ON THE COMPUTATION OF EDIT DISTANCE FUNCTIONS

RYAN MARTIN

Abstract. The edit distance between two graphs on the same labeled vertex set is the size of the symmetric difference of the edge sets. The edit distance function of hereditary property, \( \mathcal{H} \), is a function of \( p \in [0, 1] \) and is the limit of the maximum normalized distance between a graph of density \( p \) and \( \mathcal{H} \).

This paper uses localization, for computing the edit distance function of various hereditary properties. For any graph \( H \), \( \text{Forb}(H) \) denotes the property of not having an induced copy of \( H \). We compute the edit distance function for \( \text{Forb}(H) \), where \( H \) is any so-called split graph, and the graph \( H_9 \), a graph first used to describe the difficulties in computing the edit distance function.

1. Introduction

This paper uses the method of localization, introduced in \cite{12} as a way to compute edit distance functions. It uses some properties of quadratic programming, first applied by Marchant and Thomason \cite{11}. Some results on the edit distance function can be found in a variety of papers \cite{14, 5, 6, 1, 2, 3, 4, 10, 11, 13}. Much of the background to this paper can be found in a paper by Balogh and the author \cite{7}. Terminology and proofs of supporting lemmas that are suppressed here can be found in \cite{12}.

1.1. The edit distance function. A hereditary property is a family of graphs that is closed under isomorphism and the taking of induced subgraphs. The edit distance function of a hereditary property \( \mathcal{H} \), denoted \( ed_{\mathcal{H}}(p) \), measures the maximum distance of a density \( p \) graph from a hereditary property. Formally, if \( \text{Dist}(G, \mathcal{H}) = \min\{|E(G) \triangle E(G')| : |V(G')| = n, G' \in \mathcal{H}\} \), then

\[
ed_{\mathcal{H}}(p) = \lim_{n \to \infty} \max \left\{ \text{Dist}(G, \mathcal{H}) : |V(G)| = n, |E(G)| = \left\lfloor p \binom{n}{2} \right\rfloor \right\} / \binom{n}{2}.
\]

In \cite{7}, a result of Alon and Stav \cite{11} is generalized to show that the limit in \cite{7} does indeed exist for nontrivial hereditary properties and, furthermore,
that
\[ ed_\mathcal{H}(p) = \lim_{n \to \infty} \text{Dist}(G(n, p), \mathcal{H}) / \binom{n}{2}. \]
That is, the maximum edit distance is achieved, asymptotically, by the Erdős-Rényi random graph \( G(n, p) \). For any nontrivial hereditary property \( \mathcal{H} \) (that is, one that is not finite), the function \( ed_\mathcal{H}(p) \) is continuous and concave down. Hence, it achieves its maximum at a point \( (p^*_{\mathcal{H}}, d^*_{\mathcal{H}}) \). It should be noted that, for some hereditary properties, \( p^*_{\mathcal{H}} \) might be an interval.

1.2. Main results. The main results of this paper are Theorem 1 and Theorem 3.

A split graph is a graph whose vertex set can be partitioned into one clique and one independent set. If \( H \) is a split graph on \( h \) vertices with independence number \( \alpha \) and clique number \( \omega \), then \( \alpha + \omega \in \{h, h + 1\} \). The value of \( (p^*, d^*) \) had been obtained for the claw by Alon and Stav [2] and for graphs of the form \( K_\alpha + E_b \) (an \( \alpha \)-clique with \( b \) isolated vertices) by Balogh and the author [7].

**Theorem 1.** Let \( H \) be a split graph that is neither complete nor empty, with independence number \( \alpha \) and clique number \( \omega \). Then,

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

It is a trivial result (see, e.g., [12]) that \( ed_{\text{Forb}(K_\omega)}(p) = p/(\omega - 1) \) and \( ed_{\text{Forb}(K_\alpha)}(p) = (1 - p)/(\alpha - 1) \). So, we can combine Theorem 1 with the prior results for which \( H \) is either complete or empty.

**Corollary 2.** Let \( H \) be a split graph with independence number \( \alpha \) and clique number \( \omega \). Then, \( (p^*_{H}, d^*_{H}) = \left( \frac{\omega - 1}{\alpha + \omega - 2}, \frac{1}{\alpha + \omega - 2} \right) \).

![Figure 1. The graph \( H_9 \).](image-url)
The graph, $H_9$, as drawn in Figure 1.2, was given in [7] as an example of a hereditary property $\mathcal{H} = \text{Forb}(H_9)$ such that the maximum value of $ed(H)(p)$ cannot be determined by CRGs that only have gray edges. In [7] only an upper bound of $\min\{\frac{p}{3}, \frac{p}{1+4p}, \frac{1-p}{2}\}$ is provided for $ed_{\text{Forb}(H_9)}(p)$. Here we determine the function itself.

**Theorem 3.** Let $H_9$ be the graph in Figure 1.2. Then,

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

Consequently, $\left(p^*_{\text{Forb}(H_9)}, d^*_{\text{Forb}(H_9)}\right) = \left(\frac{1+\sqrt{17}}{8}, \frac{7-\sqrt{17}}{16}\right)$.

![Figure 2](image)

**Figure 2.** Plot of $ed_{\text{Forb}(H_9)}(p) = \min\{p/3, p/(1+4p), (1-p)/2\}$. The point $(p^*, d^*) = \left(\frac{1+\sqrt{17}}{8}, \frac{7-\sqrt{17}}{16}\right)$ is indicated.

The rest of the paper is organized as follows: Section 2 gives some of the general definitions for the edit distance function, such as colored regularity graphs. Section 3 defines and categorizes so-called $p$-core colored regularity graphs introduced by Marchant and Thomason [11]. Section 4 describes the method we use (localization), which is similar to the symmetrization idea in Sidorenko [15]. Section 5 proves Theorem 1 regarding split graphs. Section 6 proves Theorem 3 regarding the graph $H_9$. Section 7 is a section of acknowledgements.

2. **Background and basic facts**

2.1. **Notation.** All graphs are simple. If $S$ and $T$ are sets, then $S + T$ denotes the disjoint union of $S$ and $T$. If $v$ and $w$ are adjacent vertices in a graph, we denote the edge between them to be $vw$.

---

[1] CRGs are defined shortly in Section 2.
2.2. Colored regularity graphs. A colored regularity graph (CRG), $K$, is a simple complete graph, together with a partition of the vertices into black and white $V(K) = VW(K) + VB(K)$ and a partition of the edges into black, white and gray $E(K) = EW(K) + EG(K) + EB(K)$. We say that a graph $H$ embeds in $K$, (writing $H \mapsto K$) if there is a function $\varphi : V(H) \rightarrow V(K)$ so that if $h_1h_2 \in E(H)$, then either $\varphi(h_1) = \varphi(h_2) \in VB(K)$ or $\varphi(h_1)\varphi(h_2) \in EB(K) \cup EG(K)$ and if $h_1h_2 \notin E(H)$, then either $\varphi(h_1) = \varphi(h_2) \in VW(K)$ or $\varphi(h_1)\varphi(h_2) \in EW(K) \cup EG(K)$.

For a hereditary property of graphs, $\mathcal{H}$, we denote $K(\mathcal{H})$ to be the subset of CRGs such that no forbidden graph maps into $K$. That is, if $\mathcal{F}(\mathcal{H})$ is defined to be the minimal set of graphs so that $\mathcal{H} = \bigcap_{H \in \mathcal{F}(\mathcal{H})} \text{Forb}(H)$, then $K(\mathcal{H}) = \{K : H \not\rightarrow K, \forall H \in \mathcal{F}(\mathcal{H})\}$. A CRG $K'$ is said to be a sub-CRG of $K$ if $K'$ can be obtained by deleting vertices of $K$.

2.3. The $f$ and $g$ functions. For every CRG, $K$, we associate two functions. The function $f$ is a linear function of $p$ and $g$ is found by weighting the vertices. Let $K$ have a total of $k$ vertices $\{v_1, \ldots, v_k\}$, and let $M_K(p)$ be a matrix such that the entries are:

$$[M_K(p)]_{ij} = \begin{cases} p, & \text{if } v_i v_j \in VW(K) \cup EW(K); \\ 1-p, & \text{if } v_i v_j \in VB(K) \cup EB(K); \\ 0, & \text{if } v_i v_j \in EG(K). \end{cases}$$

Then, we can express the $f$ and $g$ functions over the domain $p \in [0, 1]$ as follows, with $VW = VW(K)$, $VB = VB(K)$, $EW = EW(K)$ and $EB = EB(K)$:

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

$$g_K(p) = \min \{ x^T M_K(p) x \mid x^T 1 = 1, x \geq 0 \}$$

If we denote $1$ to be the vector of all ones, then $f_K(p) = (\frac{1}{\kappa} 1)^T M_K(p) (\frac{1}{\kappa} 1)$. So, $f_K(p) \geq g_K(p)$.

**Theorem 4 ([7]).** For any nontrivial hereditary property $\mathcal{H}$,

$$ed_\mathcal{H}(p) = \lim_{K \in K(\mathcal{H})} g_K(p) = \lim_{K \in K(\mathcal{H})} f_K(p).$$

2.4. Basic observations on $ed_\mathcal{H}(p)$. The following is a summary of basic facts about the edit distance function. Item [iii] comes from Alon and Stav [1]. Item [iv] comes from [7].

**Theorem 5.** Let $\mathcal{H}$ be a nontrivial hereditary property with chromatic number $\chi$, complementary chromatic number $\overline{\chi}$, binary chromatic number $\chi_B$ and edit distance function $ed_\mathcal{H}(p)$.

(i) If $\chi > 1$, then $ed_\mathcal{H}(p) \leq p/(\chi - 1)$.

(ii) If $\overline{\chi} > 1$, then $ed_\mathcal{H}(p) \leq (1-p)/(\overline{\chi} - 1)$. 

(iii) \( ed_H(1/2) = 1/(2(\chi_B - 1)) \).
(iv) \( ed_H(p) \) is continuous and concave down.
(v) \( ed_H(p) = ed_P(1 - p) \).

3. The \( p \)-cores

In Marchant and Thomason \cite{11}, it is shown that

\[
\text{ed}_H(p) = \inf \{ g_K(p) : K \in \mathcal{K}(H) \} = \inf \{ f_K(p) : K \in \mathcal{K}(H) \}.
\]

Although the setting of that paper is not edit distance, the results can be translated to our setting. They show, in fact, that \( ed_H(p) = \min \{ g_K(p) : K \in \mathcal{K}(H) \} \).

That is, for any hereditary property \( H \) and \( p \in [0, 1] \), there is a CRG, \( K \in \mathcal{K}(H) \) such that \( ed_H(p) = g_K(p) \). This is found by looking at so-called \( p \)-cores. A CRG, \( K \), is a \( p \)-core CRG, or simply a \( p \)-core, if \( g_K(p) < g_{K'}(p) \) for all nontrivial sub-CRGs \( K' \) of \( K \). Marchant and Thomason prove that

\[
\text{ed}_H(p) = \min \{ g_K(p) : K \in \mathcal{K}(H) \text{ and } K \text{ is p-core} \}.
\]

4. Computing edit distance functions using localization

Upper bounds for the edit distance function of \( H \) are found by simply exhibiting some CRGs \( K \in \mathcal{K}(H) \) and computing \( g_K(p) \) by means of (4). The localization method obtains lower bounds for \( \text{ed}_H(p) \) and is similar to symmetrization from Sidorenko \cite{15}. The main tools are Lemmas 6 and 7, found in \cite{12}. We have already seen much of the theoretical underpinnings.

Given \( K \), a \( p \)-core, there is a unique optimum weight vector, \( x \), with all entries positive that is a solution to (4). For any vertex \( v \in V(K) \), \( d_G(v) \) denotes the sum of the weights of the gray neighbors of \( v \) under \( x \), \( d_W(v) \) the sum of the white neighbors (including \( v \) itself if its color is white) and \( d_B(v) \) the sum of the black neighbors (again, including \( v \) itself if its color is black). Consequently, \( d_G(v) + d_W(v) + d_B(v) = 1 \).

The fundamental concept is that we may, in many cases, assume the vertices are monochromatic (say, black) and all edges are either white or gray. The sizes of the gray neighborhoods are a function of the weight \( x(v) \). We formalize the observations below:

**Lemma 6.** Let \( H \) be a nontrivial hereditary property and \( p \in (0, 1) \), \( \mathcal{K}(H) \) the set of CRGs defined by \( H \). Then,

(i) \( ed_H(p) = \min \{ g_K(p) : K \in \mathcal{K}(H) \text{ and } K \text{ is p-core} \} \).
(ii) If \( p \leq 1/2 \) and \( K \) is a \( p \)-core CRG, then \( K \) has no black edges and white edges can only be incident to black vertices.
(iii) If \( p \geq 1/2 \) and \( K \) is a \( p \)-core CRG, then \( K \) has no white edges and black edges can only be incident to white vertices.
(iv) If \( x \) is the optimal weight function of a \( p \)-core CRG \( K \), then for all \( v \in V(K) \), \( g_K(p) = px_W(v) + (1 - p)x_B(v) \).

The overall idea is that we need only consider \( p \)-core CRGs and their special structure, then a great deal of information can be obtained by focusing
on a single vertex. This is referred to as “localization” because we can focus on one vertex at a time.

Lemma 7 has all of the elements to express $d_G(v)$ for any vertex $v$ in a $p$-core CRG. It is often useful to focus on the gray neighborhood of vertices.

**Lemma 7 (Localization).** Let $p \in (0, 1)$ and $K$ be a $p$-core CRG with optimal weight function $x$.

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

$$d_G(v) = \frac{p - g_K(p)}{p} + \frac{1 - 2p}{p} x(v),$$

for all $v \in VB(K)$.

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

$$d_G(v) = \frac{1 - p - g_K(p)}{1 - p} + \frac{2p - 1}{1 - p} x(v),$$

for all $v \in VW(K)$.

**Corollary 8.** Let $p \in (0, 1)$ and $K$ be a $p$-core CRG with optimal weight function $x$.

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

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

**Remark 9.** From this point forward in the paper, if $K$ is a CRG under consideration and $p$ is fixed, $x(v)$ will denote the weight of $v \in V(K)$ under the optimal solution of the quadratic program in equation (4) that defines $g_K$.

One more useful observation is Theorem 6 from [12]:

**Theorem 10.** A sub-CRG, $K'$, of a CRG, $K$, is a component if, for all $v \in V(K')$ and all $w \in V(K) - V(K')$, then $vw$ is gray. Let $K$ be a CRG with components $K^{(1)}, \ldots, K^{(\ell)}$. Then

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

5. Forb($H$), $H$ a split graph

We need to define a special class of graphs. For $\omega \geq 2$ and a nonnegative integer vector $(\omega; a_0, a_1, \ldots, a_{\omega})$, a $(\omega; a_0, a_1, \ldots, a_{\omega})$-clique-star is a graph $G$ such that $V(G)$ is partitioned into $A$ and $W$. The set $A$ induces an independent set, the set $W = \{w_1, \ldots, w_{\omega}\}$ induces a clique and for $i = 1, \ldots, \omega$, vertex $w_i$ is adjacent to a set of $a_i + 1$ leaves in $A$ and there are $a_0$ independent vertices. Note that this implies that $\sum_{i=0}^{\omega} a_i = \alpha - \omega$.

Colloquially, a clique-star can be partitioned into stars and independent sets such that the centers of the stars are connected by a clique and there

\[\text{We get the notation from Hung, Syslo, Weaver and West [9]. Barrett, Jepsen, Lang, McHenry, Nelson and Owens [8] define a clique-star, but it is a different type of graph.}\]
are no other edges. (If one of the stars is \(K_2\), one of the endvertices is designated to be the center.) Proving that Theorem 1 is true is much more difficult in the case where either \(H\) or its complement is a clique-star.

5.1. **Proof of Theorem 1.** Note that, because \(H\) is neither complete nor empty, \(\alpha, \omega \geq 2\). Without loss of generality, we may assume that \(\omega \leq \alpha\).

**Fact 11.** For any split graph \(H\), its chromatic number is its clique number. In notation, \(\chi(H) = \omega(H)\). Consequently, \(\chi(\overline{H}) = \alpha(H)\).

*Proof.* Trivially, \(\chi(H) \geq \omega(H)\). Let \(W \subseteq V(H)\) be a maximum-sized clique in \(H\). For every \(a \in V(H) - W\), there is some vertex \(w \in W\) such that \(a\) is nonadjacent to \(w\). Thus, \(w\) and its nonneighbors form an independent set and the union of such independent sets cover \(V(H)\). So, \(\chi(H) = |W| = \omega\). Because \(\overline{H}\) is also a split graph, \(\chi(H) = \omega(H) = \alpha(H)\).

Let \(K(w, b)\) denote the CRG with \(w\) white vertices, \(b\) black vertices and all edges gray. An immediate consequence of Fact 11 is that \(H\) cannot be embedded into \(K(\omega - 1, 0)\) and \(K(0, \alpha - 1)\) and so

\[
ed_{\text{Forb}(H)}(p) \leq \min \{g_{K(\omega - 1, 0)}(p), g_{K(0, \alpha - 1)}(p)\} = \min \left\{ \frac{p}{\omega - 1}, \frac{1 - p}{\alpha - 1} \right\}.
\]

Let \(K \in K(\text{Forb}(H))\) be a \(p\)-core CRG and denote \(g = g_K(p)\). By Lemma 6 any edge between vertices of different colors must be gray. Since \(H\) is a split graph, \(H\) would embed into any \(K\) with a pair of different-colored vertices. So, the vertices in \(K\) must be monochromatic. Furthermore, if \(K\) has only gray edges, then either \(K\) has at most \(\omega - 1\) white vertices or at most \(\alpha - 1\) black vertices. In particular, if \(p = 1/2\), then all edges must be gray and so \(ed_{\text{Forb}(H)}(1/2) = \min \left\{ \frac{1/2}{\omega - 1}, \frac{1/2}{\alpha - 1} \right\}\). Because we have assumed that \(\omega \leq \alpha\), and concavity comes from Theorem 5(iv), it is the case that

\[
ed_{\text{Forb}(H)}(p) = \frac{1 - p}{\alpha - 1}, \quad p \in [1/2, 1].
\]

If \(p < 1/2\) and \(K\) has white vertices, then Lemma 6(iii) gives that all edges must be gray. In that case, \(g_K(p) = \frac{1 - p}{\alpha - 1}\). So, we may assume that \(p < 1/2\) and \(K\) has only black vertices and only white or gray edges.

By virtue of the fact that a clique and independent set can intersect in at most one vertex, \(h \leq \alpha + \omega \leq h + 1\). This yields two cases.

**Case 1.** \(\alpha + \omega = h + 1\).

Let \(v\) be a largest-weight vertex and let \(x = x(v)\). If \(v\) has \(h - \omega = \alpha - 1\) gray neighbors, then \(H \mapsto K\).
Thus, because $x$ is the largest weight, Lemma 7(i) gives that

$$d_G(v) \leq (\alpha - 2)x$$

$$\frac{p - g}{p} + \frac{1 - 2p}{p} x \leq (\alpha - 2)x$$

$$p - g \leq (p\alpha - 1)x.$$  

If $p < 1/\alpha$, then $g > p \geq p/(\omega - 1)$. If $p \geq 1/\alpha$, then Corollary 8(i) gives that

$$p - g \leq (p\alpha - 1) \frac{g}{1 - p}$$

$$p(1 - p) \leq gp(\alpha - 1)$$

$$\frac{1 - p}{\alpha - 1} \leq g.$$  

This concludes Case 1.

**Case 2.** $\alpha + \omega = h$.

Let $p \in \left(0, \frac{\omega - 1}{h - 1}\right]$. Let $v$ be a vertex of largest weight $x = x(v)$. Lemma 7(i) gives that

$$d_G(v) \leq (h - \omega - 1)x$$

$$\frac{p - g}{p} + \frac{1 - 2p}{p} x \leq (\alpha - 1)x$$

$$p - g \leq (p(\alpha + 1) - 1)x.$$  

If $p < 1/(\alpha + 1)$, then $g > p \geq p/(\omega - 1)$. If $p \geq 1/(\alpha + 1)$, then Corollary 8(i) gives that

$$p - g \leq (p(\alpha + 1) - 1) \frac{g}{1 - p}.$$  

Then,

$$g \geq \frac{1 - p}{\alpha} \geq \frac{1 - \frac{\omega - 1}{h - 1}}{\alpha} = \frac{1}{h - 1} = \frac{\omega - 1}{\omega - 1} \geq \frac{p}{\omega - 1}.$$  

Finally, we may assume that $p \in \left(\frac{\omega - 1}{h - 1}, \frac{1}{2}\right]$. We have to split into two cases according to the structure of $H$.

**Case 2a.** $\alpha + \omega = h$ and there exists an $c \leq \omega - 1$ such that $H$ can be partitioned into $c$ cliques and an independent set of $\alpha - c$ vertices.

Because all vertices are black and all edges are either white or gray in $K$, there cannot be $\alpha$ vertices such that there are $\omega - 1$ of them are adjacent to all of the $\alpha - 1$ others via a gray edge. We will show that this condition requires that $g \geq \min\left\{\frac{p}{\omega - 1}, \frac{1 - p}{\alpha - 1}\right\}$. 


First, we show that if \( g \) is small then \( K \) must have a gray \((\omega - 1)\)-clique. Let \( v_1, \ldots, v_\ell \) be a maximal gray clique. That is, any edge between these vertices is gray and every vertex not in \( \{v_1, \ldots, v_\ell\} \) has at least one white neighbor in \( \{v_1, \ldots, v_\ell\} \). Let \( x_i = x(v_i) \) for \( i = 1, \ldots, \ell \) and let \( X = \sum_{i=1}^\ell x_i \).

Using Lemma 7(i) and maximality,

\[
\sum_{i=1}^\ell [d_G(v_i) - X + x_i] \leq (\ell - 1)(1 - X)
\]

\[
\ell \frac{p - g}{p} + \frac{1 - p}{p} X - cX \leq (\ell - 1)(1 - X)
\]

\[
p - \ell g \leq (2p - 1)X.
\]

Hence, either \( g > \frac{p}{(\omega - 1)} \) or \( K \) has a gray \((\omega - 1)\)-clique. We may thus suppose that \( K \) has a gray \((\omega - 1)\)-clique. Let one with maximum total weight be \( \{v_1, \ldots, v_{\omega - 1}\} \) with \( x_i = x(v_i) \) for \( i = 1, \ldots, \omega - 1 \) and \( X = \sum_{i=1}^{\omega - 1} x_i \). Further, let \( Y \) be the sum of the weights of the common gray neighbors of \( x_1, \ldots, x_{\omega - 1} \). Since \( X \) has the largest weight of any gray \((\omega - 1)\)-clique and \( Y \) has at most \( \alpha - \omega \) vertices,

\[
Y \leq (\alpha - \omega) \frac{X}{\omega - 1}.
\]

Therefore,

\[
\sum_{i=1}^{\omega - 1} [d_G(v_i) - X + x_i] \leq (\omega - 1)Y + (\omega - 2)(1 - X - Y)
\]

\[
(\omega - 1) \left( \frac{p - g}{p} - X \right) + \frac{1 - p}{p} X \leq Y + (\omega - 2)(1 - X)
\]

\[
(1 - X) - (\omega - 1) \frac{g}{p} + \frac{1 - p}{p} X \leq (\alpha - \omega) \frac{X}{\omega - 1}
\]

\[
1 + X \left( \frac{1}{p} - \frac{h - 2}{\omega - 1} \right) \leq \omega - 1 \frac{g}{p}
\]

(5)

If \( p < (\omega - 1)/(h - 2) \), then the term in parentheses in (5) is positive and \( g > \frac{p}{(\omega - 1)} \). If \( p \geq (\omega - 1)/(h - 2) \), then \( X \leq (\omega - 1) \frac{g}{1 - p} \) by Corollary 8. Hence,

\[
1 + \frac{g}{1 - p} \left( \frac{1}{p} - \frac{h - 2}{\omega - 1} \right) \leq \omega - 1 \frac{g}{p}
\]

\[
1 \leq g \left( \frac{\omega - 1}{p} - \frac{1}{p(1 - p)} - \frac{h - 2}{(\omega - 1)(1 - p)} \right)
\]

\[
1 \leq g \left( \frac{\alpha - 1}{1 - p} + \frac{\omega - 2}{\omega - 1} \cdot \frac{(\omega - 1) - p(h - 2)}{p(1 - p)} \right)
\]

Therefore, \( g \geq (1 - p)/(\alpha - 1) \). This concludes Case 2a.
Which graphs are in Case 2, but not Case 2a? Since \( \alpha + \omega = h \), every \( w \in W \) has at least one neighbor in \( A \). If any \( a \in A \) has more than one neighbor in \( W \), then we can greedily find at most \( \omega - 1 \) vertices in \( A \) such that the union of their neighborhoods is \( W \). Such a graph would be in Case 2a.

So, the graphs, \( H \) with \( \omega \leq \alpha \) that are in neither Case 1 nor Case 2a have the property that \( N(w) \cap N(w') \cap A = \emptyset \) for all distinct \( w, w' \in W \). This is exactly the case of a clique-star.

**Case 2b.** \( \alpha + \omega = h \) and \( G \) is a clique-star.

Let \( W = \{w_1, \ldots, w_\omega\} \) such that \( w_i \) has \( a_i + 1 \) neighbors in \( A \) for \( i = 1, \ldots, \omega \) and there are \( a_0 \) isolated vertices.

**Fact 12.** If \( \omega \geq 2 \) and \( H \) is a \( (\omega; a_0, \ldots, a_\omega) \)-clique-star and \( K \) is a black-vertex CRG such that either

- there exists a vertex with at least \( \alpha \) gray neighbors, or
- there exist vertices \( v_1, \ldots, v_\omega \) such that
  - \( \{v_1, \ldots, v_\omega\} \) is a gray clique,
  - for \( i = 1, \ldots, \omega - 1 \), \( v_i \) has \( \alpha - 1 \) gray neighbors, and
  - \( v_\omega \) has at least \( p \cdot (\alpha - \omega)/\omega \) gray neighbors (including \( v_1, \ldots, v_{\omega-1} \)).

Then, \( H \mapsto K \).

**Proof of Fact 12.** If \( K \) has a vertex, \( v \), with \( \alpha \) gray neighbors, then \( W \) can be mapped to \( v \) whereas each member of \( A = V(H) - W \) can be mapped to a different gray neighbor of \( v \). Thus \( H \mapsto K \). So, we may assume the maximum gray degree of \( K \) is at most \( \alpha - 1 \).

Without loss of generality, let \( a_1 \geq \cdots \geq a_\omega \). Our mapping is done recursively: Map \( w_\omega \) and one of its neighbors to \( v_\omega \). Map its remaining \( A \)-neighbors \( \{a_\omega \leq p \cdot (\alpha - \omega)/\omega \} \) of them to each of \( a_\omega \) gray neighbors of \( v_\omega \) that are not in \( \{v_1, \ldots, v_{\omega-1}\} \).

Having embedded \( w_\omega, \ldots, w_{i+1} \) and each of their respective \( A \)-neighbors into a total of at most \( \sum j=0^{j=\omega} (a_j + 1) \) vertices of \( K \), we map \( w_i \) and one of its \( A \)-neighbors into \( v_i \) and its remaining \( a_i \) \( A \)-neighbors into arbitrary unused gray neighbors of \( v_i \). After \( w_1 \) and its neighbors are mapped, we map the remaining \( a_0 \) isolated vertices arbitrarily into the vertices of \( K \) that were not already used.

This mapping can be accomplished because the fact that each of the \( v_i \) have at least \( \alpha - 1 \) gray neighbors ensures that, even at the last step, when \( w_1 \) and a neighbor is embedded, there are at least \( \alpha - 1 \) gray neighbors of \( v_1 \). The number of gray neighbors of \( v_1 \) that were used are the \( \omega - 1 \) vertices \( v_i \) and at most \( \sum j=2^j=\omega a_j = \alpha - \omega - a_1 - a_0 \) others, for a total of \( \alpha - 1 - a_1 - a_0 \). So, there are enough gray neighbors of \( v_1 \) to embed the \( a_1 \) neighbors of \( w_1 \) as well as the \( a_0 \) isolated vertices. Thus, \( H \mapsto K \). \( \square \)
Fact 13. Let $p \in (0, 1/2)$ and let $K$ be a black-vertex CRG. If $g_K(p) \leq \min \{p/(\omega - 1), (1 - p)/(\alpha - 1)\}$, then either

- $K \approx K(0, \alpha - 1)$, or
- there exists a vertex with at least $\alpha$ gray neighbors, or
- there exist vertices $v_1, \ldots, v_\omega$ such that
  - $\{v_1, \ldots, v_\omega\}$ is a gray clique,
  - for $i = 1, \ldots, \omega - 1$, $v_i$ has $\alpha - 1$ gray neighbors, and
  - $v_\omega$ has at least $\lfloor (\alpha - \omega)/\omega \rfloor + \omega - 1$ gray neighbors (including $v_1, \ldots, v_{\omega - 1}$).

Proof of Fact 13. Assume that no vertex has $\alpha$ gray neighbors. We find $v_1, \ldots, v_\omega$ greedily. Choose $v_1$ to be a vertex of largest weight. Stop if $i = \omega$ or if $N_G(v_1) \cap \cdots \cap N_G(v_i)$ is empty. Otherwise, let $v_{i+1}$ be a vertex of largest weight in that set. We will show later that this process creates at least $\omega$ vertices.

First, we find the number of gray neighbors of $v_1$, using the fact that $x_1$ is the largest weight.

$$|N_G(v_1)| \geq \left\lceil \frac{d_G(v_1)}{x_1} \right\rceil \geq \frac{p - g}{px_1} + \frac{1 - 2p}{p}.$$

Using Corollary 8(i), we have that $x_1 \leq g/(1 - p)$ and so

$$|N_G(v_1)| \geq \frac{1 - p - g}{g} \geq \alpha - 2.$$

Equality only occurs if $g = (1 - p)/(\alpha - 1)$ there are $\alpha - 1$ vertices, all of weight $1/(\alpha - 1)$, thus $K \approx K(0, \alpha - 1)$. So, we may assume $|N_G(v_1)| \geq \alpha - 1$.

For $i \in \{2, \ldots, \omega - 1\}$, we let $X = \sum_{j=1}^i x_j$ and consider the common gray neighborhood of $\{v_1, \ldots, v_i\}$. Its total weight is:

$$d_G(v_i) - (X - x_i) - \sum_{j=1}^{i-1} x(N_W(v_j)) = \frac{p - ig}{p} + \frac{1 - 2p}{p}X > 0,$$

because $i \leq \omega - 1$, $g \leq p/(\omega - 1)$, $p < 1/2$ and $X > x_i > 0$. Thus, $v_{i+1}$ must exist.

We use these calculations to obtain the size of $N_G(v_i)$ for $i = 2, \ldots, \omega - 1$. First note that $v_i$ has $i - 1$ gray neighbors among $\{v_1, \ldots, v_{i-1}\}$ and that every vertex that is a gray neighbor of each of $v_1, \ldots, v_i$ has weight at most $x_i$. As to the remaining vertices, partition the vertices in $N_G(v_i) - \{v_1, \ldots, v_{i-1}\} - \bigcap_{j=1}^{i-1} N_G(v_j)$ according to the least index $j$ for which the vertex is adjacent to $v_j$ via a white edge. By the choice of $v_1, \ldots, v_i$, such a vertex has weight at most $x_j = x(v_j)$. Consequently, we have a lower bound
for \( |N_G(v_i)| \):

\[
|N_G(v_i)| \geq (i - 1) + \sum_{j=1}^{i-1} \frac{1}{x_j} \left( g \left( p - \frac{1-p}{p} x_i \right) \right)
+ \frac{1}{x_i} \left( p - g \left( p - \frac{1-p}{p} x_i \right) \right).
\]

We can drop the ceilings to obtain the lower bound

\[
|N_G(v_i)| \geq (i - 1) + \frac{1}{x_i} \left( g \left( p - \frac{1-p}{p} x_i \right) \right)
+ \sum_{j=1}^{i-1} \frac{1}{x_j} \left( g \left( p - \frac{1-p}{p} x_i \right) \right).
\]

Now we look at the coefficients \( \frac{1}{x_i} \leq \frac{1}{x_2} \leq \cdots \leq \frac{1}{x_1} \). The total weight of gray neighbors with coefficient \( \frac{1}{x_i} \) is at most \( x \left( N_W(v_1) \right) \). The total weight of gray neighbors with coefficient \( \frac{1}{x_2} \) or \( \frac{1}{x_3} \) is at most \( x \left( N_W(v_1) \right) + x \left( N_W(v_2) \right) \) and so on. Therefore, we may further lower bound \( |N_G(v_i)| \) as follows:

\[
|N_G(v_i)| \geq (i - 1) + \sum_{j=1}^{i-1} \frac{1}{x_j} \left( g \left( p - \frac{1-p}{p} x_i \right) \right)
+ \frac{1}{x_i} \left( g \left( p - \frac{1-p}{p} x_i \right) \right).
\]

and observe that inequality (6) shows that the last numerator is nonnegative.

Using similar computations as before, the previous inequality simplifies to

\[
|N_G(v_i)| \geq (i - 1) + \sum_{j=1}^{i-1} \frac{1}{x_j} \left( g \left( p - \frac{1-p}{p} x_i \right) \right)
+ \frac{1}{x_i} \left( g \left( p - \frac{1-p}{p} x_i \right) \right).
\]

After some simplification

\[
|N_G(v_i)| \geq \frac{g}{p} \sum_{j=1}^{i} \frac{1}{x_j} - \frac{1-2p}{p} (i - 1) + \frac{p - (i+1)g}{px_i} + \frac{1-2p}{p} \frac{X}{x_i}.
\]

Using Jensen’s inequality, we see that

\[
\sum_{j=1}^{i} \frac{1}{x_j} \geq \frac{i}{X/i}.
\]
So, we return to (7) and subsequently use the fact that \( x_i \leq X/i \):

\[
|N_G(v_i)| \geq g \left( \frac{i^2}{X} \right) - \frac{1 - 2p}{p} (i - 1) + \frac{p - (i + 1)g}{px_i} + \frac{1 - 2p}{p} \left( \frac{X}{x_i} \right)
\]

\[
\geq \frac{gi^2}{pX} - \frac{1 - 2p}{p} (i - 1) + \frac{p - (i + 1)g + (1 - 2p)X}{pX/i}
\]

\[
= \frac{i(p - g)}{pX} + \frac{1 - 2p}{p}.
\]

Using the fact that \( X \leq ig/(1 - p) \) from Corollary 8(i), we see that

\[
|N_G(v_i)| \geq \frac{1 - p - g}{g} \geq \alpha - 2.
\]

Equality only occurs if \( g = (1 - p)/\alpha - 1 \) and \( K \approx K(0, \alpha - 1) \). So, we may assume \( |N_G(v_i)| \geq \alpha - 1 \) for \( i = 1, \ldots, \omega - 1 \).

Finally, we try to determine the number of vertices adjacent to \( v_\omega \) via a gray edge. We only need \( |N_G(v_\omega)| \geq \lfloor \frac{\alpha}{\omega} \rfloor + \omega - 2 \) in order to finish the proof. First, note that the very existence of \( v_\omega \) ensures that \( |N_G(v_\omega)| \geq \omega - 1 \).

Second, suppose that \( \omega \geq 3 \). Recalling that \( d_G(v_\omega) = \frac{p - g}{p} + \frac{1 - 2p}{p} x_\omega \) and \( x_1 \leq \frac{p}{1 - p} \), the pigeonhole principle gives that for \( v = v_\omega \) (indeed, for any vertex \( v \)),

\[
|N_G(v)| \geq \left\lfloor \frac{p - g}{p} \cdot \frac{1 - p}{g} \right\rfloor
\]

\[
= \begin{cases} \left\lfloor \frac{p - g}{p} \cdot \frac{1 - p}{p/\omega - 1} \right\rfloor, & \text{if } p \leq \frac{\omega - 1}{\omega - 2}; \\ \left\lfloor \frac{p - 1 - p}{p} \cdot \frac{1 - p}{(1 - p)/(\omega - 1)} \right\rfloor, & \text{if } p \geq \frac{\omega - 1}{\omega - 2}; \end{cases}
\]

\[
\geq \left\lfloor (\alpha - 1) \frac{\omega - 2}{\omega - 1} \right\rfloor.
\]

Let \( \alpha = q\omega + r \) with \( 0 \leq r \leq \omega - 1 \) and note that since \( \alpha \geq 2\omega \), \( q \geq 2 \).

Hence,

\[
|N_G(v)| \geq \left\lfloor (\alpha - 1) \frac{\omega - 2}{\omega - 1} \right\rfloor
\]

\[
= \left\lfloor q(\omega - 2) + (g + r - 1)(\omega - 2) \right\rfloor/\omega - 1 \right\rfloor
\]

\[
\geq q(\omega - 2) + 1
\]

\[
= q + \omega - 2 + (q - 1)(\omega - 3)
\]

Since \( q = \lfloor \alpha/\omega \rfloor \), we may conclude that \( |N_G(v)| \geq \lfloor \alpha/\omega \rfloor + \omega - 2 \), just as desired.

Third, let \( \omega = 2 \); i.e., \( H \) is a double-star (possibly with isolated vertices). Recall that \( \alpha \geq 2\omega = 4 \). Our goal is to show that \( |N_G(v_2)| \geq \lfloor \alpha/\omega \rfloor + \omega - 2 \).
$2 = \lfloor \alpha/2 \rfloor$. The computations are, by now, routine. We use the fact that $x_2 \geq d_G(v_1)/(\alpha - 1)$ and $x_1 \leq g/(1-p)$.

$$|N_G(v_2)| \geq \left\lceil \frac{d_G(v_2)}{x_1} \right\rceil \geq \left\lceil \frac{1}{x_1} \left( \frac{p - g}{p} + \frac{1 - 2p}{p} x_2 \right) \right\rceil \geq \left\lceil \frac{1}{x_1} \left( \frac{p - g}{p} + \frac{1 - 2p}{p} \cdot d_G(v_1) \right) \right\rceil \geq \left\lceil \frac{p - g}{px_1} \left( 1 + \frac{1 - 2p}{p(\alpha - 1)} \right) \right\rceil \geq \left\lceil \frac{(p - g)(1 - p)}{pg} \left( \frac{p(\alpha - 3) + 1}{p(\alpha - 1)} \right) + \left( \frac{1 - 2p}{p} \right)^2 \frac{1}{\alpha - 1} \right\rceil.$$ 

Recalling that, in the case of $\omega = 2$, $g \leq \min\{p, (1 - p)/(\alpha - 1)\},$

$$|N_G(v_2)| \geq \begin{cases} \left\lceil \frac{(1 - 2p)^2}{p(\alpha - 1)} \right\rceil, & \text{if } p \leq 1/\alpha; \\ \left\lceil \frac{p(\alpha - 3) + 1}{p(\alpha - 1)} \right\rceil + \left( \frac{1 - 2p}{p} \right)^2 \frac{1}{\alpha - 1}, & \text{if } p \geq 1/\alpha. \end{cases}$$

$$= \begin{cases} \left\lceil \frac{(1 - 2p)^2}{p(\alpha - 1)} \right\rceil, & \text{if } p \leq 1/\alpha; \\ \alpha - 2 - \frac{(1 - 2p)}{p(\alpha - 1)}, & \text{if } p \geq 1/\alpha. \end{cases}$$

In each case, the smallest value of the expression occurs when $p = 1/\alpha$, giving

$$|N_G(v_2)| \geq \left\lceil \alpha - 3 \right\rceil \geq \alpha - 2 = \left\lceil \frac{\alpha}{2} \right\rceil + \left( \left\lceil \frac{\alpha}{2} \right\rceil - 2 \right).$$

This is at least $\lfloor \alpha/2 \rfloor$ since $\alpha \geq 4$. This concludes the proof of Fact 13. $\square$

Summarizing, if $H \not\rightarrow K$, then either $g \geq p/(\omega - 1)$ or $g \geq (1 - p)/(\alpha - 1)$. This concludes the proof of Theorem 1.

5.2. Examples of split graphs. Items [i] and [ii] in Corollary 14 were proven in [7].

Corollary 14. Let $H$ be a graph on $h$ vertices.

(i) If $H \approx K_a + E_b$, then $ed_{\text{Forb}(H)}(p) = \min\left\{ \frac{p}{a-1}, \frac{1-p}{b} \right\}$.

(ii) If $H$ is a star (i.e., $H \approx E_{h-1} \vee K_1$), then $ed_{\text{Forb}(H)}(p) = \min\left\{ p, \frac{1-p}{h-2} \right\}$.

(iii) If $H$ is a double-star (i.e., there are adjacent vertices $u$ and $v$ to which every other vertex is adjacent to exactly one), then $ed_{\text{Forb}(H)}(p) = \min\left\{ p, \frac{1-p}{h-3} \right\}$. 

6. Forb($H_9$)

Marchant and Thomason [11] give the example of $\mathcal{H} = \text{Forb}(C_6^*)$, where $C_6^*$ is a 6-cycle with an additional diagonal edge, such that $ed_H(p)$ is not determined by CRGs with all gray edges. More precisely, they prove that

$$ed_{\text{Forb}(C_6^*)}(p) = \min \left\{ \frac{p}{1+2p}, \frac{1-p}{2} \right\}.$$  

The CRG which corresponds to $g_K(p) = (1-p)/2$ is $K(0,2)$. The CRG, $K$, which has $g_K(p) = p/(1+2p)$ for $p \in [0,1/2]$ consists of three vertices: two black vertices connected via a white edge and a white vertex. The remaining two edges are gray.

The graph $H_9$, shown in Figure 1.2 and cited in [7], generates a hereditary property $\mathcal{H} = \text{Forb}(H_9)$ such that $d_\mathcal{H}$ cannot be determined by CRGs of the form $K(a,c)$. Note that $d_{\text{Forb}(C_6^*)}$ can be determined by such CRGs, but the part of the function for $p \in (0,1/2)$ cannot.

6.1. Proof of Theorem 3. Upper bound. We know that $\chi(H_9) = 4$ so let $K^{(1)} = K(3,0)$ where $g_{K^{(1)}}(p) = p/3$. We also know that $\chi(H_9) = 3$ so let $K^{(4)} = K(0,2)$ where $g_{K^{(4)}}(p) = (1-p)/2$. In [7], another CRG in $\mathcal{H} = \text{Forb}(H_9)$ is given, call it $K^{(2)}$. It consists of 4 white vertices, one black edge and 5 gray edges. It has edit distance function $g_{K^{(2)}}(p) = \min\{p/3, p/(2+2p)\}$.

There is a CRG with a smaller $g$ function. We call it $K^{(3)}$, it consists of 5 white vertices, two disjoint black edges and the remaining 8 edges gray. The function $g_{K^{(3)}}(p)$ can be computed by use of Theorem 10. In the setup of that theorem, $K^{(3)}$ has 3 components. Since the components have $g$ functions either $p$ (for the solitary white vertex) or $\min\{p,1/2\}$ (for each of the other two components), the theorem gives that

$$g_{K^{(3)}}(p)^{-1} = p^{-1} + 2(\min\{p,1/2\})^{-1} = (\min\{p/3, p/(1+4p)\})^{-1}.$$  

It is easy to see that $H_9 \not\rightarrow K^{(1)}$ and $H_9 \not\rightarrow K^{(4)}$. In [7], it was shown that $H_9 \not\rightarrow K^{(3)}$. To finish the upper bound, it remains to show that $H_9 \not\rightarrow K^{(3)}$.

Let $v_0$ be the isolated vertex, $\{v_1,w_1\}$ be a black edge and $\{v_2,w_2\}$ be a black edge.

First, we show that no component of $K^{(3)}$ can have 4 vertices from $H_9$. Since there are no independent sets of size 4 and no induced stars of size 4, the only way to have a component of size 4 is to have an induced copy of $C_4$ in the component consisting of, say, $\{v_2,w_2\}$. It is not difficult to see that deleting two vertices from the set $\{0,3,6\}$ yields a $C_4$-free graph. So, any $C_4$ contains exactly two members of $\{0,3,6\}$. Without loss of generality, the induced $C_4$ is $\{1,3,6,8\}$. But the graph induced by $\{0,2,4,5,7\}$ induces a $C_5$, which cannot be mapped into the sub-CRG induced by $\{v_0,v_1,w_1\}$. Therefore, if $H_9$ were to map to $K^{(3)}$, each component must contain exactly 3 vertices. First we map to $v_0$. The only independent sets of size 3 are
\{1, 4, 7\} and \{2, 5, 8\}. Without loss of generality, assume the former. Second, we consider the graph induced by \{0, 2, 3, 5, 6, 8\}. Any partition of these vertices into two subsets of 3 vertices either has a triangle or a copy of \(P_3\), neither of which maps into \{v_1, w_1\} or \{v_2, w_2\}. So, these six vertices cannot be mapped into \{v_1, w_1, v_2, w_2\}. Hence \(H_9 \not\rightarrow K(3)\).

The CRGs \(K(1)\), \(K(3)\) and \(K(4)\) give an upper bound of \(\min \left\{ \frac{p}{3}, \frac{p}{1 + 4p}, \frac{1 - p}{2} \right\} \).

**Lower bound, for** \(p \leq 1/2\). Let \(K\) be a \(p\)-core such that \(H_9 \not\rightarrow K\). If \(K\) has at least 2 white vertices, then it has no black vertices because \(H_9 \rightarrow K(2, 1)\). (The independent sets are \{1, 4, 7\} and \{2, 5, 8\} and the clique is \{0, 3, 6\}.)

So, in this case \(g_K(p) \geq p/3\) with equality if and only if \(K \approx K(3, 0)\).

If \(K\) has exactly one white vertex, then there is no gray edge among the black vertices because \(H_9 \rightarrow K(1, 2)\). (The independent set is \{2, 7\} and the cliques are \{0, 1, 8\} and \{3, 4, 5, 6\}.) Let \(w\) be the white vertex and \(K' = K - \{w\}\) and \(k' = |V(K')|\). Since \(K'\) is a clique with all black vertices and all white edges, Proposition 8 from [12] gives that, for \(p \in (0, 1/2)\), \(g_{K'}(p) = p + \frac{1 - 2p}{k'} > p\). By Theorem 10, \(g_K(p) > 1/(1/p + 1/p) = p/2\), which is strictly larger than \(ed_{Forb(H_9)}(p)\) for \(p \in (0, 1/2)\).

If \(K\) has no white vertices, then let \(v_0\) be the vertex with largest weight and let \(v_1\) be a vertex in the gray neighborhood of \(v_0\). Let \(x_0 = x(v_0)\) and \(x_1 = x(v_1)\). Since \(K\) can have no gray triangles (\(H_9\) can be partitioned into 3 cliques), \(d_G(v_0) + d_G(v_1) \leq 1\).

\[
1 \geq d_G(v_0) + d_G(v_1) \\
\geq 2p - g + \frac{1 - 2p}{p}(x_0 + x_1) \\
g \geq \frac{p}{2} + \frac{1 - 2p}{2}(x_0 + x_1) \geq \frac{p}{2}.
\]

Equality only occurs if \(p = 1/2\).

Summarizing, if \(p \leq 1/2\) and \(K\) is a \(p\)-core such that \(H \not\rightarrow K\), then \(g_K(p) \geq p/3\) with equality only if \(K \approx K(3, 0)\).

**Lower bound, for** \(p \geq 1/2\). Let \(K\) be a \(p\)-core such that \(H_9 \not\rightarrow K\). If \(K\) has at least 2 black vertices, then there are no white vertices because \(H_9 \rightarrow K(1, 2)\) and so \(g_K(p) \geq (1 - p)/2\) with equality if and only \(K \approx K(0, 2)\).

If \(K\) has exactly one black vertex, then there is no gray edge among the white vertices because \(H_9 \rightarrow K(2, 1)\). Let \(b\) be the black vertex and \(K' = K - \{b\}\) and \(k' = |V(K')|\). Similar to the above, Proposition 8 from [12] can be used to show that, for \(p \in (1/2, 1)\), \(g_{K'}(p) = 1 - p + \frac{2p - 1}{k'} > 1 - p\). By Theorem 10, \(g_K(p) > (1 - p)/2\), which is strictly larger than \(ed_{Forb(H_9)}(p)\) for \(p \in [1/2, 1)\).

From now on, we will assume that \(K\) has only white vertices and, since it is \(p\)-core for \(p \geq 1/2\), all edges are black or gray. Fact 15 and Fact 16 establish some of the structural theorems.
Fact 15. Let \( p \in [1/2, 1) \) and \( K \) be a \( p \)-core CRG with white vertices and black or gray edges. Let \( v \) and \( v' \) be vertices connected by a gray edge. Then, \( N_G(v) \cap N_G(v') \) has at most two vertices.

Proof. If \( N_G(v) \cap N_G(v') \) has three vertices, then map \( H_0 \) vertices 0, 3 and 6 to each of them, map \( \{1,4,7\} \) to \( v \) and \( \{2,5,8\} \) to \( v' \). This is a map demonstrating that \( H_0 \leftrightarrow K \).

Fact 16. Let \( p \in [1/2, 1) \) and \( K \) be a \( p \)-core CRG with white vertices and black or gray edges. Let \( v_0 \) be a vertex of largest weight and \( v_1 \) be a vertex that has largest weight among those in \( N_G(v_0) \). Then, either \( N_G(v_0) \cap N_G(v_1) \) has exactly two vertices or \( g_K(p) > (1 - p)/2 \) or \( g_K(p) \geq p/3 \) with equality if and only if \( K \approx K(3,0) \).

Proof. Let \( g = g_K(p) \). If the statement of Fact 16 is not true, then \( N_G(v_0) \cap N_G(v_1) \) has at most one vertex which, by the choice of \( v_1 \), has weight at most \( x(v_1) \) and, by inclusion-exclusion, has weight at least \( d_G(v_0) + d_G(v_1) - 1 \). Therefore,

\[
x(v_1) \geq d_G(v_0) + d_G(v_1) - 1 \\
\geq 2 \frac{1 - p - g}{1 - p} + \frac{2p - 1}{1 - p} (x(v_0) + x(v_1)) - 1
\]

(8)

\[
g \geq \frac{1 - p}{2} + \frac{2p - 1}{2} x(v_0) - \frac{2 - 3p}{2} x(v_1).
\]

If \( p \geq 2/3 \), then \( g > (1 - p)/2 \). If \( p < 2/3 \), then use \( x(v_1) \leq x(v_0) \) in (8).

(9)

\[
g \geq \frac{1 - p}{2} + \frac{5p - 3}{2} x(v_1)
\]

If \( p \geq 3/5 \), then \( g > (1 - p)/2 \). If \( p < 3/5 \), then use the fact that Corollary 11 gives \( x(v_0) \leq g/p \), which we use in (9).

\[
g \geq \frac{1 - p}{2} + \frac{5p - 3}{2} x(v_1) \geq \frac{1 - p}{2} + \frac{5p - 3}{2} \left( \frac{g}{p} \right)
\]

\[
g \geq \frac{p}{3}.
\]

It is easy to see that equality can only occur if \( x(v_2) = x(v_1) = g/p = 1/3 \) and their common gray neighborhood is a vertex of weight \( 1 - 2g/p = 1/3 \). □

Given Fact 15 and Fact 16, we can identify \( v_0 \), a vertex of maximum weight, \( v_1 \) a vertex of maximum weight among those in \( N_G(v_0) \) and \( \{v_2, w_2\} = N_G(v_0) \cap N_G(v_1) \). Without loss of generality, let \( x(v_2) \geq x(w_2) \). For ease of notation, let \( x_i = x(v_i) \) for \( i = 0, 1, 2 \). If \( N_G(v_0) \cap N_G(v_2) - \{v_1\} \) is nonempty, then let its unique vertex be denoted \( w_1 \). (Uniqueness is a consequence of Fact 15)

Case 1. The vertex \( w_1 \) does not exist.
Most of our observations come from inclusion-exclusion: $|A| + |B| = |A \cup B| + |A \cap B|$. Inequality (10) comes from the fact that $N_G(v_0) \cap N_G(v_1) = \{v_2, w_2\}$. Inequality (11) comes from the fact that $N_G(v_0) \cap N_G(v_2) = \{v_1\}$. Hence,

\begin{align*}
(10) \quad d_G(v_0) + d_G(v_1) &\leq 1 + 2x_2 \\
(11) \quad d_G(v_0) + d_G(v_2) &\leq 1 + x_1.
\end{align*}

Solve for $x_2$ in each case, recalling that Lemma 7(ii) gives that $d_G(v_2) = 1 - p + 2p - 1 - x_2$. Inequality (10) gives a lower bound for $x_2$ and inequality (11) gives an upper bound:

\[
\frac{1}{2} (d_G(v_0) + d_G(v_1) - 1) \leq x_2 \leq \frac{1 - p}{2p - 1} \left( 1 + x_1 - d_G(v_0) - \frac{1 - p - g}{1 - p} \right).
\]

Some simplification gives

\[
2g \geq d_G(v_0) + (2p - 1)d_G(v_1) - 2(1 - p)x_1 - 2p + 1
\geq 2p - \frac{1 - p - g}{1 - p} x_0 + \frac{2p - 1}{1 - p} x_1 - 2p + 1
\geq \frac{1 - p}{2} + \frac{2p - 1}{2} x_0 + \frac{2p - 1}{2} x_1.
\]

If $2p^2 - 1 > 0$ (i.e., $p > 1/\sqrt{2}$), then $g > (1 - p)/2$. Otherwise, we use the bound $x_1 \leq x_0$.

\[
g \geq \frac{1 - p}{2} + \frac{2p - 1}{2} x_0 + \frac{2p - 1}{2} x_0
\geq \frac{1 - p}{2} + (p^2 + p - 1)x_0.
\]

If $p^2 + p - 1 > 0$ (i.e., $p > (\sqrt{5} - 1)/2$), then $g > (1 - p)/2$. Otherwise, we use the bound from Corollary 8(ii) that $x_0 \leq g/p$.

\[
g \geq \frac{1 - p}{2} + (p^2 + p - 1)x_0
\geq \frac{1 - p}{2} + (p^2 + p - 1)\frac{g}{p}
\geq \frac{p}{2(1 + p)}.
\]

Equality occurs only if $x_0 = x_1 = g/p$ and $x(w_2) = x_2 = 1/2 - g/p$. This is precisely the CRG denoted $K^{(2)}$.

**Case 2.** The vertex $w_1$ exists.

Inequality (12) comes from the fact that $N_G(v_0) \cap N_G(v_1) = \{v_2, w_2\}$ and $x(w_2) \leq x(v_2) = x_2$. Inequality (13) comes from the fact that $N_G(v_0) \cap
\[ N_G(v_2) = \{v_1, w_1\} \] and \( x(w_1) \leq x(v_1) = x_1 \). Observe that \( x(w_2) \leq x_2 \) and \( x(w_1) \leq x_1 \), hence,

\begin{align*}
(12) \quad d_G(v_0) + d_G(v_1) & \leq 1 + 2x_2 \\
(13) \quad d_G(v_0) + d_G(v_2) & \leq 1 + 2x_1.
\end{align*}

Adding \((12)\) and \((13)\) gives

\[ 2d_G(v_0) + d_G(v_1) + d_G(v_2) \leq 2 + 2(x_1 + x_2) \quad \text{(14)} \)

If \( p \geq 3/4 \), then \((14)\) gives that \( 2d_G(v_0) - \frac{2g}{1-p} \leq 0 \). Consequently, \( \frac{p-2}{p} - \frac{g}{1-p} < 0 \) and so \( g > \frac{(1-p)/2}{3} \) easily. Thus, we assume \( p < 3/4 \).

Next, we use Fact 17 to conclude that \( v_0 \) is the only common gray neighbor of \( v_1 \) and \( v_2 \).

**Fact 17.** Let \( p \geq 1/2 \) and \( K \) be a \( p \)-core with white vertices and black or gray edges. Let \( a_0, a_1, a_2, b_0, b_1, b_2 \in V(K) \) such that \( \{a_0, a_1, a_2\} \) is a gray triangle and \( \{b_i, a_j\} \) is a gray edge as long as \( i \) and \( j \) are distinct. Then, \( H_0 \rightarrow K \).

**Proof.** The following map shows the embedding:

\[
\begin{align*}
2, 7 & \mapsto a_0 & 1, 5 & \mapsto a_1 & 4, 8 & \mapsto a_2 \\
0 & \mapsto b_0 & 3 & \mapsto b_1 & 6 & \mapsto b_2.
\end{align*}
\]

If \( v_1 \) and \( v_2 \) have a gray neighbor in \( K \) other than \( v_0 \), call it \( w_0 \) and observe that by setting \( a_i := v_i \) and \( b_i := w_i \) for \( i = 0, 1, 2 \), Fact 17 would imply that \( H_0 \rightarrow K \).

Since \( v_0 \) is the only common gray neighbor of \( v_1 \) and \( v_2 \)

\[ d_G(v_1) + d_G(v_2) \leq 1 + x_0 \]

\[ \frac{2p-1}{1-p}(x_1 + x_2) \leq 1 + x_0 - 2\frac{1-p-g}{1-p}. \quad \text{(15)} \]

Inequality \((14)\) gives a lower bound for \( x_1 + x_2 \) and inequality \((15)\) gives an upper bound. Recall that Lemma 7(ii) gives that \( d_G(v) = \frac{1-p-g}{1-p} + 2p-1 \) \( x(v) \) for any vertex \( v \in V(K) \). Recall that we assume \( p < 3/4 \).

\[
\frac{1-p}{3-4p} \left( 2d_G(v_0) - \frac{2g}{1-p} \right) \leq x_1 + x_2 \leq \frac{1-p}{2p-1} \left( 1 + x_0 - 2\frac{1-p-g}{1-p} \right).
\]

Some simplification gives

\[ 2(2p-1)((1-p)d_G(v_0) - g) \leq (3-4p)((1-p)(1+x_0) - 2(1-p-g)) \]

and so

\[ g \geq \frac{1-p}{2} + \frac{4p^2 - p - 1}{2} x_0. \]
If \(4p^2 - p - 1 > 0\) (i.e., \(p > (\sqrt{17} + 1)/8\)), then \(g > (1 - p)/2\). Otherwise, we use the bound \(x_0 \leq g/p\) from Corollary 8(ii).

\[
g \geq \frac{1 - p}{2} + \frac{4p^2 - p - 1}{2} \left(\frac{g}{p}\right)
\]

Equality occurs only if \(x_0 = g/p, x_1 = x_2 = \frac{p}{1+4p}\) and \(x(w_i) = x_i\) for \(i = 1, 2\). This is precisely the CRG denoted \(K^{(3)}\).

Therefore, for \(p \in [1/2, 1]\) and in each case, \(g \geq \min\{p/(1 + 4p), (1 - p)/2\}\). Combining this with the fact that for \(p \in [0, 1]\) that \(g \geq p/3\). This concludes the proof of the lower bound. Consequently, \(ed_{\text{Forb}}(H_0)(p) = \min\{p/3, p/(1 + 4p), (1 - p)/2\}\). This concludes the proof of Theorem 3.

7. Thanks

I would like to thank Maria Axenovich and József Balogh for conversations which have improved the results. Thanks to Andrew Thomason for some useful conversations and for directing me to [11]. Thanks also to Tracy McKay for conversations that helped deepen my understanding and to Doug West for answering my question about clique-stars.

A very special thanks to Ed Marchant for finding an error in the original formulation of Theorem 1 and I am indebted to an anonymous referee whose detailed comments resulted in correcting some errors and provided a much better exposition of the proofs.

Figures are made by Mathematica and WinFIGQT.

References

[1] N. Alon and A. Stav, What is the furthest graph from a hereditary property? Random Structures Algorithms 33 (2008), no. 1, pp. 87–104.
[2] N. Alon and A. Stav, The maximum edit distance from hereditary graph properties. J. Combin. Th. Ser. B 98 (2008), no. 4, pp. 672–697.
[3] N. Alon and A. Stav, Stability type results for hereditary properties. J. Graph Theory 62 (2009), no. 1, 65–83.
[4] N. Alon and A. Stav, Hardness of edge-modification problems. Theoret. Comput. Sci. 410 (2009), no. 47-49, 4920–4927.
[5] M. Axenovich, A. Kézdy and R. Martin, On the editing distance of graphs, J. Graph Theory 58 (2009), no. 2, 123–138.
[6] M. Axenovich and R. Martin, Avoiding patterns in matrices via a small number of changes. SIAM J. Discrete Math. 20 (2006), no. 1, 49–54 (electronic).
[7] J. Balogh and R. Martin, Edit distance and its computation. Electron. J. Combin. 15 (2008), no. 1, Research paper 20, 27pp.
[8] W. Barrett, C. Jepsen, R. Lang, E. McHenry, C. Nelson and K. Owens, Inertia sets for graphs on six or fewer vertices 20 (2010), 53–78.
[9] L.T.Q. Hung, M. Syslo, M. Weaver and D. West, Bandwidth and density for block graphs, Discrete Math. 189 (1998), no. 1-3, 163–176.
[10] E. Marchant, (in preparation).
[11] E. Marchant and A. Thomason, Extremal graphs and multigraphs with two weighted colours, preprint.

[12] R. Martin, Edit distance and localization, submitted, arXiv:1007.1897v3.

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

[14] D.C. Richer, Ph.D. thesis, University of Cambridge (2000).

[15] A.F. Sidorenko, Boundedness of optimal matrices in extremal multigraph and digraph problems, Combinatorica 13 (1993), no. 1, 109–120.

Department of Mathematics, Iowa State University, Ames, Iowa 50011

E-mail address: rymartin@iastate.edu