ON THE EDIT DISTANCE FROM $K_{2,t}$ -FREE GRAPHS II: CASES $t \ge 5$

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 with edge density p from \mathcal{H} under this metric.

The edit distance function has proven to be difficult to compute for many hereditary properties. Some surprising connections to extremal graph theory problems, such as strongly regular graphs and the problem of Zarankiewicz, have been uncovered in attempts to compute various edit distance functions. In this paper, we address the hereditary property $\operatorname{Forb}(K_{2,t})$ when $t \geq 5$, the property of having no induced copy of the complete bipartite graph with 2 vertices in one class and t in the other.

This work continues from a prior paper by the authors. Employing an assortment of techniques and colored regularity graph constructions, we are able to extend the interval over which the edit distance function for this hereditary property is generally known and determine its maximum value for all odd t. We also explore several constructions to improve upon known upper bounds for the function.

1. INTRODUCTION

Motivated by applications, including property testing applications in computer science, the problem of determining graph edit distances was developed independently in papers by Alon and Stav [1] and Axenovich, Kézdy and the first author [2].

A hereditary property is a set of graphs closed under vertex deletion and isomorphism, and the edit distance from a graph G to a hereditary property \mathcal{H} , denoted by $\text{Dist}(G, \mathcal{H})$, is the minimum number of edge additions or deletions necessary to make G a member of \mathcal{H} .

The edit distance function for a hereditary property \mathcal{H} , which is denoted $ed_{\mathcal{H}}(p)$, is presented in [3] by Balogh and the first author. It models with high probability the limiting behavior as $n \to \infty$ of the edit distance from an Erdős-Rényi random graph on n vertices with edge probability $p \in [0, 1]$ to a hereditary property normalized by $\binom{n}{2}$.

Marchant and Thomason explore the value of $1 - ed_{\mathcal{H}}(p)$ for various hereditary properties in [10], developing some insightful results for determining the value of the function in general. One discovery from [10] of particular interest is a relationship between the problem of determining the edit distance function for Forb($K_{3,3}$) and constructions by Brown in [6] for $K_{3,3}$ -free graphs, related to the Zarankiewicz problem.

Meanwhile in [12], the precursor to this paper and to which we refer the reader for further background on the problem, constructions that are used to address the Zarankiewicz problem proved to be immaterial to the edit distance function for $\operatorname{Forb}(K_{2,t})$ when t = 3 or 4. In this paper, however, we show that not only does their relevance reemerge as a potential player in the problem of determining the value of the edit distance function for $\operatorname{Forb}(K_{2,t})$ when t is sufficiently large, but that exploring the relationship between $ed_{\operatorname{Forb}(K_{2,t})}(p)$ and colored regularity graph (CRG) constructions

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.

involving strongly regular graphs and cycles also gives insight into the general behavior of these functions.

As described in [3], there are two different ways that colored regularity graphs may be used to define the edit distance function for a given hereditary property \mathcal{H} .

The first method involves the function

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

where VW(K), VB(K), EW(K) and EB(K) are the sets of white vertices, black vertices, white edges and black edges in K respectively. The CRG, K, does not permit embedding of any of the forbidden induced subgraphs associated with \mathcal{H} -we refer to the set of all such K as $\mathcal{K}(\mathcal{H})$, or just \mathcal{K} , when \mathcal{H} is evident.

The second method involves the function

(2)
$$g_K(p) = \min\{\mathbf{u}^T M_K(p)\mathbf{u} : \mathbf{u}^T \mathbf{1} = 1 \text{ and } \mathbf{u} \ge 0\}.$$

Defined in [3], the matrix $M_K(p)$ is a weighted adjacency matrix for the CRG, $K \in \mathcal{K}$, with black, white and gray vertices/edges receiving weights (1 - p), p and 0, respectively.

Both equations (1) and (2) can readily supply an upper bound for the edit distance function, since $ed_{\mathcal{H}}(p) = \inf\{f_K(p) : K \in \mathcal{K}\} = \inf\{g_K(p) : K \in \mathcal{K}\} = \min\{g_K(p) : K \in \mathcal{K}\}$, where this final equality is a result from [10].

The maximum value of the edit distance function, $d_{\mathcal{H}}^*$, was originally defined as follows:

$$d_{\mathcal{H}}^* = \lim_{n \to \infty} \operatorname{Dist}(n, \mathcal{H}) / {n \choose 2}.$$

That is, it is the limit as $n \to \infty$ of the maximum possible distance of any graph on n vertices from a hereditary property \mathcal{H} normalized by the total number of possible edges in such a graph. We refer to the value of p, or the set of values of p, for which $ed_{\mathcal{H}}(p) = d^*_{\mathcal{H}}$, as $p^*_{\mathcal{H}}$. It follows from the concavity of the edit distance function that $p^*_{\mathcal{H}}$ must be a closed, though potentially degenerate, interval.

In practice, the edit distance function can be difficult to determine for a given hereditary property, and most methods depend heavily on the structures of related colored regularity graphs (or *types*, if one employs the paradigm used in [10]) introduced in [1]. In the case of Forb $(K_{2,t})$, the hereditary property of forbidding an induced $K_{2,t}$ subgraph, however, much can be said about the edit distance function, especially in the cases when t is small or p > 1/2. The following theorem summarizes some known results for the edit distance function of Forb $(K_{2,t})$.

Theorem 1. Let $\mathcal{H} = Forb(K_{2,t})$, and $ed_{\mathcal{H}}(p)$ be the edit distance function for \mathcal{H} , then

- (1) (Marchant-Thomason [10]) If t = 2, then $ed_{\mathcal{H}}(p) = p(1-p)$.
- (2) (Marchant-Thomason [10]) If $p \ge 1/2$, then $ed_{\mathcal{H}}(p) = (1-p)/(t-1)$.
- (3) ([12]) For p < 1/2,

• If
$$t = 3$$
, then $ed_{\mathcal{H}}(p) = p(1-p)$

• If t = 4, then $ed_{\mathcal{H}}(p) = p(1-p)$. • If t = 4, then $ed_{\mathcal{H}}(p) = \min\left\{p(1-p), \frac{7p+1}{15}, \frac{1-p}{3}\right\}$.

The part of the function $ed_{\text{Forb}(K_{2,4})}(p)$ with value (7p + 1)/15 corresponds to a construction derived from the so-called (15, 6, 1, 3)-strongly regular graph, which is commonly called the generalized quadrangle GQ(2, 2). We will discuss more about the connection between $ed_{\text{Forb}(K_{2,t})}(p)$ and strongly regular graphs later.

In this paper we continue the exploration of what can be said about the edit distance function of Forb $(K_{2,t})$ when $t \ge 5$. In particular, we will extend the interval from [10] for which this function is generally known, and explore some new colored regularity graph constructions that reduce the known upper bound for these edit distance functions on certain intervals. One of these constructions originates from a construction in [9] by Füredi for $K_{2,t}$ -free graphs. When t is large enough,

we demonstrate a similar result to that in [10] from Brown's construction. Other constructions arise from powers of cycles and strongly regular graphs.

1.1. New results.

In [10], it is established that for $p \ge 1/2$ and $\mathcal{H} = \operatorname{Forb}(K_{2,t})$, the edit distance function $ed_{\mathcal{H}}(p) = (1-p)/(t-1)$. We extend this result to hold true for $p \ge 2/(t+1)$.

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

This extension along with a new CRG construction results in the determination of $d_{\mathcal{H}}^*$, the maximum value of the edit distance function, for all odd t. Using the general lower bound in Theorem 3 below, we also demonstrate, via Theorem 4, that this maximum value occurs on a nondegenerate *interval* of values for p. That is, $p_{\mathcal{H}}^*$ is not a single value for all odd $t \geq 5$.

Theorem 3. Let $t \ge 3$ and p < 1/2. If K is a black-vertex, p-core CRG with white and gray edges such that the gray edges have neither a $K_{2,t}$ nor a B_{t-2} (as defined in Lemma 12), then

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

Theorem 4. For odd $t \geq 5$ and $\mathcal{H} = Forb(K_{2,t})$,

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

For small p and t large enough we demonstrate a "Zarankiewicz effect" similar to that discovered in [10] for Forb $(K_{3,3})$ and rejected for t = 3 and 4 in [12].

Theorem 5. For $\mathcal{H} = Forb(K_{2,t})$, the edit distance function $ed_{\mathcal{H}}(p) \leq \frac{t-1+p(2q^2-q(t-1)-2t)}{2(q^2-1)}$ for any prime power q such that t-1 divides q-1.

Corollary 6. For $t \ge 9$, there exists a value q_0 , so that if $q > q_0$, then $\frac{t-1+p(2q^2-q(t-1)-2t)}{2(q^2-1)} < p(1-p)$ for some values of p, which approach 0 as q increases. That is, arbitrarily close to p = 0, there is some value for p such that $ed_{\mathcal{H}}(p) < p(1-p)$.

A strongly regular graph construction provides a key upper bound for $ed_{Forb(K_{2,4})}(p)$. Such constructions continue to be relevant for larger t values.

Theorem 7. For any (k, d, λ, μ) strongly regular graph, there exists a corresponding CRG, K, such that

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

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

(4)
$$f_K(p) = \frac{t-1}{t-1+d(d+1)} + \left(1 - \frac{(d+2)(t-1)}{t-1+d(d+1)}\right)p.$$

The following upper bound arises from a CRG construction involving the second power of cycles. **Theorem 8.** For $\mathcal{H} = Forb(K_{2,t})$,

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

There is a very close connection between the result in Theorem 3 and strongly regular graphs. If we take the expression in (4) and minimize it with respect to d (see expression (5)), then we obtain the expression on the right-hand side of (3). In particular, we show that if a (k, d, t - 3, t - 1)strongly regular graph exists, then the corresponding CRG, K, has $f_K(p)$ tangent to the curve $p - \frac{t-1}{4t-5}[3p - 2 + 2\sqrt{1-3p} + (t+1)p^2]$ for some value of p. Thus, we obtain the exact value of $ed_{\text{Forb}(K_{2,t})}$ for that value of p.

Below are known upper bounds for $5 \le t \le 8$. It should be noted that as our knowledge of existing strongly regular graphs increases, new upper bounds are also likely to be discovered.

Theorem 9. Let $\mathcal{H} = Forb(K_{2,t})$.

 $\begin{aligned} \text{ If } t &= 5, \ then \\ &ed_{\mathcal{H}}(p) \leq \min\left\{p(1-p), \frac{1+75p}{96}, \frac{1+26p}{40}, \frac{1+5p}{13}, \frac{1}{6}, \frac{1-p}{4}\right\}. \\ \text{ If } t &= 6, \ then \\ &ed_{\mathcal{H}}(p) \leq \min\left\{p(1-p), \frac{1+63p}{85}, \frac{1+14p}{26}, \frac{1+7p}{17}, \frac{1+2p}{10}, \frac{1-p}{5}\right\}. \\ \text{ If } t &= 7, \ then \\ &ed_{\mathcal{H}}(p) \leq \min\left\{p(1-p), \frac{1+124p}{156}, \frac{1+76p}{100}, \frac{1+44p}{64}, \frac{1+31p}{49}, \frac{1+20p}{36}, \frac{1+5p}{16}, \frac{1}{8}, \frac{1-p}{6}\right\}. \\ \text{ If } t &= 8, \ then \\ &ed_{\mathcal{H}}(p) \leq \min\left\{p(1-p), \frac{1+124p}{156}, \frac{1+95p}{125}, \frac{1+53p}{76}, \frac{1+20p}{36}, \frac{1+11p}{25}, \frac{1+5p}{16}, \frac{3p+1}{13}, \frac{1-p}{7}\right\}. \end{aligned}$

We compare these upper bounds to the lower bound in Theorem 3 via the figures in Appendix A.

1.2. Organization.

Section 2 reviews some relevant definitions and results that also appear in [12]. Section 3 addresses the proofs of Theorems 2, 3 and 4. In Section 4, we present several new CRG constructions that yield upper bounds for $ed_{Forb(K_{2,t})}(p)$ in general. Sections 5 and 6 are conclusions and acknowledgements, respectively. Section A contains the figures that compare our upper and lower bounds.

2. Pertinent definitions and past results

In order to continue our exploration of $Forb(K_{2,t})$ begun in [10] by Marchant and Thomason, and pursued in [12], we first need to review some important results and definitions also used in [12].

Colored regularity graphs that do not permit colored homomorphisms from forbidden induced subgraphs associated with a hereditary property are imperative for determining the edit distance function. If there does not exist a colored homomorphism from a simple graph H to a CRG, K, then we say that K forbids an H embedding. In the case of $Forb(K_{2,t})$ we are, not surprisingly, looking for CRGs that forbid a $K_{2,t}$ embedding.

To this end, it is sometimes convenient to refer only to the subgraph induced by edges of a particular color in K. We will refer to these subgraphs as the white, black and gray subgraphs of K. We also define a *sub-CRG of* K as any complete subgraph of K retaining its original coloring (i.e., a sub-CRG of K is formed by deleting some vertices of K). Frequently, the color of the vertices in the CRGs we examine may be restricted to black due to several results from [10] with regard to a subset of CRGs known as *p*-cores.

Definition 10 (Marchant-Thomason [10]). 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 $g_{K'}(p) > g_K(p)$.

As before, we let K(w, b) denote the CRG with w white vertices, b black vertices and only gray edges.

Theorem 11 (Marchant-Thomason [10]). Let $\mathcal{H} = Forb(K_{2,t})$ and $t \geq 3$. For $p \leq \frac{1}{2}$, either

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

Since the value of the edit distance function for $\operatorname{Forb}(K_{2,t})$ is already known to be $\frac{1-p}{t-1}$ for p > 1/2, the restriction of vertex color to black holds in most instances explored in this paper. Hence we also make repeated use of the following lemma.

Lemma 12 (Marchant-Thomason [10]). A CRG, K, on all black vertices with only white and gray edges forbids a $K_{2,t}$ 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 [7]. 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.

From [10], we know that 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 [10]). 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.

Two propositions from [11], which follow easily from [10], relate the optimal weight function and the function described in equation (2). They are used extensively in [12] and will also play a significant role in the proofs of Theorems 2, 3 and 4.

Proposition 14 ([11]). 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 ([11]). Let K be a p-core CRG with all vertices black, then for $p \in [0, 1/2]$ and optimal weighting \mathbf{x} ,

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

Recall from [12] that since $ed_{\mathcal{H}}(p) \leq p(1-p)$ for $\mathcal{H} = \operatorname{Forb}(K_{2,t})$, Proposition 14 gives the lower bound $d_G(v) \geq p + \frac{1-2p}{p} \mathbf{x}(v)$, and Proposition 15 restricts the optimal weights of all vertices in Kto be no more than p.

3. Proofs of Theorems 2, 3 and 4

In this section we extend the generally known interval for $ed_{\text{Forb}(K_{2,t})}(p)$ from $p \in [1/2, 1]$ to $p \in [\frac{2}{t+1}, 1]$. With a new CRG construction, this extension is sufficient to determine $d_{\mathcal{H}}^*$ and a subset of $p_{\mathcal{H}}^*$ for odd t. Subsection 3.1 contains the proof of Theorem 2, while the remaining subsections address Theorems 3 and 4.

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

Proof of Theorem 2. Let K be a p-core CRG for $p \in \left[\frac{2}{t+1}, 1\right]$ that does not permit $K_{2,t}$ embedding for $t \geq 5$. If we assume that $g_K(p) < g_{K(0,t-1)}(p) = (1-p)/(t-1)$, then by Theorem 11, K has only black vertices and no black edges.

As in [12], partition the vertices of K into three sets $\{u_0\}, U = \{u_1, \ldots, u_\ell\}$ and W, where u_0 is a fixed vertex with maximum weight x; U is the set of all vertices in the gray neighborhood of u_0 with u_1 a vertex of maximum weight x_1 in U; and W is the set of all remaining vertices, or those vertices adjacent to u_0 via white edges. Finally, let $d_G(u_i)$ signify the sum of the weights of all vertices in the gray neighborhood of u_i .

Then by Lemma 12, the total weight of the vertices in W is greater than

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

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

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

Applying Proposition 15 and letting $g_K(p) = g$ for ease of notation,

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

$$2(p-g) - p + (1-2p)x \leq [p(t-1)-1]x_1$$

$$2(p-g) - p + (1-2p)x \leq [p(t-1)-1]x$$

$$p - 2g \leq [p(t+1)-2]x.$$

Since $p \ge \frac{2}{t+1}$ and $x \le \frac{g}{1-p}$ by Proposition 15,

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

By Theorem 11, $g \leq \frac{1-p}{t-1}$, for $p \in [\frac{2}{t+1}, 1]$, so $ed_{Forb(K_{2,t})}(p) = \frac{1-p}{t-1}$.

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

3.2. A construction for odd t.

Proposition 16. Let $\mathcal{H} = Forb(K_{2,t})$ for odd t. Then $ed_{\mathcal{H}}(p) \leq 1/(t+1)$.

Proof. Let K be the CRG consisting of t+1 black vertices with white subgraph forming a perfect matching and all other edges gray. The CRG, K, does not contain a gray $K_{2,t}$ or book B_{t-2} , and so by Lemma 12, K forbids a $K_{2,t}$ embedding.

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

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

 $\leq 1/(t+1).$

Therefore, $ed_{\mathcal{H}}(p)$ L/(t

Since by Theorem 2, $ed_{\text{Forb}(K_{2,t})}(\frac{2}{t+1}) = \frac{1}{t+1}$, and by Proposition 16, $ed_{\text{Forb}(K_{2,t})} \leq \frac{1}{t+1}$, we have that $d^*_{\operatorname{Forb}(K_{2,t})} = \frac{1}{t+1}$ for odd $t \ge 5$.

3.3. A general lower bound for t.

We conclude this section by determining a general lower bound for the edit distance function of Forb $(K_{2,t})$. It is the lower bound from Theorem 3, and it allows us to make the claim in Theorem 4 that, in the case of odd t, there is a nondegenerate interval $p_{\mathcal{H}}^*$ that achieves the maximum value of the function.

Proof of Theorem 3. Here we use the standard bounds from Propositions 14 and 15. Let $g = g_K(p)$, where K is a black-vertex, p-core CRG, and let $N_G(v)$ denote the gray neighborhood of a given vertex v in K. Then if u_1, \ldots, u_ℓ are the vertices in the gray neighborhood, U, of a fixed vertex of maximum weight, u_0 ,

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

The left-hand side of this inequality calculates the weight of the total gray neighborhood of each vertex in U that must be contained in W, the set of all vertices not in U or u_0 . On the right-hand side we make use of the facts that $\mathbf{x}(W) = 1 - x - d_G(u_0)$ and that no vertex in W may be adjacent to more than (t-1) vertices in U without violating Lemma 12 by forming a gray $K_{2,t}$ with u_0 . Thus, applying Proposition 14,

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

Again considering Lemma 12 reveals that no vertex $u_i \in U$ can have more than t-3 gray neighbors in U without inducing a gray book B_{t-2} with u_0 . Therefore,

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

Recalling that by Proposition 15, $\frac{p-g}{p} \ge x$, we use the bound $\ell \ge d_G(u_0)/x$ (which follows from the pigeon-hole principle) to get

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

By Proposition 14,

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

Collecting terms yields,

$$\left(\frac{p-g}{p}\right)^2 + \left[\left(\frac{p-g}{p}\right)\left(\frac{2-3p}{p}\right) - (t-1)\right]_7 x + \left[\left(\frac{1-2p}{p}\right)\left(\frac{1-p}{p}\right) + (t-1)\right] x^2 \le 0,$$

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

$$\left(\frac{p-g}{p}\right)^2 - \frac{\left[\left(t-1\right) - \left(\frac{p-g}{p}\right)\left(\frac{2-3p}{p}\right)\right]^2}{4\left[\left(\frac{1-2p}{p}\right)\left(\frac{1-p}{p}\right) + \left(t-1\right)\right]} \le 0$$

$$\left(\frac{p-g}{p}\right)^2 \left(4t-5\right) + 2\left(\frac{p-g}{p}\right)\left(t-1\right)\left(\frac{2-3p}{p}\right) - \left(t-1\right)^2 \le 0.$$

Using the quadratic formula,

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

The function in (3) achieves its maximum at $p = \frac{2t-1}{t^2+t}$, and that maximum is, in fact, $\frac{1}{t+1}$. Hence $ed_{\text{Forb}(K_{2,t})}(p)$ is at least $\frac{1}{t+1}$ at $p = \frac{2t-1}{t(t+1)}$ and is at least $\frac{1}{t+1}$ at $p = \frac{2}{t+1}$. As a result of concavity,

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

Equality holds whenever t is odd because, in that case, Proposition 16 gives that $ed_{\text{Forb}(K_{2,t})}(p) \leq 1/(t+1)$, so $p_{\mathcal{H}}^*$ must be an interval. This concludes the proof of Theorem 4.

As can be seen from an analysis of the first and second derivatives of

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

with respect to p, the maximum difference between p(1-p) and the lower bound in Theorem 3 on the interval $[0, \frac{2}{t+1}]$ occurs when $p = \frac{2}{t+1}$ as long as $t \ge 5$. The ratio of this difference to the value of p(1-p) at this point increases with t, and approaches 1/2 as $t \to \infty$.

4. Upper bound constructions

That we have been able to determine the entire edit distance function for $\operatorname{Forb}(K_{2,3})$ and $\operatorname{Forb}(K_{2,4})$ raises the question of whether it might be possible to do something similar for $\operatorname{Forb}(K_{2,t})$ when $t \geq 5$. That is, can we always find a few simple upper bounds that determine the entire edit distance function? In this section we show that when $t \geq 5$, the number and types of known upper bounds for the function increases significantly, though this does not necessarily preclude the possibility that a few, yet to be discovered, CRG constructions could determine the entire function. We start with the following CRG construction, inspired by the theory in [4].

we start with the following Orto construction, inspired by the theory

4.1. Results from strongly regular graph constructions.

Recall that a strongly regular graph with parameters (k, d, λ, μ) is a d-regular graph on k vertices such that each pair of adjacent vertices has λ common neighbors, and each pair of nonadjacent vertices has μ common neighbors. Here we develop a function based on the existence of a strongly regular graph. Suppose that K is a CRG with all vertices black and all edges white or gray that is derived from a (k, d, λ, μ) -strongly regular graph so that the edges of the strongly regular graph correspond to gray edges of K. In such a case we recall from [12] that

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

As is commonly known (see [13], for instance), if a strongly regular graph with parameters (k, d, λ, μ) exists then it is necessary, though not sufficient, for

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

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

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

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

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

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

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

which is equal to the lower bound from (3) in Theorem 3.

Of course, in order to even have a chance of actually attaining (5) with a strongly regular graph construction, both d and $k = \frac{t-1+d(d+1)}{t-1}$ must be integers. This equation, however, provides something of a best case scenario for strongly regular graphs, and if there is a CRG, K, derived from a (k, d, t-3, t-1)-strongly regular graph that realizes equation (5), then $f_K(p)$ is tangent to the lower bound in (3) at

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

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

The remaining upper bounds in Theorem 9 are the result of checking constructions from the known strongly regular graphs listed at [5]. Figure 1 is a chart of the relevant parameters and $f_K(p)$ functions for $5 \le t \le 8$.

To our knowledge, it is not known whether, for fixed t, there are a finite or infinite number of (k, d, t-3, t-1)-strongly regular graphs. See Elzinga [8] for values of λ and μ for which the number of strongly regular graphs with parameters (k, d, λ, μ) is known to be finite or infinite.

There is an additional construction defining the upper bound for t = 8 in Theorem 8, described in the following section.

4.2. Cycle construction.

Definition 17 ([13], p.296). For two vertices $x, y \in V(G)$, where G is a simple connected graph, let dist(x, y) denote the length of the minimum path from x to y. The rth power of G, G^r , is the graph with vertex set $V(G^r) = V(G)$, and edge set $E(G^r) = \{xy : x \neq y \text{ and } dist(x, y) \leq r\}$.

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

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

Proposition 18. $C_{5+t,2}$ forbids a $K_{2,t}$ embedding, and therefore $ed_{Forb(K_{2,t})}(p) \leq \frac{3p+1}{5+t}$.

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

Second, we check that $C_{5+t,2}$ does not contain a gray B_{t-2} . If u_1 and u_2 are any two nonadjacent vertices in C_{5+t}^2 , then $|(N(u_1) \cup N(u_2)) - \{u_1, u_2\}| \ge 6$. Therefore, by reasoning similar to above, $|N(u_1) \cap N(u_2)| \le t-3$ in the complement of C_{5+t}^2 , implying $C_{5+t,2}$ does not contain a gray B_{t-2} . Thus, by Lemma 12, $C_{5+t,2}$ forbids a $K_{2,t}$ embedding, and therefore $ed_{\text{Forb}(K_{2,t})}(p) \le f_{C_{5+t,2}}(p) = \frac{3p+1}{5+t}$.

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

For $t \ge 5$, $f_{C_{5+t,2}}(p)$ is always an improvement on the bound $\min\{p(1-p), \frac{1-p}{t-1}\}$ from Theorem 11, though it is improved upon or made irrelevant by bounds from strongly regular graphs, for $t \le 7$. When t = 4, the function $f_{C_{9,2}}(p)$ is tangent to the edit distance function at p = 1/3, where the edit distance function achieves its maximum value.

4.3. Füredi constructions.

As is observed in Lemma 12 and used in the exploration of the past two constructions, graphs that forbid $K_{2,t}$ and B_{t-2} as subgraphs are of interest when looking for CRGs that forbid $K_{2,t}$ embedding. The following results come from examining the bipartite versions of $K_{2,t}$ -free graph constructions described by Füredi [9]. This strategy mimics the one used in [10] with Brown's $K_{3,3}$ -free construction.

Proof of Theorem 5. We take the construction described in [9] for a $K_{2,t}$ -free graph G on $n = (q^2 - 1)/(t - 1)$ vertices, each with degree q, where q is a prime power so that t - 1 divides q - 1. We should note here that in the original construction from [9], loops were omitted, reducing the degree of some vertices to q - 1. It is to our advantage, however, to leave the loops in so that the final construction will be q-regular. By the same proof as in [9], the graph with loops still retains the property that no two vertices have a common neighborhood greater than t - 1 even when a looped vertex is considered to be in its own neighborhood.

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

The gray subgraph of K is bipartite, so it cannot contain a B_{t-2} , and since no two vertices v_i and v_j from the original construction have more than t-1 common neighbors, the common neighborhood of two vertices in the gray subgraph of K is also at most t-1. Thus by Lemma 12, K forbids a $K_{2,t}$ embedding.

The CRG, K, has $k = 2n = 2(q^2 - 1)/(t - 1)$ vertices and $q(q^2 - 1)/(t - 1)$ gray edges, so by equation (1), $f_K(p)$ is as described in the statement of Theorem 5.

Remark 19. Though the property of being bipartite is sufficient to exclude a B_{t-2} subgraph, using a bipartite $K_{2,t}$ -free construction may not be the optimal choice. A more efficient CRG may be constructed from another graph that has a gray subgraph that is both $K_{2,t}$ - and B_{t-2} -free, but, for instance, still contains triangles.

Nevertheless, we can discover more about the potential for these constructions to improve upon the bounds for $ed_{Forb(K_{2,t})}(p)$ by fixing p and considering the general formula in Theorem 5 as a continuous function with respect to q.

Lemma 20. Let $t \ge 3$, and let $q_0 < q$ be prime powers such that t - 1 divides both $q_0 - 1$ and q - 1. If the CRG, K_0 , is constructed according to the proof of Theorem 5 with parameter q_0 and if the CRG, K, is constructed according to the proof of Theorem 5 with parameter q, then $f_{K_0}(p) \le f_K(p)$ for $p \in \left[\frac{2}{4+q_0}, \frac{1}{3}\right)$.

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

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

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

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

Additionally, we can make some statements about when we can expect constructions that originate from the $K_{2,t}$ -free graphs described by Füredi [9] to improve upon the bound p(1-p) for any q.

Lemma 21. Fix $t \ge 9$, and let q be a prime power such that t - 1 divides q - 1. Let K be the CRG with parameter q described in the proof of Theorem 5, hence $f_K(p) = \frac{t-1+p(2q^2-q(t-1)-2t)}{2(q^2-1)}$. Then for any sufficiently large prime power q and corresponding K, there is an interval of values of p on which $f_K(p) < p(1-p)$. Moreover as $q \to \infty$ the left-hand endpoints of these open intervals approach 0.

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

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

(6)
$$p - \frac{p(q(t-1)+2t-2)-(t-1)}{2(q^2-1)}
$$2p^2(q^2-1) < p(q(t-1)+2(t-1))-(t-1)$$
$$2p^2(q^2-1) - p(t-1)(q+2) + (t-1) < 0.$$$$

The minimum value of $2p^2(q^2-1) - p(t-1)(q+2) + (t-1)$ occurs when $p = \frac{(t-1)(q+2)}{4(q^2-1)}$. Therefore, the inequality above is satisfied for some q and p values if and only if

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

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

Lemma 22. Fix $5 \le t \le 8$, and let q be a prime power such that t - 1 divides q - 1. Let K be the CRG with parameter q described in the proof of Theorem 5, hence $f_K(p) = \frac{t-1+p(2q^2-q(t-1)-2t)}{2(q^2-1)}$. Then

(7)
$$q < \frac{(t-1) + \sqrt{(t-1)^2 + (9-t)(t+1)}}{\frac{1}{2}(9-t)}.$$

Proof. Returning to inequality (6) and performing a similar analysis to that in the proof of Lemma 21, we see that if $t \le 8$, then $2p^2(q^2 - 1) - p(t - 1)(q + 2) + (t - 1) < 0$ for some value of p if and only if

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

The lower bound for q described above is immaterial since for $t \leq 8$ it is always negative. The upper bound completes the proof of Lemma 22.

Using Lemma 22, we generated the following table of possible q values that obey the inequality in (7). Since we have already determined the entire edit distance function for Forb $(K_{2,3})$ and Forb $(K_{2,4})$, only t = 5, 6, 7, 8 needed to be considered:

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

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

5. Conclusions

• While we have yet to determine the entire edit distance function $ed_{Forb(K_{2,t})}(p)$ when $t \ge 5$, strongly regular graph constructions have the potential to determine its value exactly, at

least for certain values of p. It is likely that with the development of knowledge of strongly regular graphs, we will see improved upper bounds for $ed_{K_{2,t}}(p)$. Already, they provide significant improvements to previously known upper bounds, in some cases realizing the function value exactly.

• For $t \ge 9$, Füredi's $K_{2,t}$ -free construction leads to improvements to the bound p(1-p) for values of p arbitrarily close to 0. In fact, analyzing inequality (6) with respect to q, indicates that for fixed p and $t \ge 9$, if an appropriate q exists such that

$$\frac{t-1-\sqrt{(t-1)^2-8(t-1)(1-2p)-16p^2}}{4p} < q < \frac{t-1+\sqrt{(t-1)^2-8(t-1)(1-2p)-16p^2}}{4p},$$

then there is an improvement.

Meanwhile, for $t \leq 8$, the upper bounds from these constructions are inferior to those from alternative constructions. It is unknown, whether or not these improvements are best possible, or if there is an unknown construction that could render them irrelevant too.

6. Acknowledgements

Special thanks to Sung-Yell Song for his knowledge and expertise on strongly regular graphs and Maria Axenovich for providing references on Füredi's constructions. Thank you to Ed Marchant and Andrew Thomason for helpful conversations and to Stephen Hartke for providing some references on strongly regular graphs.

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] Stephan Brandt. On the structure of dense triangle-free graphs. Combin. Probab. Comput., 8(3):237–245, 1999.
- [5] Andries E. Brouwer. Parameters of strongly regular graphs. http://www.win.tue.nl/~aeb/graphs/srg/srgtab.html (accessed July 1, 2010).
- [6] W. G. Brown. On graphs that do not contain a Thomsen graph. Canad. Math. Bull., 9:281-285, 1966.
- [7] Guantao Chen, Ralph J. Faudree, and Ronald J. Gould. Saturation numbers of books. Electron. J. Combin., 15(1):Research Paper 118, 12, 2008.
- [8] Randall J. Elzinga. Strongly regular graphs: values of λ and μ for which there are only finitely many feasible (v, k, λ, μ) . Electron. J. Linear Algebra, 10:232–239 (electronic), 2003.
- [9] Zoltán Füredi. New asymptotics for bipartite Turán numbers. J. Combin. Theory Ser. A, 75(1):141–144, 1996.
- [10] 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.
- [11] Ryan Martin. The edit distance function and localization. arXiv:1007.1897v3.
- [12] Ryan Martin and Tracy McKay. On the edit distance from $K_{2,t}$ -free graphs I: Cases t = 3, 4. Preprint.
- [13] Douglas B. West. Introduction to Graph Theory. Prentice Hall, Upper Saddle River, NJ, 2nd edition, 2001.

Appendix A. Figures

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

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

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

FIGURE 1. Above are the known parameters (see [5]) and $f_K(p)$ functions from strongly regular graphs that provide an improvement upon the known upper bound for $ed_{\mathcal{H}}(p)$ for some interval of p values, where $\mathcal{H} = \operatorname{Forb}(K_{2,t})$ for $5 \leq t \leq 8$. Parameters with resulting bounds surpassed by other strongly regular graph constructions are omitted.



