Discussiones Mathematicae
Graph Theory xx (xxxx) 1–15
doi:10.7151/dmgt.2154

THE EDIT DISTANCE FUNCTION OF SOME GRAPHS

Yumei Hu[a], Yongtang Shi[b]

and

Yarong Wei[1,a]

[a]*School of Mathematics*
Tianjin University, Tianjin 300072, China

[b]*Center of Combinatorics and LPMC*
Nankai University, Tianjin 300071, China

e-mail: huyumei@tju.edu.cn
shi@nankai.edu.cn

Abstract

The edit distance function of a hereditary property $\mathscr{H}$ is the asymptotically largest edit distance between a graph of density $p \in [0,1]$ and $\mathscr{H}$. Denote by P_n and C_n the path graph of order n and the cycle graph of order n, respectively. Let C_{2n}^* be the cycle graph C_{2n} with a diagonal, and $\widetilde{C_n}$ be the graph with vertex set $\{v_0, v_1, \ldots, v_{n-1}\}$ and $E(\widetilde{C_n}) = E(C_n) \cup \{v_0v_2\}$. Marchant and Thomason determined the edit distance function of C_6^*. Peck studied the edit distance function of C_n, while Berikkyzy *et al.* studied the edit distance of powers of cycles. In this paper, by using the methods of Peck and Martin, we determine the edit distance function of C_8^*, $\widetilde{C_n}$ and P_n, respectively.

Keywords: edit distance, colored regularity graphs, hereditary property, clique spectrum.

2010 Mathematics Subject Classification: 05C35.

1. Introduction

The edit distance in graphs was introduced by Axenovich, Kézdy and Martin [5] and by Alon and Stav [4] independently. The edit distance problem considered here is "How many edges need to be added or deleted (edited) in a graph G

[1]Corresponding author.

so that it will have a certain property?" The presence or absence of edges in a certain graph corresponds to pairs of genes which activate or deactivate one another in evolutionary biology. In evolutionary theory, the gene reconstruction avoiding forbidden induced subgraphs is studied [9], which is equivalent to the edit distance problem. The edit distance problem is also important to the algorithmic aspects of property testing [1–4].

The *edit distance* between a graph G and a property $\mathscr{H}$ is

$$dist(G, \mathscr{H}) = \min\left\{|E(G) \triangle E(G')|/\binom{n}{2} : V(G) = V(G'), G' \in \mathscr{H}\right\}.$$

The *edit distance function* of a property $\mathscr{H}$, denoted $ed_{\mathscr{H}}(p)$, measures the maximum distance of a graph with density p from $\mathscr{H}$. Formally,

$$ed_{\mathscr{H}}(p) = \lim_{n\to\infty} \max\left\{dist(G, \mathscr{H}) : |V(G)| = n, |E(G)| = \left\lfloor p\binom{n}{2}\right\rfloor\right\}. \tag{1}$$

if this limit exists.

A *hereditary property* is a family of graphs that is closed under the taking of induced subgraphs. For a given graph H, the property of having no H as an induced subgraph is called a *principal hereditary property*, denoted by $Forb(H)$. Clearly, $Forb(H)$ is a hereditary property for any graph H. In fact, for every hereditary property $\mathscr{H}$ there exists a family of graphs $\mathscr{F}(\mathscr{H})$ such that $\mathscr{H} = \bigcap_{H\in\mathscr{F}(\mathscr{H})} Forb(H)$. A hereditary property is said to be *nontrivial* if there is an infinite sequence of graphs that is in the property. The properties for which we study the edit distance are usually hereditary property.

Balogh and Martin [6] showed that the limit in (1) exists and the edit distance function has a number of interesting properties.

Proposition 1 [11]**.** *Let $\mathscr{H}$ be a nontrivial hereditary property. For $p \in [0, 1]$,*

(a) *$ed_{\mathscr{H}}(p)$ is continuous.*
(b) *$ed_{\mathscr{H}}(p)$ is concave down.*

In [4], Alon and Stav proved that for every hereditary property $\mathscr{H}$, there exists a $p^* = p^*(\mathscr{H}) \in [0, 1]$ such that the maximum distance of a graph G on n vertices from $\mathscr{H}$ is asymptotically the same as that of the Erdös-Rényi random graph $G(n, p^*)$. Namely,

$$\max\{dist(G, \mathscr{H}) : |V(G)| = n\} = \mathbb{E}[dist(G(n, p^*), \mathscr{H})] + o(1). \tag{2}$$

We denote the limit in (2) by $d^*_{\mathscr{H}}$.

The edit distance functions of some kinds of graphs have been investigated in recent years, including complete graphs [13] and split graphs [12]. Actually, complete bipartite graphs are also studied. Marchant and Thomason [10] studied the edit distance functions of $K_{2,2}$ and $K_{3,3}$, respectively. Balogh and Martin [6] established the value of $p^*_{Forb(K_{3,3})}$ and $d^*_{Forb(K_{3,3})}$. Martin and McKay studied the edit distance function of $K_{2,t}$ in [14]. Recently, Berikkyzy *et al.* [7] settled the edit distance function for many powers of cycles.

Denote by P_n and C_n the path graph of order n and the cycle graph of order n, respectively. Let C^*_{2n} be the cycle graph C_{2n} with a diagonal, and $\widetilde{C_n}$ be the graph with vertex set $\{v_0, v_1, \ldots, v_{n-1}\}$ and $E(\widetilde{C_n}) = E(C_n) \cup \{v_0v_2\}$.

In [10], Marchant and Thomason studied the edit distance function of the graph C^*_6. Motivated by this result, we study the edit distance function of the graph C^*_8 and prove the following result.

Theorem 2. *Let $\mathscr{H} = Forb(C^*_8)$.*

$$ed_{\mathscr{H}}(p) = \min\left\{\frac{p}{2}, \frac{p(1-p)}{1+p}, \frac{1-p}{3}\right\}, \text{ for } p \in [0,1].$$

Peck [15] in her Master's thesis calculated the edit distance function of C_n. The result is as follows.

Theorem 3 [15]**.** *Let $\mathscr{H} = Forb(C_n)$.*

(a) *If n is odd, then* $ed_{\mathscr{H}}(p) = \min\left\{\frac{p}{2}, \frac{p(1-p)}{1+(\lceil\frac{n}{3}\rceil-2)p}, \frac{1-p}{\lceil\frac{n}{2}\rceil-1}\right\}$, *for* $p \in [0,1]$.

(b) *If n is even, then* $ed_{\mathscr{H}}(p) = \min\left\{\frac{p(1-p)}{1+(\lceil\frac{n}{3}\rceil-2)p}, \frac{1-p}{\lceil\frac{n}{2}\rceil-1}\right\}$, *for* $p \in \left[\lceil n/3\rceil^{-1}, 1\right]$.

Motivated by this result, we study the edit distance function of $\widetilde{C_n}$ and P_n.

Theorem 4. *Let $\mathscr{H} = Forb(\widetilde{C_n})$ and $n \geq 9$.*

$$ed_{\mathscr{H}}(p) = \min\left\{\frac{p}{2}, \frac{p(1-p)}{1+\left(\lceil\frac{n-1}{3}\rceil-2\right)p}, \frac{1-p}{\lceil\frac{n-3}{2}\rceil}\right\}, \text{ for } p \in [0,1].$$

Theorem 5. *Let $\mathscr{H} = Forb(P_n)$ and $n \geq 3$.*

$$ed_{\mathscr{H}}(p) = \min\left\{\frac{p(1-p)}{1+\left(\lceil\frac{n-1}{3}\rceil-2\right)p}, \frac{1-p}{\lceil\frac{n}{2}\rceil-1}\right\}, \text{ for } p \in \left[\lceil(n-1)/3\rceil^{-1}, 1\right].$$

Our paper is organized as follows. Some definitions and tools are explained in Section 2. We prove Theorems 2, 4 and 5 in Sections 3, 4 and 5, respectively.

2. Definitions and Tools

All graphs considered in this paper are simple. The standard graph theory notation not defined here will conform to that in [8]. The edit distance notation not defined here will conform to that in [11].

In order to estimate the edit distance function, Alon and Stav [4] defined a *colored regularity graph* (CRG) K as follows. Let K be a simple complete graph, together with a partition of the vertices into white and black, and a partition of the edges into white, gray, and black. Denote by $VW(K)$ and $VB(K)$ the set of white vertices and the set of black vertices, respectively. Then $V(K) = VW(K) \cup VB(K)$. Denote by $EW(K)$, $EG(K)$ and $EB(K)$ the set of white edges, the set of gray edges and the set of black edges, respectively. Then we have $E(K) = EW(K) \cup EG(K) \cup EB(K)$. A CRG K' is said to be a *sub*-CRG of K if K' can be obtained by deleting vertices of K and is a *proper sub*-CRG if $K' \neq K$.

We say that a graph H *embeds in* K (writing $H \mapsto K$), if there is a function $\varphi : V(H) \to V(K)$ so that if $h_1h_2 \in E(H)$, then either $\varphi(h_1) = \varphi(h_2) \in VB(K)$ 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 $\mathscr{H}$, we denote by $\mathscr{K}(\mathscr{H})$ the subset of CRGs K such that any graph $H \in \mathscr{F}(\mathscr{H})$ does not embed in K. That is, $\mathscr{K}(\mathscr{H}) = \{K : H \not\mapsto K, \forall H \in \mathscr{F}(\mathscr{H})\}$.

For a hereditary property $\mathscr{H}$, we can use the g function of each CRG K to compute the edit distance function, where g function is defined by

$$g_K(p) = \min\left\{\mathbf{x}^T M_K(p)\mathbf{x} : \mathbf{x}^T\mathbf{1} = 1, \mathbf{x} \geq \mathbf{0}\right\}, \tag{3}$$

and

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

Marchant and Thomason in [10] proved that for every $p \in [0,1]$, there is a CRG $K \in \mathscr{K}(\mathscr{H})$ such that $ed_{\mathscr{H}}(p) = g_K(p)$. That is

Proposition 6 [10]**.** *Let $\mathscr{H}$ be a nontrivial hereditary property. For $p \in [0,1]$,*

$$ed_{\mathscr{H}}(p) = \min\{g_K(p) : K \in \mathscr{K}(\mathscr{H})\}.$$

In [10], the authors also proved that in order to find such CRGs, we only need to look at all p-core CRGs. A CRG K is p-core if, for any proper sub-CRG K' of K, we have $g_{K'}(p) > g_K(p)$.

The *gray-edge* CRG $K(r,s)$ is the CRG K with r white vertices, s black vertices and all edges gray. The *clique spectrum* of $\mathscr{H}$ is the set $\Gamma(\mathscr{H}) := \{(r,s) : H \not\mapsto K(r,s),\ \forall H \in \mathscr{F}(\mathscr{H})\}$. Clearly, we obtain

Proposition 7 [11]**.** *Let $\mathscr{H}$ be a nontrivial hereditary property and $\Gamma(\mathscr{H})$ denote the clique spectrum of $\mathscr{H}$. If we define*

$$\gamma_{\mathscr{H}}(p) := \min_{(r,s)\in\Gamma(\mathscr{H})} g_{K(r,s)}(p) = \min_{(r,s)\in\Gamma(\mathscr{H})} \frac{p(1-p)}{r(1-p)+sp},$$

then $ed_{\mathscr{H}}(p) \leq \gamma_{\mathscr{H}}(p)$.

Let K be a p-core CRG, $v \in V(K)$, and let $\mathbf{x}$ be an optimal weight vector in the quadratic program (3) that defines $g_K(p)$. The *weight* of v, denoted by $\mathbf{x}(v)$, is the entry corresponding to v of the vector $\mathbf{x}$. We denote the *gray neighborhood* of v by $N_G(v) = \{v' \in V(K) : vv' \in EG(K)\}$. The *weighted gray degree* of vertex $v \in V(K)$ is $d_G(v) = \Sigma_{v'\in N_G(v)}\mathbf{x}(v')$ and the number of vertices adjacent to v via gray edges is denoted by $\deg_G(v)$, i.e., $\deg_G(v) = |N_G(v)|$. We use similar notation for the white and black cases. Now we get $d_G(v) + d_W(v) + d_B(v) = 1$ for each $v \in V(K)$.

The *weighted gray codegree* of vertices v and v', denoted by $d_G(v, v')$, is the sum of the weights of the common gray neighbors of v and v'. Denote the number of common gray neighbors of vertices v and v' by $\deg_G(v, v')$.

Marchant and Thomason [10] gave the following characterization of all p-core CRGs.

Proposition 8 [10]**.** *Let K be a p-core CRG.*

(a) *If $p \leq 1/2$, then there are no black edges, and the white edges are only incident to black vertices.*
(b) *If $p \geq 1/2$, then there are no white edges, and the black edges are only incident to white vertices.*

Martin [13] gave a formula for $d_G(v)$ for all $v \in V(K)$ and a bound on the weight of each v.

Proposition 9 [13]**.** *Let $p \in (0, 1)$ and K be a p-core CRG with optimum weight vector $\mathbf{x}$.*

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

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

(b) *If $p \geq 1/2$, then $\mathbf{x}(u) = g_K(p)/(1-p)$ for all $u \in VB(K)$ and*

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

The following results will be used in this paper.

Proposition 10 [13]**.** *Let $p \in (0, 1/2)$ and K be a p-core CRG with black vertices and white or gray edges.*

(a) *If K has no gray 3-cycle, then $g_K(p) > p/2$.*

(b) *If K has a gray 3-cycle, but no gray C_4^+ (that is, four vertices that induce 5 gray edges), then $g_K(p) \geq \min\{2p/3, (1-p)/3\}$.*

Proposition 11 [7]**.** *Let F be a connected graph. If some path of maximum length forms a cycle, then F is Hamiltonian.*

Proposition 12 [7]**.** *Let F be a graph on n vertices with no cycle of length longer than $\lceil \frac{n}{2} \rceil - 1$, with every vertex having degree at least $\lceil \frac{n-1}{3} \rceil \geq 2$ and with every pair of vertices having at least one common neighbor. Furthermore, let F have the property that no maximum length path forms a cycle.*

Let $v_1 \cdots v_\ell$ be a path of maximum length in F. Then v_1 and v_ℓ have exactly one common neighbor v_c on this path. Furthermore, $N(v_1) \subseteq \{v_2, \ldots, v_c\}$ and $N(v_\ell) \subseteq \{v_c, \ldots, v_{\ell-1}\}$.

3. Proof of Theorem 2

In this section, we consider the edit distance function for the hereditary property that forbids C_{2n}^* where n is even and prove that $ed_{Forb(C_8^*)}(p) = \gamma_{Forb(C_8^*)}(p)$ for all $p \in [0, 1]$.

First, we obtain the value of $\gamma_{Forb(C_{2n}^*)}(p)$ for $p \in [0, 1]$ and restrict $ed_{Forb(C_{2n}^*)}(p)$ to $p \in [0, 1/2)$ and CRGs K with only black vertices. Finally, we determine the edit distance $ed_{Forb(C_8^*)}(p) = \gamma_{Forb(C_8^*)}(p)$ and then prove Theorem 2.

Lemma 13. *Let $\mathscr{H} = Forb(C_{2n}^*)$, $p \in [0, 1]$ and $n \geq 4$ be even.*

$$\gamma_{\mathscr{H}}(p) = \min\left\{\frac{p}{2},\ \frac{p(1-p)}{1 + \left(\lceil \frac{2n-1}{3} \rceil - 2\right)p},\ \frac{1-p}{n-1}\right\}.$$

Furthermore, if there is a p-core CRG $K \in K(Forb(C_{2n}^*))$ such that $g_K(p) < \gamma_{Forb(C_{2n}^*)}(p)$ for any $p \in [0, 1]$, then $p < 1/2$ and K has all black vertices.

Proof. If n is even, the extreme points of the clique spectrum of $Forb(C_{2n}^*)$ are $(2, 0)$, $\left(1, \lceil \frac{2n-1}{3} \rceil - 1\right)$ and $(0, n-1)$. Then $\gamma_{\mathscr{H}}(p) = \min\left\{\frac{p}{2}, \frac{p(1-p)}{1+\left(\lceil \frac{2n-1}{3} \rceil - 2\right)p}, \frac{1-p}{n-1}\right\}$.

Since $ed_{\mathscr{H}}(1/2) = \gamma_{\mathscr{H}}(1/2)$ for any hereditary property and $\gamma_{\mathscr{H}}(1) = 0$, we may use continuity and concavity to conclude that $ed_{\mathscr{H}}(p) = \gamma_{\mathscr{H}}(p) = \frac{1-p}{n-1}$ for $p \in [1/2, 1]$. Now we suppose $p \in [0, 1/2)$ and K is a p-core CRG such that $g_K(p) < \gamma_{\mathscr{H}}(p)$.

If K has only white vertices, then $|V(K)| \leq 2$ and $g_K(p) \geq \frac{p}{2} \geq \gamma_{\mathscr{H}}(p)$ since $C^*_{2n} \mapsto K(3,0)$. If K has both white and black vertices, then it has 1 white vertex ω since $C^*_{2n} \mapsto K(2,1)$. Furthermore, it can have at most $\lceil \frac{2n-1}{3} \rceil - 1$ black vertices.

To see this, denote the vertices of C^*_{2n} by $\{0,\ldots,2n-1\}$ where $i \sim i+1$ for $0 \leq i \leq 2n-2$, $2n-1 \sim 0$ and $0 \sim n$. If n is not divisible by 3, then let S consist of the members of $\{0,1,\ldots,2n-1\}$ that are divisible by 3. The graph $C^*_{2n} - S$ has $\lceil \frac{2n-1}{3} \rceil$ connected components, each of which are cliques of size 1 or 2. If n is divisible by 3, then let $S = \{i : i \in \{0,1,\ldots,2n-1\}, i-1 \text{ is divisible by } 3\}$. The graph $C^*_{2n} - S$ has $\lceil \frac{2n-1}{3} \rceil$ connected components, each of which are cliques of size 2 except three edges $n-1 \sim n$, $n \sim 0$ and $0 \sim 2n-1$.

If $d_G(v_i) = \mathbf{x}(\omega)$ for any $v_i \in VB(K)$, then by Proposition 9(a), we have $\frac{g_K(p)}{p} = \frac{p-g_K(p)}{p} + \frac{1-2p}{p}\mathbf{x}(v_i) > \frac{p-g_K(p)}{p}$. Rearranging the terms, we obtain $g_K(p) > \frac{p}{2} \geq \gamma_{\mathscr{H}}(p)$, a contradiction. So, there are two black vertices v_1, v_2 in K such that $v_1v_2 \in EG(K)$. Let v_1 receive $n-1 \sim n$ and v_2 receive $0 \sim 2n-1$. Then $C^*_{2n} \mapsto K$. Thus, regardless of whether the edges are white or gray, there are at most $\lceil \frac{2n-1}{3} \rceil - 1$ black vertices in K and $g_K(p) \geq \frac{p(1-p)}{1+(\lceil \frac{2n-1}{3} \rceil - 2)p} \geq \gamma_{\mathscr{H}}(p)$.

So, if $p \in [0,1/2)$ and $g_K(p) = ed_{Forb(C^*_{2n})}(p)$, then K is either $K(2,0)$, $K\left(1, \lceil \frac{2n-1}{3} \rceil - 1\right)$, $K(0,n-1)$ or K has all black vertices (and white or gray edges). ■

Proof of Theorem 2. Now, we calculate $ed_{\mathscr{H}}(p)$ where $\mathscr{H} = Forb(C^*_8)$. By Lemma 13, we know $\gamma_{\mathscr{H}}(p) = \min\left\{\frac{p}{2}, \frac{1-p}{3}, \frac{p(1-p)}{1+p}\right\}$ and only need to consider the p-core CRGs K with only black vertices for some $p \in [0,1/2)$.

If K has only black vertices, then K has no gray C^+_4 otherwise $C^*_8 \mapsto K$. By Proposition 10, we know either $g_K(p) > p/2 \geq \gamma_{\mathscr{H}}(p)$ or $g_K(p) \geq \min\{2p/3, (1-p)/3\} > \gamma_{\mathscr{H}}(p)$. By straightforward calculations, this contradicts to $g_K(p) < \gamma_{\mathscr{H}}(p)$ for all $p \in [0,1/2)$. ■

4. PROOF OF THEOREM 4

In this section, we consider the edit distance function for hereditary property that forbids $\widetilde{C_n}$. Let $\mathscr{H} = Forb(\widetilde{C_n})$. First, we obtain the value of $\gamma_{\mathscr{H}}(p)$ for $p \in [0,1]$. Then we suppose there is a p-core CRG $K \in \mathscr{K}(Forb(\widetilde{C_n}))$ such that $g_K(p) < \gamma_{\mathscr{H}}(p)$ and establish some characterizations of such a p-core CRG K. Finally, we obtain a contradiction to such a CRG existing in $\mathscr{K}(Forb(\widetilde{C_n}))$ for our desired range of p values, establishing $\gamma_{\mathscr{H}}(p) \leq ed_{\mathscr{H}}(p)$.

Lemma 14. *Let $\mathscr{H} = Forb(\widetilde{C_n})$, and $n \geq 6$. Then*

$$\gamma_{\mathscr{H}}(p) = \min\left\{\frac{p}{2}, \frac{p(1-p)}{1+\left(\lceil\frac{n-1}{3}\rceil - 2\right)p}, \frac{1-p}{\lceil\frac{n-3}{2}\rceil}\right\}, \text{ for } p \in [0,1].$$

Furthermore, if there is a p-core CRG $K \in \mathscr{K}(\mathscr{H})$ such that $g_K(p) < \gamma_{\mathscr{H}}(p)$ for any $p \in [0,1]$, then $p < \frac{1}{2}$ and K has all black vertices.

Proof. The extreme points of the clique spectrum of $Forb(\widetilde{C_n})$ are $(2,0)$, $\left(1, \lceil\frac{n-1}{3}\rceil - 1\right)$ and $\left(0, \lceil\frac{n-3}{2}\rceil\right)$, which establishes the value of $\gamma_{\mathscr{H}}(p)$.

Since $ed_{\mathscr{H}}(1/2) = \gamma_{\mathscr{H}}(1/2)$ for any hereditary property and $\gamma_{\mathscr{H}}(1) = 0$, we may use continuity and concavity to conclude that $ed_{\mathscr{H}}(p) = \frac{1-p}{\lceil\frac{n-3}{2}\rceil}$ for $p \in [1/2, 1]$.

Now, let $p \in [0, 1/2)$ and K be a p-core CRG such that $\widetilde{C_n} \not\mapsto K$. If K has at most two vertices, then $g_K(p) \geq \frac{p}{2}$ since $\widetilde{C_n} \mapsto K(3,0)$. If K has both white and black vertices, then it has at most one white vertex since $\widetilde{C_n} \mapsto K(2,1)$. Furthermore, it can have at most $\lceil\frac{n-1}{3}\rceil - 1$ black vertices.

To see this, denote the vertices of $\widetilde{C_n}$ by $\{0, 1, \ldots, n-1\}$ where $i \sim i+1$ for $0 \leq i \leq n-2$, $n-1 \sim 0$ and $0 \sim 2$. Let S consist of the members of $\{3, \ldots, n-1\}$ that are divisible by 3. If $n-1$ is not divisible by 3, then add 0 to S. The graph $\widetilde{C_n} - S$ has $\lceil\frac{n-1}{3}\rceil$ connected components, each of which are cliques of size 1 or 2 or 3. Thus, regardless of whether the edges are white or gray, there are at most $\lceil\frac{n-1}{3}\rceil - 1$ black vertices in K and $g_K(p) \geq \frac{p(1-p)}{1+(\lceil\frac{n-1}{3}\rceil - 2)p}$, with equality if and only if $K \cong K\left(1, \lceil\frac{n-1}{3}\rceil - 1\right)$.

Summarizing, if $p \in [0, 1/2)$ and $g_K(p) = ed_{\mathscr{H}}(p)$, then K is either $K(2,0)$, $K\left(1, \lceil\frac{n-1}{3}\rceil - 1\right)$, $K\left(0, \lceil\frac{n-3}{2}\rceil\right)$, or K has all black vertices (and white or gray edges). ■

We only need to consider the $K \in \mathscr{K}(Forb(\widetilde{C_n}))$ with all black vertices such that $g_K(p) < \gamma_{Forb(\widetilde{C_n})}(p)$. Now, we establish some characterizations of such a p-core CRG K.

Proposition 15. *Let $p \in [0, 1/2)$ and K be a p-core CRG such that K has only black vertices and white and gray edges. If $\widetilde{C_n} \not\mapsto K$ then K has no gray cycle of length $l \in \left\{\lceil\frac{n-1}{2}\rceil, \ldots, n-1\right\}$.*

Proof. Suppose K has some gray cycle of length $l \in \left\{\lceil\frac{n-1}{2}\rceil, \ldots, n-1\right\}$. Partition the vertices of $\widetilde{C_n}$ into l parts so that one part is the triangle and each of the others parts is either a set of two consecutive vertices (an edge) or single vertex. Because of the structure of $\widetilde{C_n}$ and the fact that $\lceil\frac{n-1}{2}\rceil \leq l \leq n-1$, it is always possible to do so. This partition witnesses an embedding of $\widetilde{C_n}$ into the l-cycle of K because we can map consecutive parts to consecutive vertices on the l-cycle. Since non-consecutive parts do not have edges between them and Proposition 8(a) gives that the edges of K are either white or gray, this map is an embedding that demonstrates $\widetilde{C_n} \mapsto K$, a contradiction. ■

Proposition 16. *Let* $p \in \left[\frac{1}{\lceil\frac{n-1}{3}\rceil}, \frac{1}{2}\right)$, *and K be a p-core CRG with all black vertices such that* $g_K(p) < \gamma_{Forb(\widetilde{C_n})(p)}$. *Then*

(a) *for every* $v \in V(K)$, $\deg_G(v) \geq \lceil\frac{n-1}{3}\rceil$, *and*
(b) *for every* $v, w \in V(K)$, $\deg_G(v, w) \geq 1$.

Proof. (a) Let $v, w \in V(K)$. By using Proposition 9(a),

$$\begin{aligned}
\deg_G(v) &\geq \left\lceil \frac{d_G(v)}{\max\{\mathbf{x}(w)\}} \right\rceil \geq \frac{\frac{p-g_K(p)}{p} + \frac{1-2p}{p}\mathbf{x}(v)}{\frac{g_K(p)}{1-p}} \\
&\geq \frac{(p-g_K(p))(1-p)}{pg_K(p)} = \frac{1-p}{g_K(p)} - \frac{1-p}{p} \\
&> \frac{(1-p) + \left(\lceil\frac{n-1}{3}\rceil - 1\right)p}{p} - \frac{1-p}{p} = \left\lceil \frac{n-1}{3} \right\rceil - 1.
\end{aligned}$$

(b) By the inclusion-exclusion principle, $d_G(v) + d_G(w) - d_G(v, w) \leq 1$, and by using Proposition 9(a), we have $d_G(v, w) \geq 2\frac{p-g_K(p)}{p} + \frac{1-2p}{p}(\mathbf{x}(v) + \mathbf{x}(w)) - 1 \geq \frac{p-g_K(p)}{p} \geq \frac{p-2g_K(p)}{p}$ and for all $u \in V(K)$, $\mathbf{x}(u) \leq g_K(p)/(1-p)$. Therefore,

$$\begin{aligned}
\deg_G(v, w) &\geq \left\lceil \frac{d_G(v, w)}{\max\{\mathbf{x}(u)\}} \right\rceil \geq \left\lceil \frac{\frac{p-2g_K(p)}{p}}{\frac{g_K(p)}{1-p}} \right\rceil = \frac{1-p}{g_K(p)} - \frac{2(1-p)}{p} \\
&> \frac{(1-p) + \left(\lceil\frac{n-1}{3}\rceil - 1\right)p}{p} - \frac{2(1-p)}{p} = \left\lceil \frac{n-1}{3} \right\rceil - \frac{1}{p}.
\end{aligned}$$

Since $p \geq \frac{1}{\lceil\frac{n-1}{3}\rceil}$, we have $\deg_G(v, w) \geq 1$. ■

We consider the value of $ed_{Forb(\widetilde{C_n})}(p)$ from the perspective of the gray subgraphs of CRGs K. Let F be a graph such that $V(F) = V(K)$ and $E(F) = EG(K)$, where $K \in \mathscr{K}(Forb(\widetilde{C_n}))$ is a p-core CRG with all black vertices such that $g_K(p) < \gamma_{Forb(\widetilde{C_n})}(p)$. By Proposition 16, F is a connected graph and each pair of vertices has at least one common neighbor.

Proposition 17. *Let $n \geq 9$ and F be a graph with no cycle with length in* $\left\{\lceil\frac{n-1}{2}\rceil, \ldots, n-1\right\}$ *and every pair of vertices having at least one common neighbor. Then F has no cycle of with length greater than* $\lceil\frac{n-1}{2}\rceil - 1$.

Proof. Let $v_1 \cdots v_\ell v_1$ be a shortest cycle in F among all those with length greater than $n-1$. Consider the path $v_1 \cdots v_{\lceil \frac{n-1}{2} \rceil -1}$ on the cycle $v_1 \cdots v_\ell v_1$.

Assume v_i is a common neighbor of v_1 and $v_{\lceil \frac{n-1}{2} \rceil -1}$, then either $v_1 v_i v_{i+1} \cdots v_\ell v_1$ or $v_1 \cdots v_i v_{\lceil \frac{n-1}{2} \rceil -1} \cdots v_\ell v_1$ has length less than ℓ. Without loss of generality, we assume $v_1 v_i v_{i+1} \cdots v_\ell v_1$ has length less than ℓ, which implies

$$\left\lceil \frac{n-1}{2} \right\rceil - 1 \geq \ell - i + 2 \geq \ell - \left(\left\lceil \frac{n-1}{2} \right\rceil - 2 \right) + 2 \geq n - \left\lceil \frac{n-1}{2} \right\rceil + 4.$$

Thus,

$$2 \left\lceil \frac{n-1}{2} \right\rceil - 1 - n - 4 \geq 0,$$

a contradiction, since $2 \left\lceil \frac{n-1}{2} \right\rceil - 1 - n - 4 < 2 \left(\frac{n-1}{2} + 1 \right) - 1 - n - 4 < 0$.

Therefore, F has no cycle of with length greater than $\left\lceil \frac{n-1}{2} \right\rceil - 1$. ■

Then, we consider the maximum-length path in the graph F. If this path forms a cycle, then Proposition 11 gives that F must be Hamiltonian. By Proposition 17, $|V(K)| \leq \left\lceil \frac{n-1}{2} \right\rceil - 1$ and $g_K(p) \geq \frac{1-p}{\lceil \frac{n-1}{2} \rceil - 1}$, a contradiction. Thus, no maximum-length path in F forms a cycle. By Proposition 17, F has no cycle of with length greater than $\left\lceil \frac{n-1}{2} \right\rceil - 1$. And, by Proposition 16, every vertex in F has degree at least $\left\lceil \frac{n-1}{3} \right\rceil \geq 2$ and every pair of vertices has at least one common neighbor.

Let $v_1 \cdots v_\ell$ be a maximum-length path in F such that the sum $\mathbf{x}(v_1) + \mathbf{x}(v_\ell)$ is largest among all such paths. Then by Proposition 12, we have v_1 and v_ℓ have a unique common neighbor v_c and $N(v_1) \subseteq \{v_2, \ldots, v_c\}$. Let v_1 have d neighbors in F. Since v_1 cannot have neighbors outside of this path, $d_G(v_1) \leq \mathbf{x}(v_2) + \cdots + \mathbf{x}(v_c)$. And if $v_i \in \{v_1, \ldots, v_{c-1}\}$ is a predecessor of a neighbor of v_1 in F, then it is an endpoint of a path containing the same ℓ vertices, namely $v_i v_{i-1} \cdots v_1 v_{i+1} v_{i+2} \cdots v_c \cdots v_\ell$. Hence all d predecessors of gray neighbors of v_1 (including v_1 itself) have weight at most $\mathbf{x}(v_1)$. By Proposition 9, $\frac{p - g_K(p)}{p} + \frac{1-p}{p}\mathbf{x}(v_1) = \mathbf{x}(v_1) + d_G(v_1) \leq \mathbf{x}(v_1) + \cdots + \mathbf{x}(v_c) \leq d\mathbf{x}(v_1) + (c-d)\frac{g}{1-p}$, which implies

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

By Propositions 15 and 16, we have $c \leq \left\lceil \frac{n-1}{2} \right\rceil - 1$ and $d > \left\lceil \frac{n-1}{3} \right\rceil - 1$. So when $p \geq \left\lceil \frac{n-1}{3} \right\rceil^{-1}$, by Proposition 9(a), we have $\mathbf{x}(v) \leq g_K(p)/(1-p)$, hence

$$g_K(p) \geq \frac{1-p}{c} \geq \frac{1-p}{\left\lceil \frac{n-1}{2} \right\rceil - 1} \geq \gamma_{\mathscr{H}}(p),$$

a contradiction. So $ed_{\mathscr{H}}(p) = \gamma_{\mathscr{H}}(p)$ for all $p \in \left[\frac{1}{\lceil \frac{n-1}{3} \rceil}, \frac{1}{2} \right)$.

Finally, $ed_{\mathscr{H}}(p) = \gamma_{\mathscr{H}}(p) = \frac{p}{2}$ for $p = \frac{1}{\lceil\frac{n-1}{3}\rceil}$, and $ed_{\mathscr{H}}(p) = \gamma_{\mathscr{H}}(p) = \frac{p}{2}$ for $p = 0$. Then, since the function $\gamma_{\mathscr{H}}(p)$ is linear over this interval and $ed_{\mathscr{H}}(p)$ is continuous and concave down, we have $ed_{\mathscr{H}}(p) = \gamma_{\mathscr{H}}(p)$ for $p \in \left[0, \frac{1}{\lceil\frac{n-1}{3}\rceil}\right]$. Hence the two functions are equal for all $p \in [0,1]$.

5. Proof of Theorem 5

Similarly as Section 4, but it also involves some crucial differences. We first prove the following lemma.

Lemma 18. *Let $\mathscr{H} = Forb(P_n)$ where P_n denotes the path on $n \geq 3$ vertices.*

$$\gamma_{\mathscr{H}}(p) = \min\left\{\frac{p(1-p)}{1+\left(\lceil\frac{n-1}{3}\rceil - 2\right)p}, \frac{1-p}{\lceil\frac{n}{2}\rceil - 1}\right\}, \quad \textit{for} \quad p \in [0,1].$$

Furthermore, if there is a p-core CRG $K \in \mathscr{K}(\mathscr{H})$ such that $g_K(p) < \gamma_{\mathscr{H}}(p)$ for any $p \in (0,1)$, then $p < \frac{1}{2}$ and K has all black vertices.

Proof. The extreme points of the clique spectrum of $Forb(P_n)$ are $\left(1, \lceil\frac{n-1}{3}\rceil - 1\right)$ and $\left(0, \lceil\frac{n}{2}\rceil - 1\right)$, which establishes the value of $\gamma_{\mathscr{H}}(p)$.

Since $ed_{\mathscr{H}}(1/2) = \gamma_{\mathscr{H}}(1/2)$ for any hereditary property and $\gamma_{\mathscr{H}}(1) = 0$, we may use continuity and concavity to conclude that $ed_{\mathscr{H}}(p) = \frac{1-p}{\lceil\frac{n}{2}\rceil - 1}$ for $p \in [1/2, 1]$.

Now, let $p \in [0, 1/2)$ and K be a p-core CRG such that $P_n \not\mapsto K$. If K has only white vertices, then $K \approx K(1,0)$ and $g_K(p) = p > \gamma_{\mathscr{H}}(p)$. If K has both white and black vertices, then it has at most one white vertex since $P_n \mapsto K(2,1)$. Furthermore, it can have at most $\lceil\frac{n-1}{3}\rceil - 1$ black vertices. To see this, denote the vertices of P_n by $\{0, 1, \ldots, n-1\}$ where $0 \sim 1 \sim 2 \sim \cdots \sim n-1$. Let S consist of the members of $\{0, 1, \ldots, n-1\}$ that are divisible by 3. The graph $P_n - S$ has $\lceil\frac{n-1}{3}\rceil$ connected components, each of which are cliques of size 1 or 2. Thus, regardless of whether the edges are white or gray, there are at most $\lceil\frac{n-1}{3}\rceil - 1$ black vertices in K and $g_K(p) \geq \frac{p(1-p)}{1+\left(\lceil\frac{n-1}{3}\rceil - 2\right)p}$, with equality if and only if $K \approx K\left(1, \lceil\frac{n-1}{3}\rceil - 1\right)$.

Summarizing, if $p \in [0, 1/2)$ and $g_K(p) = ed_{\mathscr{H}}(p)$, then K is either $K\left(1, \lceil\frac{n-1}{3}\rceil - 1\right)$, $K\left(0, \lceil\frac{n}{2}\rceil - 1\right)$ or K has all black vertices (and white or gray edges). ∎

When $n < 5$, $\gamma_{\mathscr{H}}(p) = \min\{p, 1-p\}$. This observation plus continuity and concavity give that $ed_{\mathscr{H}}(p) = \gamma_{\mathscr{H}}(p)$ for all $p \in [0,1]$. From now on, we assume $n \geq 5$.

We only need to consider the $K \in \mathscr{K}(Forb(P_n))$ with all black vertices such that $g_K(p) < \gamma_{Forb(P_n)}(p)$. Now, we establish some characterizations of such a p-core CRG K.

Proposition 19. *Let $p \in \left[\frac{1}{\lceil \frac{n-1}{3} \rceil}, \frac{1}{2}\right)$, and K be a p-core CRG with all black vertices such that $g_K(p) < \gamma_{Forb(P_n)}(p)$. Then*
(a) *for every $v \in V(K)$, $\deg_G(v) \geq \lceil \frac{n-1}{3} \rceil$, and*
(b) *for every $v, w \in V(K)$, $\deg_G(v, w) \geq 1$.*

Proof. (a) Let $v, w \in V(K)$. By using Proposition 9(a),

$$\begin{aligned}
\deg_G(v) &\geq \left\lceil \frac{d_G(v)}{\max\{\mathbf{x}(w)\}} \right\rceil \geq \frac{\frac{p-g_K(p)}{p} + \frac{1-2p}{p}\mathbf{x}(v)}{\frac{g_K(p)}{1-p}} \\
&\geq \frac{(p-g_K(p))(1-p)}{pg_K(p)} = \frac{1-p}{g_K(p)} - \frac{1-p}{p} \\
&> \frac{(1-p) + \left(\lceil \frac{n-1}{3} \rceil - 1\right)p}{p} - \frac{1-p}{p} = \left\lceil \frac{n-1}{3} \right\rceil - 1.
\end{aligned}$$

(b) By the inclusion-exclusion principle, $d_G(v) + d_G(w) - d_G(v, w) \leq 1$, and by using Proposition 9(a), we have $d_G(v, w) \geq 2\frac{p-g_K(p)}{p} + \frac{1-2p}{p}(\mathbf{x}(v) + \mathbf{x}(w)) - 1 \geq \frac{p-g_K(p)}{p} \geq \frac{p-2g_K(p)}{p}$ and for all $u \in V(K)$, $\mathbf{x}(u) \leq g_K(p)/(1-p)$. Therefore,

$$\begin{aligned}
\deg_G(v, w) &\geq \left\lceil \frac{d_G(v, w)}{\max\{\mathbf{x}(u)\}} \right\rceil \geq \left\lceil \frac{\frac{p-2g_K(p)}{p}}{\frac{g_K(p)}{1-p}} \right\rceil = \frac{1-p}{g_K(p)} - \frac{2(1-p)}{p} \\
&> \frac{(1-p) + \left(\lceil \frac{n-1}{3} \rceil - 1\right)p}{p} - \frac{2(1-p)}{p} = \left\lceil \frac{n-1}{3} \right\rceil - \frac{1}{p}.
\end{aligned}$$

Since $p \geq \frac{1}{\lceil \frac{n-1}{3} \rceil}$, we have $\deg_G(v, w) \geq 1$. ■

Proposition 20. *Let $p \in [0, 1/2)$ and K be a p-core CRG such that K has only black vertices and white and gray edges. If $P_n \not\mapsto K$ then K has no gray path with length greater than $\lceil \frac{n}{2} \rceil - 1$.*

Proof. Suppose K has some gray path of length $l > \lceil \frac{n}{2} \rceil - 1$. Partition the vertices of P_n into l parts so that each of parts is either a set of two consecutive vertices (an edge) or single vertex. Because of the structure of P_n and the fact that $l > \lceil \frac{n}{2} \rceil - 1$, it is always possible to do so. This partition witnesses an embedding of P_n into l-path of K because we can map consecutive parts to consecutive vertices on the l-path. Since non-consecutive parts do not have edges between them and Proposition 8(a) gives that the edges of K are either white or gray, this map is an embedding that demonstrates $P_n \mapsto K$, a contradiction. ■

We consider the value of $ed_{Forb(P_n)}(p)$ from the perspective of the gray subgraphs of CRGs K. Let F be a graph, $V(F) = V(K)$, $E(F) = EG(K)$ where $K \in \mathscr{K}(Forb(P_n))$ is a p-core CRG with all black vertices such that $g_K(p) < \gamma_{Forb(P_n)}(p)$. By Proposition 19, we obtain F is a connected graph.

Suppose a maximum-length path forms a cycle in the graph F. Then Proposition 11 implies that F must be Hamiltonian. By Proposition 20, $|V(K)| \leq \lceil \frac{n}{2} \rceil - 1$ and $g_K(p) \geq \frac{1-p}{\lceil \frac{n}{2} \rceil - 1}$, a contradiction, and so we may assume that no maximum-length path in F forms a cycle. By Proposition 20, F has no path with length greater than $\lceil \frac{n}{2} \rceil - 1$, so F has no cycle with length greater than $\lceil \frac{n}{2} \rceil - 1$. And, by Proposition 19, every vertex in F has degree at least $\lceil \frac{n-1}{3} \rceil \geq 2$ and every pair of vertices has at least one common neighbor.

Let $v_1 \cdots v_\ell$ be such a maximum-length path in K such that the sum $\mathbf{x}(v_1) + \mathbf{x}(v_\ell)$ is the largest among all such paths. By Proposition 12, v_1 and v_ℓ have a unique common neighbor v_c and $N(v_1) \subseteq \{v_2, \ldots, v_c\}$. Let v_1 have d neighbors in F. Since v_1 cannot have neighbors outside of this path, the sum of the weights of the neighbors of v_1 satisfies $d_G(v_1) \leq \mathbf{x}(v_2) + \cdots + \mathbf{x}(v_c)$ in K. And if $v_i \in \{v_1, \ldots, v_{c-1}\}$ is a predecessor of a neighbor of v_1, then it is an endpoint of a path containing the same ℓ vertices, namely $v_i v_{i-1} \cdots v_1 v_{i+1} v_{i+2} \cdots v_c \cdots v_\ell$. Hence all d predecessors of gray neighbors of v_1 (including v_1 itself) have weight at most $\mathbf{x}(v_1)$. By Proposition 9, $\frac{p-g_K(p)}{p} + \frac{1-p}{p}\mathbf{x}(v_1) = \mathbf{x}(v_1) + d_G(v_1) \leq \mathbf{x}(v_1) + \cdots + \mathbf{x}(v_c) \leq d\mathbf{x}(v_1) + (c-d)\frac{g_K(p)}{1-p}$, which implies

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

By Propositions 19 and 20, we have $c \leq \lceil \frac{n}{2} \rceil - 1$ and $d \geq \lceil \frac{n-1}{3} \rceil$. And, when $\lceil \frac{n-1}{3} \rceil^{-1} \leq p \leq \frac{1}{2}$, by Proposition 9(a), we have $\mathbf{x}(v) \leq g_K(p)/(1-p)$, hence

$$g_K(p) \geq \frac{1-p}{c} \geq \frac{1-p}{\lceil \frac{n}{2} \rceil - 1} \geq \gamma_{\mathscr{H}}(p),$$

a contradiction.

So we can get $ed_{\mathscr{H}}(p) = \gamma_{\mathscr{H}}(p)$ for $p \in \left[\frac{1}{\lceil \frac{n-1}{3} \rceil}, 1\right]$. The proof is thus complete.

Acknowledgements

The authors would like to thank the two referees for their careful reading of the manuscript and for their constructive suggestions provided in the reports, in particular, for pointing out the error in the proof. Yongtang Shi was partially supported by National Natural Science Foundation of China and Natural Science Foundation of Tianjin (No. 17JCQNJC00300).

References

[1] N. Alon, *Testing subgraphs in large graphs*, Random Structures Algorithms **21** (2002) 359–370.
doi:10.1002/rsa.10056

[2] N. Alon, A. Duke, H. Lefmann, V. Rödl and R. Yuster, *The algorithmic aspects of the regularity lemma*, J. Algorithms **16** (1994) 80–109.
doi:10.1006/jagm.1994.1005

[3] N. Alon, M. Krivelevich, E. Fischer and M. Szegedy, *Efficient testing of large graphs*, in: Proc. 40th Symposium on Foundations of Computer Science) (IEEE Press, 1999) 656–666.

[4] N. Alon and U. Stav, *What is the furthest graph from a hereditary property?*, Random Structures Algorithms **33** (2008) 87–104.
doi:10.1002/rsa.20209

[5] M. Axenovich, A. Kézdy and R. Martin, *On the editing distance of graphs*, J. Graph Theory **58** (2008) 123–138.
doi:10.1002/jgt.20296

[6] J. Balogh and R. Martin, *Edit distance and its computation*, Electron. J. Combin. **15** (2008) #R20.

[7] Z. Berikkyzy, R. Martin and C. Peck, *On the edit distance of powers of cycles*, submitted.
arXiv:1509.07438

[8] J.A. Bondy and U.S.R. Murty, Graph Theory (New York, Springer, 2008).

[9] D. Chen, O. Eulenstein, D. Fernández-Baca and M. Sanderson, *Supertrees by flipping*, in: O.H. Ibarra, L. Zhang (Eds.) Computing and Combinatorics, Lectures Notes in Comput. Sci. **2387** (2002) 391–400.
doi:10.1007/3-540-45655-4_42

[10] E. Marchant and A. Thomason, *Extremal graphs and multigraphs with two weighted colours*, in: Fete of Combinatorics and Computer Science, Springer (2010) 239–286.
doi:10.1007/978-3-642-13580-4_10

[11] R. Martin, *The edit distance in graphs: Methods, results, and generalizations*, in: Recent Trends in Combinatorics, Springer (2016) 31–62.
doi:10.1007/978-3-319-24298-9_2

[12] R. Martin, *On the computation of edit distance functions*, Discrete Math. **338** (2015) 291–305.
doi:10.1016/j.disc.2014.09.005

[13] R. Martin, *The edit distance function and symmetrization*, Electron. J. Combin. **20** (2013) #P26.

[14] R. Martin and T. McKay, *On the edit distance from $K_{2,t}$-free graphs*, J. Graph Theory **77** (2014) 117–143.
doi:10.1002/jgt.21777

[15] C. Peck, On the Edit Distance from a Cycle- and Squared Cycle-Free Graph (Master's Thesis, Iowa State University, 2013).

Received 5 November 2017
Revised 11 May 2018
Accepted 11 May 2018