from [8], as well as previous work: [1], [2], [3], [7]; however, there are a number of other excellent resources on related topics. For a more extensive review of this literature, the reader may wish to consult Thomason [10]. We now introduce some important definitions and state our results more rigorously.

1.1. **Definitions.**

We begin by recalling some important definitions relating to edit distance.

Definition 1 (Alon-Stav [1]; Axenovich-Kézdy-RM [2])**.** *Let G and H be simple graphs on the same labeled vertex set, and let $\mathcal{H}$ be a hereditary property, then*

1. $\mathit{Dist}(G, H) = |E(G)\Delta E(H)|$ *is the edit distance from G to H,*
2. $\mathit{Dist}(G, \mathcal{H}) = \min\{\mathit{Dist}(G, H) : H \in \mathcal{H}\}$ *is the edit distance from G to $\mathcal{H}$ and*
3. $\mathit{Dist}(n, \mathcal{H}) = \max\{\mathit{Dist}(G, \mathcal{H}) : |G| = n\}$ *is the maximum edit distance from the set of all n-vertex graphs to the hereditary property $\mathcal{H}$.*

Since $\mathcal{H}$ is by definition closed under isomorphism, vertex labels may be ignored when considering $\mathrm{Dist}(G, \mathcal{H})$. In fact, $\mathrm{Dist}(G, \mathcal{H})$ could be defined equivalently as the minimum number of edge changes necessary to make G a member of $\mathcal{H}$.

The limit of the maximum edit distance from an n-vertex graph to a hereditary property $\mathcal{H}$ normalized by the total number of potential edges in an n-vertex graph,

$$d^*_{\mathcal{H}} = \lim_{n\to\infty} \mathrm{Dist}(n, \mathcal{H})/\binom{n}{2},$$

is demonstrated in [1] to exist and to be realized asymptotically with high probability by the random graph $G(n, p^*)$, where $p^* \in [0, 1]$ is a probability that depends on $\mathcal{H}$ and is not necessarily unique.

Definition 2 ([3])**.** *The* ***edit distance function*** *of a hereditary property $\mathcal{H}$ is defined as follows:*

$$\mathit{ed}_{\mathcal{H}}(p) = \lim_{n\to\infty} \max\left\{\mathit{Dist}(G, \mathcal{H}) : |V(G)| = n, |E(G)| = \lfloor p\tbinom{n}{2}\rfloor\right\}/\tbinom{n}{2}.$$

This function has also been denoted as $g_{\mathcal{H}}(p)$ in, for example, [3]. The limit above was proven to exist in [3], and furthermore, $\mathit{ed}_{\mathcal{H}}(p)$ is both **continuous** and **concave down**. As a result, the edit distance function attains a maximum value that is equal to $d^*_{\mathcal{H}}$. The point, or interval, at which $d^*_{\mathcal{H}}$ is attained is denoted $p^*_{\mathcal{H}}$, and when it is evident from context, the subscript $\mathcal{H}$ may be omitted from both.

1.2. **New Results.**

In this paper, we prove the following results for the hereditary properties $\mathrm{Forb}(K_{2,3})$ and $\mathrm{Forb}(K_{2,4})$. The case of $K_{2,2}$ is mentioned in Section 5.3 of [7].

Theorem 3. *Let $\mathcal{H} = \mathit{Forb}(K_{2,3})$. Then $\mathit{ed}_{\mathcal{H}}(p) = \min\{p(1-p), \frac{1-p}{2}\}$ with $p^*_{\mathcal{H}} = \frac{1}{2}$ and $d^*_{\mathcal{H}} = \frac{1}{4}$.*

Theorem 4. *Let $\mathcal{H} = \mathit{Forb}(K_{2,4})$. Then $\mathit{ed}_{\mathcal{H}}(p) = \min\{p(1-p), \frac{7p+1}{15}, \frac{1-p}{3}\}$ with $p^*_{\mathcal{H}} = \frac{1}{3}$ and $d^*_{\mathcal{H}} = \frac{2}{9}$.*

It should be noted that $p^*_{\mathcal{H}}$ and $d^*_{\mathcal{H}}$ from Theorem 3 could be found using alternative methods from previous literature as well. In fact, they are a direct result of Lemma 5.14 in [7], as is the value of $\mathit{ed}_{\mathcal{H}}(p)$ for $p \geq 1/2$ in both theorems. The $p^*_{\mathcal{H}}$ and $d^*_{\mathcal{H}}$ values in Theorem 4, however, are not so easily found. The techniques used to prove both theorems also have the potential to yield some results for $\mathit{ed}_{\mathcal{H}}(p)$ when $\mathcal{H} = \mathrm{Forb}(K_{2,t})$ and $t \geq 5$, as discussed in [9].

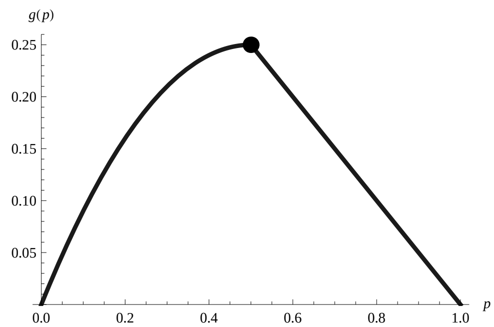

FIGURE 1. Plot of $ed_{\mathrm{Forb}(K_{2,3})}(p) = \min\{p(1-p),(1-p)/2\}$. The point $(p^*,d^*) = (1/2,1/4)$ is indicated.

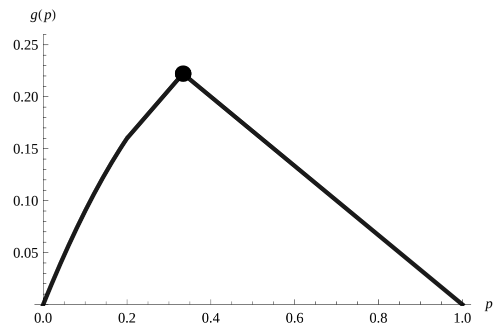

FIGURE 2. Plot of $ed_{\mathrm{Forb}(K_{2,4})}(p) = \min\{p(1-p),(1+7p)/15,(1-p)/3\}$. The point $(p^*,d^*) = (1/3,2/9)$ is indicated.

1.3. **CRGs and Background.**

To help describe how $ed_{\mathcal{H}}(p)$ may be calculated, some definitions from [1] are required.

Definition 5 (Alon-Stav [1]). *A **colored regularity graph (CRG)**, K, is a complete graph with vertices colored black or white, and with edges colored black, white or gray.*

At times, it may be convenient to refer to the graph induced by edges of a particular color in a CRG, K. We shall refer to these graphs as the black, white and gray subgraphs of K. The investigation of edit distance in [7] and in [10] uses a different paradigm with an analogous structure called a *type*. Essentially, our black, white and gray are their red, blue and green, respectively.

Definition 6 (Alon-Stav [1]). *A **colored homomorphism** from a (simple) graph H to a colored regularity graph K is a mapping $\phi : V(H) \mapsto V(K)$, which satisfies the following:*

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

Basically, a colored homomorphism is a map from a simple graph to a CRG so that black is only associated with adjacency, white is only associated with nonadjacency and gray is associated with adjacency, nonadjacency or both. We will refer to a colored homomorphism from a simple graph H to a CRG K as *an embedding of H in K*, and we denote the set of all CRGs that only allow the embedding of simple graphs in a hereditary property $\mathcal{H}$ as $\mathcal{K}(\mathcal{H})$ or merely $\mathcal{K}$ when $\mathcal{H}$ is clear from the context. Since any hereditary property may be described by a set of forbidden induced subgraphs, an equivalent description of $\mathcal{K}(\mathcal{H})$ is the set of all CRGs that do not permit the embedding of any of the forbidden induced subgraphs associated with $\mathcal{H}$. For instance, $\mathcal{K}(\mathrm{Forb}(K_{2,3}))$ is the set of all CRGs that do not admit $K_{2,3}$ embedding.

In order to calculate $ed_{\mathcal{H}}(p)$, colored regularity graphs are used in [3] in order to define the following functions:

$$
\begin{aligned}
f_K(p) &= \frac{1}{k^2}\left[p(|\mathrm{VW}(K)| + 2|\mathrm{EW}(K)|) + (1-p)(|\mathrm{VB}(K)| + 2|\mathrm{EB}(K)|)\right] \\
g_K(p) &= \min\{\mathbf{u}^T M_K(p)\mathbf{u} : \mathbf{u}^T\mathbf{1} = 1 \text{ and } \mathbf{u} \geq 0\}.
\end{aligned}
$$

Here K is a CRG with k vertices. $\mathrm{VW}(K)$, $\mathrm{VB}(K)$, $\mathrm{EW}(K)$ and $\mathrm{EB}(K)$ represent the sets of white vertices, black vertices, white edges and black edges in K respectively. M_K is essentially a weighted adjacency matrix for K with black vertices and edges receiving weight $1-p$, white vertices and edges receiving weight p and gray edges receiving weight 0. From [1], it is known that $ed_{\mathcal{H}}(p) = \inf_{K\in\mathcal{K}}\{f_K(p)\} = \inf_{K\in\mathcal{K}}\{g_K(p)\}$. Moreover, Alon and Stav [1] show that if χ_B is the binary chromatic number of $\mathcal{H}$, then $ed_{\mathcal{H}}(1/2) = 1/(\chi_B - 1)$. Marchant and Thomason, demonstrate in [7] that $ed_{\mathcal{H}}(p) = \min_{K\in\mathcal{K}}\{g_K(p)\}$. That is, given p there exists at least one CRG, $K \in \mathcal{K}$, such that $ed_{\mathcal{H}}(p) = g_K(p)$.

If we say a CRG, K, is a *sub-CRG* of another CRG, K', when $\mathrm{VW}(K) \subseteq \mathrm{VW}(K')$, $\mathrm{VB}(K) \subseteq \mathrm{VB}(K')$, $\mathrm{EW}(K) \subseteq \mathrm{EW}(K')$ and $\mathrm{EB}(K) \subseteq \mathrm{EB}(K')$, then it may be observed that $g_K(p) \geq g_{K'}(p)$. Furthermore, as was noted in [7], if $g_K(p) = g_{K'}(p)$, then there is no need to consider both K and K' when attempting to determine $\min_{K\in\mathcal{K}}\{g_K(p)\}$. Thus, a special subset of CRGs is defined as follows.

Definition 7 (Marchant-Thomason [7]). *A* ***p-core CRG*** *is a CRG K' such that for no nontrivial sub-CRG K of K' is it the case that $g_K(p) = g_{K'}(p)$. In other words, if K' is a p-core CRG, and K is a nontrivial sub-CRG of K', then $g_K(p) > g_{K'}(p)$.*

It can be shown (see [7]) that a CRG, K, is p-core if and only if $g_K(p) = \mathbf{x}^T M_K(p)\mathbf{x}$ for a unique vector $\mathbf{x}$ with positive entries summing to 1. Any CRG, K, that is not p-core contains at least one p-core sub-CRG K' so that $g_{K'}(p) = g_K(p)$. Thus we could also say that

$$ed_{\mathcal{H}}(p) = \min\{g_K(p) : K \in \mathcal{K} \text{ and } K \text{ is } p\text{-core}\}.$$

That is, when looking for CRGs to determine $ed_{\mathcal{H}}(p)$, the search may be limited to the important subset of CRGs, p-cores. This observation is especially helpful for determining lower bounds for the edit distance function.

To prove the main results in this paper, we first show that for each p there exists a CRG, $K \in \mathcal{K}$, so that either $f_K(p)$ or $g_K(p)$ is equal to the function value in the theorem, giving an upper bound for the value of $ed_{\mathcal{H}}(p)$. Then, employing a method called "localization" in [8], features of the graphs $K_{2,t}$, and the concavity of the edit distance function, we demonstrate that for no p-core CRG, $K \in \mathcal{K}$, can $g_K(p)$ be less than the value in the theorem, justifying our conclusions.

By the continuity of the edit distance function, if we know the value of the function on an open interval, then we also know the value on its closure. Hence, for convenience, most of our proofs will only address the value of the function on the interior of a given interval. We also note that, in [8], whenever the edit distance function of a hereditary property is computed, there is an attempt to determine all of the p-core CRGs that achieve the value of the edit distance function. In this paper we only concern ourselves with the value of the edit distance function itself and do not address the issue of multiple defining constructions, though careful study of the proof techniques should enable others to classify all of the relevant CRGs.

1.4. **The Zarankiewicz problem and strongly regular graphs.**

One reason for our interest in the edit distance function for $\mathrm{Forb}(K_{2,t})$ is its relation to the Zarankiewicz problem. This problem addresses the question of how many edges a graph can have before it must contain a $K_{s,t}$ subgraph for fixed s and t. In an intriguing result from [7], a construction from Brown [5] for $K_{3,3}$-free graphs is applied to construct an infinite set of new CRGs that improve upon the previously known bounds for $ed_{\mathrm{Forb}(K_{3,3})}(p)$ on certain intervals for arbitrarily small p.

Marchant and Thomason [7] establish that it is sufficient to consider only p-core CRGs for which the gray subgraph has neither a K_3 nor a $K_{3,3}$. Brown's constructions are not K_3-free, but a bipartite graph can be created from the construction that has no copy of $K_{3,3}$. Similarly, for $K_{2,t}$, it is sufficient to have no gray $K_{2,t}$ or B_{t-2} to forbid $K_{2,t}$ embedding. The graph B_{t-2} is a "book" as defined in [6]. We define books precisely in Lemma 12.

Although the Brown constructions show that the edit distance function for Forb($K_{3,3}$) is strictly less than $p(1-p)/(1+p)$ for sufficiently small p, known constructions for dense $K_{2,t}$-free graphs do not play a role in the computation of the edit distance function for Forb($K_{2,3}$) or Forb($K_{2,4}$) in the same way. However, the edit distance function for Forb($K_{2,4}$) is achieved over the interval $[1/5, 1/3]$ by a construction formed from a strongly regular graph, namely a generalized quadrangle, often denoted GQ(2, 2). This graph is a so-called $(15, 6, 1, 3)$-strongly regular graph.

1.5. **Organization.**

In Section 2, we discuss some results from [7] and [8] for the edit distance function and how they may be applied to the problem of determining the function for Forb($K_{2,t}$). We then proceed to some general results and observations in Section 3 that will be useful throughout the remainder of the paper. Sections 4 and 5 contain the proofs of our results for Forb($K_{2,3}$) and Forb($K_{2,4}$), respectively. The final sections are reserved for conclusions and acknowledgements.

2. Applications of past results to Forb($K_{2,t}$)

If a CRG is p-core, one can say some interesting things about its overall structure. From [7], for instance, we have the following useful result.

Theorem 8 (Marchant-Thomason [7])**.** *If K is a p-core CRG, then all edges of K are gray except*

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

The CRGs with all edges gray are useful in bounding the edit distance function, as we see in [8].

Definition 9. *Let $K(w, b)$ denote the CRG with w white vertices, b black vertices and only gray edges. In particular:*

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

Theorem 10 settled the case of $K_{2,2}$, and Theorem 11 permits us to focus on black-vertex CRGs.

Theorem 10 (Marchant-Thomason [7])**.** *Let $\mathcal{H}$ =Forb($K_{2,2}$). Then $ed_{\mathcal{H}}(p) = g_{K(1,1)}(p) = p(1-p)$ with $p^*_{\mathcal{H}} = \frac{1}{2}$ and $d^*_{\mathcal{H}} = \frac{1}{4}$.*

Theorem 11 (Marchant-Thomason [7])**.** *Let $\mathcal{H}$ = Forb($K_{2,t}$), $t > 2$. Then*

(1) *For $p > \frac{1}{2}$, $ed_{\mathcal{H}}(p) = g_{K(0,t-1)}(p) = \frac{1-p}{t-1}$ and*
(2) *For $p \le \frac{1}{2}$, either*
    - *$ed_{\mathcal{H}}(p) = \min\{g_{K(1,1)}(p), g_{K(0,t-1)}(p)\}$, or*
    - *$ed_{\mathcal{H}}(p) = g_K(p) < \min\{g_{K(1,1)}(p), g_{K(0,t-1)}(p)\}$, where K is a p-core CRG with only black vertices and, consequently, no black edges.*

The following lemma is about the structure of the p-core CRGs described in the second part of Theorem 11. It was originally observed in Example 5.16 of [7]. The proof is a straightforward case analysis

Lemma 12 (Marchant-Thomason [7])**.** *A CRG, K, on all black vertices with only white and gray edges forbids $K_{2,t}$ embedding if and only if its gray subgraph contains no $K_{2,t}$ or B_{t-2} as a subgraph, where B_{t-2} is a book as described in* [6]. *That is, the graph B_{t-2} is defined to be the graph consisting of $t-2$ triangles that all share a single common edge.*

As demonstrated in [7], for a p-core CRG, K, there is a *unique* vector $\mathbf{x}$ so that $g_K(p) = \mathbf{x}^T M_K(p)\mathbf{x}$.

Definition 13 (Marchant-Thomason [7])**.** *For a p-core CRG K with optimal weight vector $\mathbf{x}$, the entry of $\mathbf{x}$ corresponding to a vertex, $v \in V(K)$, is denoted by $\mathbf{x}(v)$. This is the **weight of** v, and the function $\mathbf{x}(v)$ is the **optimal weight function**.*

With this in mind, we have two propositions from [8], which follow easily from [7].

Proposition 14 ([8])**.** *Let K be a p-core CRG with all vertices black. Then for any $v \in V(K)$ and optimal weighting $\mathbf{x}$, $d_G(v) = \frac{p-g_K(p)}{p} + \frac{1-2p}{p}\mathbf{x}(v)$, where $d_G(v)$ is the sum of the weights of the vertices adjacent to v via a gray edge.*

Proposition 15 ([8])**.** *Let K be a p-core CRG with all vertices black, then for $p \in [0, 1/2]$ and optimal weighting $\mathbf{x}$,*

$$\mathbf{x}(v) \leq \frac{g_K(p)}{1-p}, \ \forall v \in V(K).$$

Because of Theorem 11, we may restrict our attention to those CRGs, K, for which $g_K(p) \leq p(1-p)$. As a result, Proposition 14 gives the lower bound $d_G(v) \geq p + \frac{1-2p}{p}\mathbf{x}(v)$. Meanwhile, Proposition 15 restricts the optimal weights of all vertices in K to be no more than p. These two restrictions are useful when attempting to prove lower bounds for $ed_{\mathrm{Forb}(K_{2,t})}(p)$.

3. Preliminary results and observations

We begin with some notation used throughout the paper for convenience.

Definition 16. *Let K be a black-vertex, p-core CRG with $g_K(p) \leq p(1-p)$ and optimal weight function $\mathbf{x}$:*

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

Partitioning the vertices in a black-vertex, p-core CRG that forbids a $K_{2,t}$ embedding into the three sets $\{u_0\}$, U and W as seen in Figure 3, illustrates some interesting features of its optimal weight function when the gray neighborhoods of these vertices are examined. One such feature is the upper bounds in Proposition 17 for x_1.

Proposition 17. *Let $K \in [\mathcal{K}(Forb(K_{2,3})) \cup \mathcal{K}(Forb(K_{2,4}))]$ be a black-vertex, p-core CRG. If either $p < 1/3$ or both $p < 1/2$ and the gray sub-CRG of K is triangle-free, then*

$$x_1 \leq x \qquad \textit{and} \qquad x_1 \leq p - x$$

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

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

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

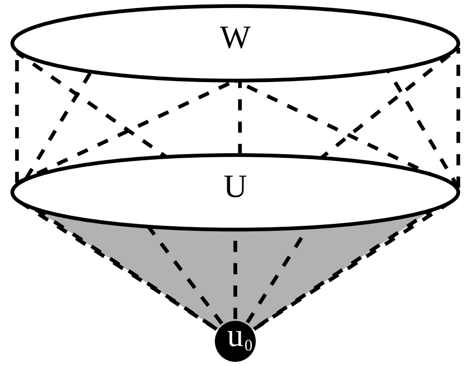

FIGURE 3. A partition of the vertices in a black-vertex, p-core CRG, K. Dashed lines and gray background represent gray edges. White edges are omitted, as are edges within subsets.

Recall that u_1 is a vertex with maximum weight in the gray neighborhood of u_0, a vertex with maximum weight in all of K and assume that $x + x_1 > p$. Then applying Proposition 14 and Theorem 11,

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

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

This completes the proof of Proposition 17 for $K \in \mathcal{K}(\mathrm{Forb}(K_{2,3}))$ since, in this case, no $K \in \mathcal{K}$ contains a gray triangle. So we may assume that $K \in \mathcal{K}(\mathrm{Forb}(K_{2,4}))$.

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

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

Applying similar reasoning to that in Case 1,

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

So,

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

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

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

Fact 18. *In a CRG, if u_0 is a vertex with maximum weight, $x = \mathbf{x}(u_0)$, the maximum weight in the gray neighborhood of u_0 is x_1 and the order of the gray neighborhood of u_0 is ℓ, then $\ell \geq d_G(u_0)/x_1 \geq d_G(u_0)/x$.*

While simple, when combined with Propositions 14 and 17 along with the observation that $\mathbf{x}(u_0) + \mathbf{x}(U) + \mathbf{x}(W) = 1$, this fact forces a delicate balance between the weights of the vertex u_0, the vertices in U and the vertices in W.

4. Proof of Theorem 3

In this section, we establish the value of $ed_{\mathrm{Forb}(K_{2,3})}(p)$ for $p \in (0, 1/2)$, determining the entire function through continuity and Theorem 11, which gives that $ed_{\mathrm{Forb}(K_{2,3})}(p) = (1-p)/2$ for $p \in [1/2, 1]$.

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

Lemma 19. *Let K be a black-vertex, p-core CRG with $p \in (0, 1/2)$, no gray triangles, no gray $K_{2,3}$ and $g_K(p) \leq p(1-p)$. If u_0 is a vertex of maximum weight, x, in K, and $\ell = |N_G(u_0)|$, then*

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

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

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

Using Proposition 14 and the fact that $\mathbf{x}(W) = 1 - x - d_G(u_0)$,

$$\begin{aligned} \sum_{i=1}^{\ell} \left(p - x + \frac{1-2p}{p} \mathbf{x}(u_i) \right) &\leq 2\mathbf{x}(W) \\ \ell(p-x) + \frac{1-2p}{p} d_G(u_0) &\leq 2\left(1 - x - d_G(u_0)\right) \\ \ell(p-x) &\leq 2 - 2x - \frac{1}{p} d_G(u_0) \\ \ell &\leq \frac{2(1-x) - \frac{1}{p} d_G(u_0)}{p-x}. \end{aligned}$$

□

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

Lemma 20. *Let K be a black-vertex, p-core CRG for $p \in (0, 1/2)$ with no gray triangles, no gray $K_{2,3}$ and $g_K(p) \leq p(1-p)$. If x and x_1 are defined as in Proposition 17, then*

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

Proof. By Fact 18, $\ell \geq \frac{d_G(u_0)}{x_1}$, and by Lemma 19, $\ell \leq \frac{2(1-x)-\frac{1}{p}d_G(u_0)}{p-x}$. Therefore,

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

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

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

and then applying Proposition 14,

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

□

We now turn to the proof of the main theorem for this section.

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

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

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

We start with the inequality from Lemma 20,

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

and apply the bound $x_1 \leq p-x$ from Proposition 17 to get

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

From Proposition 15, $p-x > 0$, and so

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

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

Case 2: $x < p/2$.

We again apply Lemma 20, only now we employ the trivial bound $x_1 \leq x$ from Proposition 17:

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

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

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

If $p < (6-2\sqrt{2})/7$, then neither x' nor x'' is real. For $p \in \left[\frac{6-2\sqrt{2}}{7}, \frac{1}{2}\right)$, routine calculations show that $\frac{p}{2} < x'$, a contradiction to the assumption that $x < p/2$.

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

5. Proof of Theorem 4

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

5.1. **Upper bounds.**

Recall that from Theorem 11 we already know that $ed_{\mathrm{Forb}(K_{2,4})}(p) \leq \min\{p(1-p), \frac{1-p}{3}\}$.

For the remaining upper bound, we turn to strongly regular graphs:

Definition 21. *A (k,d,λ,μ)-**strongly regular graph** is a graph on k vertices such that each vertex has degree d, each pair of adjacent vertices has exactly λ common neighbors and each pair of nonadjacent vertices has exactly μ common neighbors.*

Lemma 22. *Let $\mathcal{H} = Forb(K_{2,t})$. If there exists a (k,d,λ,μ)-strongly regular graph with $\lambda \leq t-3$ and $\mu \leq t-1$, then*

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

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

If no pair of adjacent vertices has $t-2$ common neighbors, then there is no book B_{t-2} in the gray subgraph. If no pair of vertices has t common neighbors, then there is no $K_{2,t}$ in the gray subgraph. By Lemma 12, $K_{2,t} \not\mapsto K$. Furthermore,

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

$\square$

In fact, there is a $(15,6,1,3)$-strongly regular graph [4]. It is a so-called "generalized quadrangle," $\mathrm{GQ}(2,2)$. As a result,

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

A list of known strongly regular graphs and their parameters has been compiled by Andries Brouwer [4]. Their implications for $ed_{\mathrm{Forb}(K_{2,t})}(p)$ in general are explored further in [9].

5.2. **Lower bounds.**

Because the edit distance function is both continuous and concave down, it is sufficient to verify that $ed_{\mathrm{Forb}(K_{2,4})}(p) \geq p(1-p)$ for $p \in (0,1/5)$ and that $ed_{\mathrm{Forb}(K_{2,4})}(p) \geq (1-p)/3$ for $p \in (1/3,1/2)$. This is because the line determined by the bound $\frac{1+7p}{15}$ passes through the points $(1/5,4/25)$ and $(1/3,2/9)$. Furthermore, by Theorem 11, we need only consider CRGs that have black vertices and white and gray edges.

Lemmas 23 and 26 settle the cases where $p \in (1/3,1/2)$ and where $p \in (0,1/5)$, respectively.

Lemma 23. *Let $p \in (1/3,1/2)$. If K is a black-vertex, p-core CRG that does not contain a gray book B_2 or a gray $K_{2,4}$, then $g_K(p) \geq \frac{1-p}{3}$, with equality occurring only if K is a gray triangle (i.e., $K \approx K(0,3)$).*

Proof. We break this into two cases, as to whether or not K has a gray triangle.

Case 1: K has a gray triangle.

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

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

Then, applying Proposition 14,

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

and so

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

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

Case 2: K has no gray triangle.

Let u_0 be a vertex of largest weight, $x = \mathbf{x}(u_0)$, and let $U = N_G(u_0)$. The absence of a gray triangle means that there are no gray edges between pairs of vertices in U. Furthermore, no vertex in W can be adjacent to more than three vertices in U via a gray edge, since by Lemma 12, the gray subgraph of K does not contain a $K_{2,4}$.

Let $u_1, \ldots, u_\ell$ be an enumeration of the vertices in U with weights $x_1, \ldots, x_\ell$, respectively, and $g = g_K(p)$. Then

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

and applying Proposition 14 to compute $d_G(u_i)$,

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

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

First, suppose $\ell \geq 5$. Then, from inequality (1), we have

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

and applying Proposition 14 again,

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

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

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

If $4p^2 + p - 1 \geq 0$, then we use Proposition 15 and substitute $x = g/(1-p)$,

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

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

Second, suppose $\ell \leq 2$. Then by Fact 18 we have $\ell \geq \mathbf{x}(U)/x$, yielding

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

and so bounding $\mathbf{x}(U)$ using Proposition 14,

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

Using Proposition 15, $x \le g/(1-p)$ yields

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

and so if $\ell \le 2$, then $g \ge (1-p)/3$, with equality if and only if $x = g/(1-p)$, and consequently, K is a gray triangle. So, we may further assume that $\ell \in \{3, 4\}$.

Third, suppose $\ell = 3$.

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

Returning to inequality (1), we have

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

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

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

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

Returning to inequality (1), we have

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

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

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

If $3p^2+p-1 \geq 0$, then we use Fact 18 to bound $x \leq g/(1-p)$,

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

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

This ends Case 2 and the proof of the lemma. □

Before we prove Lemma 26, there are two propositions that are necessary and used in several cases.

Proposition 24. *Let $p \in (0, 1/2)$, and let K be a black-vertex, p-core CRG with no gray book B_2 and no gray $K_{2,4}$. If $g = g_K(p)$, $U = N_G(u_0)$, $\ell = |U|$ and $U_1 \subseteq U$ is the set of vertices in U that are incident to a gray edge in U, then*

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

Proof. Let $u_1, \ldots, u_\ell$ be an enumeration of the vertices of U. Then

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

and applying Proposition 14,

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

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

Proposition 25. *Let $p \in (0, 1/2)$, and let K be a black-vertex, p-core CRG with no gray book B_2 and no gray $K_{2,4}$. If $g_K(p) \leq p(1-p)$, then both*

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

Proof. We begin with Proposition 24 and then use $\ell \geq \mathbf{x}(U)/x$ from Fact 18:

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

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

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

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

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

Some routine but tedious calculations demonstrate that, for $p \in \left[(9-4\sqrt{3})/11, 1/2\right)$, this expression is at least $1/25$, achieving that value uniquely at $p=1/5$. $\square$

Lemma 26. *Let $p \in (0,1/5)$. If K is a black-vertex, p-core CRG that does not contain a gray book B_2 or a gray $K_{2,4}$, then $g_K(p) > p(1-p)$.*

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

Case 1: $\ell \geq 8$.

According to Proposition 24,

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

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

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

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

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

Using Fact 18, and then Proposition 14

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

a contradiction.

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

First we bound ℓ:

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

So, $\ell = 7$. Since ℓ is odd, $\mathbf{x}(U_1) \leq 6x$. By Lemma 24,

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

and applying Proposition 14,

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

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

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

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

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

and applying Proposition 14,

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

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

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

If $\ell_1 = |U_1|$, then $\mathbf{x}(U_1) \leq (\ell_1/2)(p-x)$.

We can also count the number of vertices in $U-U_1$ by using the fact that $(\ell-\ell_1)x \geq \mathbf{x}(U)-\mathbf{x}(U_1)$. Returning to Proposition 24,

$$\begin{aligned}
\ell\left(\frac{p-g}{p}-x\right) &\leq 3-3x-\frac{1+p}{p}\mathbf{x}(U)+\mathbf{x}(U_1) \\
\left[\ell_1+\frac{1}{x}\mathbf{x}(U)-\frac{1}{x}\mathbf{x}(U_1)\right]\left(\frac{p-g}{p}-x\right) &\leq 3-3x-\frac{1+p}{p}\mathbf{x}(U)+\mathbf{x}(U_1) \\
\mathbf{x}(U)\left(\frac{p-g}{px}-1+\frac{1+p}{p}\right)-3+3x &\leq \mathbf{x}(U_1)\left(1+\frac{p-g}{px}-1\right)-\ell_1\left(\frac{p-g}{p}-x\right) \\
\left[\frac{p-g}{p}+\frac{1-2p}{p}x\right]\left(\frac{p-g}{px}+\frac{1}{p}\right)-3+3x &\leq \frac{\ell_1}{2}(p-x)\left(\frac{p-g}{px}\right)-\ell_1\left(\frac{p-g}{p}-x\right) \\
\left[\frac{p-g}{p}+\frac{1-2p}{p}x\right]\left(\frac{p-g}{px}+\frac{1}{p}\right)-3+3x &\leq \ell_1\left[x-\frac{p-g}{p}\cdot\frac{3x-p}{2x}\right] \\
\left[p+\frac{1-2p}{p}x\right]\left(\frac{p}{x}+\frac{1}{p}\right)-3+3x &\leq \ell_1\left[x-\frac{p(3x-p)}{2x}\right] \\
p^2-(1+2p)x+\frac{1-2p+3p^2}{p^2}x^2 &\leq \ell_1\frac{(p-x)(p-2x)}{2}.
\end{aligned}$$

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

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

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

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

By Proposition 25, we may restrict our attention to $p \geq (9-4\sqrt{3})/11 > 1/7$ and so we may substitute the smallest possible value for x, which still maintains the inequality.

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

a contradiction.

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

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

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

This, too, is a contradiction for $p \in (0, 1/5)$, completing the proof of Lemma 26. $\square$

6. Conclusion

- While there are many calculations in the proofs, the underlying idea is this: $x + \mathbf{x}(W)$ must be large enough to accommodate $d_G(U)$. However, if x is too large, $d_G(U)$ is too large (as a result of Proposition 14) to be accommodated by the weights of the vertices not in U. Meanwhile, if x is too small, then $N_G(U)$ has too many vertices: by Fact 18, $\ell \geq \frac{d_G(U)}{x}$, and by Proposition 14, the gray neighborhoods of those vertices must be large and mostly in W.
- Although we determine all of $ed_{\mathrm{Forb}(K_{2,4})}(p)$, convexity allows $d^*_{\mathrm{Forb}(K_{2,4})}$ to be determined with only Lemma 23. Furthermore, the generalized quadrangle $\mathrm{GQ}(2,2)$ was unnecessary to compute this quantity, since $p(1-p) < 2/9$ for $p \in [0, 1/3)$, and $p(1-p)$ is an upper bound for every function $ed_{\mathrm{Forb}(K_{2,t})}(p)$, $t \geq 2$.
- Proposition 25 gives a nontrivial lower bound for a black-vertex, p-core CRG, K, that forbids a $K_{2,4}$ embedding. If we take inequality (2) and solve for g, then we see that, if $p \in (0, 1/2)$, then

$$g_K(p) \geq \frac{2p + 6 - 6\sqrt{1 - 3p + 5p^2}}{11} = \frac{2p(3-4p)}{p + 3 + 3\sqrt{1 - 3p + 5p^2}},$$

  which is strictly larger than $p(1-p)$ for $p \in (0, (9 - 4\sqrt{3})/11)$. So, in particular, there is a positive gap between the $g_K(p)$ functions for black-vertex CRGs and the CRG with one white and one black vertex.
- Ed Marchant reports having also proven that $ed_{\mathrm{Forb}(K_{2,3})}(p) = \min\{p(1-p), (1-p)/2\}$, using different methods.
- Analysis of $ed_{\mathrm{Forb}(K_{2,t})}(p)$ for $t \geq 5$ is continued in [9].

7. Acknowledgements

Special thanks to Sung-Yell Song for his knowledge and expertise on strongly regular graphs, Maria Axenovich, Ed Marchant and Andrew Thomason for helpful conversations.

References

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

[2] Maria Axenovich, André Kézdy, and Ryan Martin. On the editing distance of graphs. *J. Graph Theory*, 58(2):123–138, 2008.

[3] József Balogh and Ryan Martin. Edit distance and its computation. *Electron. J. Combin.*, 15(1):Research Paper 20, 27, 2008.

[4] Andries E. Brouwer. Parameters of strongly regular graphs. `http://www.win.tue.nl/~aeb/graphs/srg/srgtab.html` (accessed July 1, 2010).

[5] W. G. Brown. On graphs that do not contain a Thomsen graph. *Canad. Math. Bull.*, 9:281–285, 1966.

[6] Guantao Chen, Ralph J. Faudree, and Ronald J. Gould. Saturation numbers of books. *Electron. J. Combin.*, 15(1):Research Paper 118, 12, 2008.

[7] Edward Marchant and Andrew Thomason. Extremal graphs and multigraphs with two weighted colours. In G.O.H. Katona, A. Schrijver, and T. Szönyi, editors, *Fete of Combinatorics and Computer Science*, volume 20 of *Bolyai Soc. Math. Stud.* Springer, Berlin, 2010. In press.

[8] Ryan Martin. The edit distance function and localization. `arXiv:1007.1897v3`.

[9] Ryan Martin and Tracy McKay. On the edit distance from $K_{2,t}$-free graphs II: Cases $t \geq 5$. Preprint.

[10] Andrew Thomason. Graphs, colours, weights and hereditary properties. Preprint.

Department of Mathematics, Iowa State University, Ames, Iowa 50011
E-mail address: `rymartin@iastate.edu`

Department of Mathematics, Iowa State University, Ames, Iowa 50011
E-mail address: `tmckay16@iastate.edu`