Counting Paths in Digraphs

Paul Seymour [1] Blair D. Sullivan [*,2]

Princeton University, Princeton, NJ 08544

1 Introduction

We begin with some terminology. All digraphs in this paper are finite. For a digraph G, we denote its vertex and edge sets by $V(G)$ and $E(G)$, respectively. Unless otherwise stated, we assume $|V(G)| = n$. The members of $E(G)$ are ordered pairs of vertices. We use the notation uv to denote an ordered pair of vertices (u, v) (whether or not u and v are adjacent). We only consider digraphs which have no loop edges uu, and at most one directed edge uv for all pairs of vertices $u \neq v$ (are *simple*). A *non-edge* in G is an unordered pair of distinct vertices u, v so that uv, vu are both not in $E(G)$. We say a simple digraph G is a *tournament* if for all pairs of vertices $u \neq v$, exactly one of uv, vu is an edge.

Given a vertex $v \in V(G)$, we define the set of *out-neighbors* to be $N^+(v) =$

* Corresponding author.
Email addresses: `pds@math.princeton.edu` (Paul Seymour),
`bdowling@math.princeton.edu` (Blair D. Sullivan).
[1] Supported by ONR grant N00014-04-1-0062, and NSF grant DMS03-54465.
[2] Partially supported by a Department of Homeland Security (DHS) Dissertation Grant.

$\{u : vu \in E(G)\}$ and analogously $N^-(v) = \{u : uv \in E(G)\}$ to be the set of *in-neighbors*. Let $\delta^+(v) = |N^+(v)|$ and $\delta^-(v) = |N^-(v)|$ denote the *out-degree* and *in-degree*, respectively.

A *directed cycle of length* t is a digraph whose vertices and edges can be ordered as $v_1, e_1, v_2, \ldots, e_{t-1}, v_t, e_t$ with $v_1, \ldots, v_t$ distinct vertices, e_i the directed edge $v_i v_{i+1}$ for $i = 1, \ldots, t-1$, and $e_t = v_t v_1$. We may denote such a cycle as v_1-v_2-$\cdots$-v_t-v_1. For an integer $k \geq 0$, let us say a digraph G is *k-free* if there is no directed cycle of G with length at most k. A digraph is *acyclic* if it has no directed cycle.

A *directed walk* in a digraph is a sequence $v_1, e_1, v_2, \ldots, e_{t-1}, v_t$ where $v_1, \ldots, v_t$ are vertices, and $e_i = v_i v_{i+1}$ is an edge for $i = 1, \ldots, t-1$; its *length* is $t-1$. A *directed path* in a digraph is a directed walk where $v_1, \ldots, v_t$ are distinct vertices (its length is $t-1$). We may denote a directed walk (or path) as v_1-v_2-$\cdots$-v_t. We say a directed path is *induced* if every edge $v_i v_j$ satisfies $j = i+1$ for $0 \leq i, j \leq t$. We say a digraph G *is a directed path* if its vertex set can be labeled $v_1, \ldots, v_n$ and its edges $e_1, \ldots, e_{n-1}$ so that $v_1, e_1, v_2, \ldots, e_{n-1}, v_n$ is an induced directed path in G. Let $W_s(G)$ be the number of distinct directed s-vertex walks in a digraph G, $P_s(G)$ the number of distinct s-vertex directed paths, and $\tilde{P}_s(G)$ the number of distinct induced s-vertex directed paths.

The first result of this paper concerns a conjecture of Thomassé that the number of induced 3-vertex directed paths in a 2-free digraph on n vertices is at most $(n-1)n(n+1)/15$. The best known approximate result is due to Bondy, and is presented in Section 2. We thank him for allowing us to include his proof in this paper. In this paper, we prove a strengthening of Thomassé's conjecture for “circular interval digraphs”.

A digraph G is a *circular interval digraph* if its vertices can be arranged in a circle such that for every triple u, v, w of distinct vertices, if u, v, w are in clockwise order and $uw \in E(G)$, then $uv, vw \in E(G)$. This is equivalent to saying that the vertex set of G can be numbered as $v_1, \ldots, v_n$ such that for $1 \leq i \leq n$, the set of out-neighbors of v_i is $\{v_{i+1}, \ldots, v_{i+a}\}$ for some $a \geq 0$, and the set of in-neighbors of v_i is $\{v_{i-b}, \ldots, v_{i-1}\}$ for some $b \geq 0$, reading subscripts modulo n.

In Section 3, we show:

Theorem 1. *If G is a 2-free circular interval digraph on n vertices, then $\tilde{P}_3(G) \leq n^3/16$.*

The second result of this paper was motivated by the following problem. For integer t, let α_t be the minimum constant so that all n-vertex digraphs with minimum out-degree at least $\alpha_t n$ have a directed cycle of length at most t (it can be proved that α_t exists). The Caccetta-Häggkvist conjecture [1] is that $\alpha_t = 1/t$. A number of papers have focused on the special case of getting an upper bound on α_3 that is as close to $1/3$ as possible. The most recent result by Shen [3] slightly tightens an argument of Hamburger, Haxell, and Kostochka [2] and proves $\alpha_3 \leq .3530381$.

One possible approach for finding upper bounds on α_3 is to find bounds on the number of short directed walks in 3-free digraphs. If G is a digraph on n vertices with minimum out-degree d, then $W_s(G) \geq d^{s-1}n$, and hence a bound of the form $W_s(G) \leq (c_s n)^s$ for 3-free digraphs G would prove there is a vertex of out-degree at most $(c_s)^{\frac{s}{s-1}} n$.

We observe that if G is 3-free, then $W_4(G) = P_4(G)$. We will show:

Theorem 2. *If G is a 3-free digraph on n vertices, then $P_4(G) \leq \frac{4}{75}n^4$.*

Note that there exists an infinite family of 3-free graphs where $P_4(G)/n^4 \to \frac{25}{512} \approx .0488$ as $n \to \infty$. These graphs are given by taking four acyclic tournaments $S_1, \ldots, S_4$, each on $n/4$ vertices and adding the edges uv where $u \in S_i$ and $v \in S_{i+1}$ for $i = 1, 2, 3$, as well as those from S_4 to S_1. This shows that using an upper bound on c_4 to imply a bound on α_3 will not lead to an improvement of Shen's result. Theorem 2 implies that any 3-free digraph on n vertices has minimum out-degree at most $\sqrt[3]{4/75}n \approx .3764n$.

2 Thomassé's Conjecture and Bondy's Result

There was a workshop on the Caccetta-Häggkvist Conjecture at the American Institute of Mathematics (AIM) in January of 2005. In discussions at that workshop, Thomassé proposed the following conjecture, and Bondy proved a partial result that we use in Section 4.

Conjecture 3 (Thomassé)**.** *If G is a 2-free digraph on n vertices, then*

$$\tilde{P}_3(G) \leq \frac{(n-1)n(n+1)}{15}.$$

This is tight on the following infinite family of digraphs: Let G_0 be the digraph consisting of a single vertex and no edges. Define G_i for $i \geq 1$ to be the digraph obtained by taking four disjoint copies of G_{i-1} (call them D_1, D_2, D_3, D_4) and

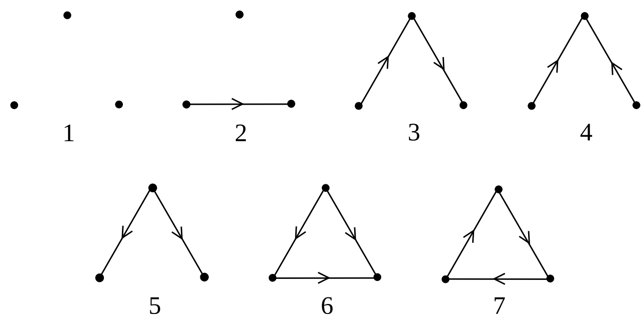

Fig. 1. The 3-vertex digraphs.

forming the digraph with vertex set $V(G_i) = \bigcup_{j=1}^{4} V(D_j)$ and edge set

$$E(G_i) = \left(\bigcup_{j=1}^{4} E(D_j) \right) \cup \{uv : u \in D_j, v \in D_{j+1}, j = 1, 2, 3, 4\},$$

where D_5 means D_1. In other words, arrange four copies of G_{i-1} in a square and put in all edges between consecutive copies in a clockwise direction. It is easy to check inductively that $\tilde{P}_3(G_i) = (n_i-1)n_i(n_i+1)/15$, where $n_i = 4^i = |V(G_i)|$.

The best known result for general 2-free digraphs is due to Bondy, whom we thank for permission to include his result here.

Theorem 4 (Bondy)**.** *If G is a 2-free digraph on n vertices, then $\tilde{P}_3(G) \leq \frac{2}{25}n^3$.*

Proof. There are seven digraphs on three vertices up to isomorphism, which we call types $1, \ldots, 7$ as shown in Figure 1. Given a digraph G with vertex set $\{v_i : 1 \leq i \leq n\}$, let d_i^- and d_i^+ denote the in-degree and out-degree of v_i $(1 \leq i \leq n)$ and s_j the number of induced subgraphs of type j in G $(1 \leq j \leq 7)$.

The following five equations hold:

$$
\begin{aligned}
s_1 + s_2 + s_3 + s_4 + s_5 + s_6 + s_7 &= \binom{n}{3} \\
s_2 + 2s_3 + 2s_4 + 2s_5 + 3s_6 + 3s_7 &= \frac{1}{2}(n-2)\sum_i (d_i^- + d_i^+) \\
s_3 + s_6 &= \sum_i \binom{d_i^-}{2} \\
s_4 + s_6 + 3s_7 &= \sum_i d_i^- d_i^+ \\
s_5 + s_6 &= \sum_i \binom{d_i^+}{2}.
\end{aligned}
$$

We prove an upper bound on $s_4 = \tilde{P}_3(G)$ as follows:

$$
\begin{aligned}
s_4 &\le \frac{2}{5}s_2 + \frac{1}{10}s_3 + s_4 + \frac{1}{10}s_5 + \frac{9}{5}s_7 \\
&= \frac{2}{5}(s_2 + 2(s_3 + s_4 + s_5) + 3(s_6 + s_7)) - \frac{7}{10}(s_3 + s_5 + 2s_6) + \frac{1}{5}(s_4 + s_6 + 3s_7) \\
&= \frac{n-2}{5}\sum_i (d_i^- + d_i^+) - \frac{7}{20}\sum_i (((d_i^-)^2 - d_i^-) + (d_i^+)^2 - d_i^+)) + \frac{1}{5}\sum_i d_i^- d_i^+ \\
&= \frac{n}{5}\sum_i (d_i^- + d_i^+) - \frac{7}{20}\sum_i ((d_i^-)^2 + (d_i^+)^2) + \frac{1}{5}\sum_i d_i^- d_i^+ - \frac{1}{20}\sum_i (d_i^+ + d_i^-) \\
&= \frac{2n^3}{25} - \frac{1}{10}\sum_i (d_i^- - d_i^+)^2 - \frac{1}{4}\sum_i \left(\frac{2n}{5} - d_i^-\right)^2 - \frac{1}{4}\sum_i \left(\frac{2n}{5} - d_i^+\right)^2 \\
&\le \frac{2n^3}{25},
\end{aligned}
$$

which proves Theorem 4. $\square$

3 Induced 3-vertex Paths in Circular Interval Digraphs

The main result of this section is:

Theorem 5. *If G is a 2-free circular interval digraph on n vertices, then $\tilde{P}_3(G) \le n^3/16$.*

We first show this is best possible. Let G be a 2-free circular interval digraph.

For $u, v \in V(G)$ let

$$d(u,v) = \begin{cases} 1 + |\{w \in V(G) : u, w, v \text{ distinct, in clockwise order}\}| & \text{if } u \neq v \\ 0 & \text{if } u = v. \end{cases}$$

For every pair uv, we say its *length* is $d(u, v)$. For integer β, let G_β be the circular interval digraph on n vertices with $E(G_\beta) = \{uv : 0 < d(u, v) \leq \beta\}$.

Lemma 6. For infinitely many values of n, there are circular interval digraphs on n vertices with exactly $n^3/16$ induced 3-vertex paths.

Proof. Let n be chosen so that $\beta = (3n-4)/8$ is an integer. A straightforward computation shows the number of induced 3-vertex paths in G_β is $n(n-2\beta-1)(2\beta-n/2+1)$. Then $G_{(3n-4)/8}$ has $(n-(3n-4)/4-1)((3n-4)/4-n/2+1) = n^3/16$ induced 3-vertex paths. □

To prove Theorem 5, we first need a few definitions and lemmas. Given $X \subseteq V(G)$, define $G|X$ to be the digraph with vertex set X and edge set $\{uv \in E(G) : u, v \in X\}$. For $Y \subseteq E(G)$, we write $G \setminus Y$ for the digraph with vertex set $V(G)$ and edge set $E(G) \setminus Y$. If $Y = \{e\}$ where $e = uv$, then we may abbreviate as $G \setminus Y = G \setminus e = G \setminus uv$. If Z is a set of non-edges of G, we write $G+Z$ for the digraph with vertex set $V(G)$ and edge set $E(G) \cup \{uv : uv \in Z\}$. Analogously, if $Z = \{f\}$ with $f = uv$, we may write $G + f = G + uv = G + Z$.

We define α_G to be the length of a shortest non-edge in G (if G has a non-edge, and otherwise we let $\alpha_G = \infty$). We also define β_G to be the length of a longest edge in G (if G has an edge, and otherwise we let $\beta_G = 0$).

Lemma 7. Let $G = (V, E)$ be a 2-free circular interval digraph. Let X be a set of longest edges in G and Y a set of shortest non-edges in G so that for all $u, v \in V(G)$, uv and vu are not both in Y. Then $G \setminus X$ and $G + Y$ are 2-free circular interval digraphs. Additionally, if $\alpha_G \leq \beta_G$, then the digraph $(G \setminus X) + Y$ is also a 2-free circular interval digraph.

Let $\xi(G)$ denote the number of pairs (uv, wx) where uv is an edge of G, wx is a non-edge and $d(u, v) > d(w, x)$ (u, v are not necessarily distinct from w, x). For a fixed $n \geq 4$, say a digraph G is *optimal* if among all 2-free circular interval digraphs on n vertices, it has the maximum number of 3-vertex induced directed paths and subject to this, $\xi(G)$ is minimum.

We now show optimal digraphs do not have edges with length at least $n/2$.

Lemma 8. If G is an optimal digraph on n vertices, then $\beta_G < n/2$.

Proof. If G has no edges, then $\beta_G = 0 < n/2$, so we may assume $E(G) \neq \emptyset$. Suppose $\beta = \beta_G \geq n/2$ and let $e = uv$ be an edge of length β. Let $G' = G \setminus e$, which is also a 2-free circular interval digraph by Lemma 7. Define $c = |N^+(v) \cap N^-(u)|$. Then

$$\tilde{P}_3(G') = \tilde{P}_3(G) - (\delta^-(u) + \delta^+(v) - 2c) + (\beta - 1 + c). \tag{1}$$

Since $N^+(v), N^-(u) \setminus N^+(v), \{u, v\}$, and $\{w : u, w, v$ are in clockwise order, $w \neq u, v\}$ are disjoint sets in $V(G)$, we have

$$\delta^+(v) + (\delta^-(u) - c) + 2 + (\beta - 1) \leq n. \tag{2}$$

Rearranging (2) gives $\delta^-(u) + \delta^+(v) \leq n - 1 - \beta + c$, and substituting for $\delta^-(u) + \delta^+(v)$ in (1) gives $\tilde{P}_3(G') \geq \tilde{P}_3(G) - (n - 1 - \beta + c - 2c) + (\beta - 1 + c)$ or

$$\tilde{P}_3(G') \geq \tilde{P}_3(G) - (n - 2\beta - 2c).$$

Since $\tilde{P}_3(G') \leq \tilde{P}_3(G)$ because G is optimal, we have $n - 2\beta - 2c \geq 0$. Since $\beta \geq n/2$, it follows that $\beta = n/2$, $c = 0$, and $\tilde{P}_3(G') = \tilde{P}_3(G)$. Hence $d(v, u) = n - d(u, v) = n/2 \geq 2$ ($n \geq 4$ since G is optimal). Then there is at least one vertex w so that u, v, w appear in clockwise order. Since $c = 0$, one of vw, wu must be a non-edge, and thus $\alpha_G < n/2$. But then $\xi(G') < \xi(G)$, contradicting the optimality of G. This proves Lemma 8. $\square$

We now prove a straightforward lemma giving upper and lower bounds on the vertex degrees in an optimal digraph.

Lemma 9. For every vertex v in an optimal digraph G,

$$\alpha_G - 1 \leq \delta^+(v), \delta^-(v) \leq \beta_G.$$

Proof. Let v be a vertex in an optimal digraph G on n vertices. Certainly α_G is finite, as otherwise G has no induced directed paths of length greater than one. Suppose $\delta^+(v) < \alpha_G - 1$. We first show that v has a non-neighbor. We have $\alpha_G \leq \beta_G + 1$, since if $v_i v_j$ is a shortest non-edge then $v_i v_{j-1}$ is an edge, and therefore has length at most β_G. Then since $\beta_G < n/2$ by Lemma 8, it follows that $\alpha_G < n/2 + 1$. Now $\delta^+(v) \leq \alpha_G - 2$ implies $\delta^+(v) < n/2 - 1$. Since v has in-degree at most $(n-1)/2$ by Lemma 8, it has less than $(n/2-1)+(n-1)/2 = n - 3/2$ neighbors. Consequently v has a non-neighbor, and we let u be the first vertex following v in the clockwise order for which vu is a non-edge. Then vu has length $\delta^+(v)+1 < \alpha_G$, a contradiction to the definition of α_G. Analogously, $\delta^-(v) \geq \alpha_G - 1$. Now, suppose $\delta^+(v) > \beta_G$. Then the edge from v to its last clockwise out-neighbor has length $1 + (\delta^+(v) - 1) > \beta_G$. This contradicts the definition of β_G. Again, $\delta^-(v) \leq \beta_G$ by an analogous argument. $\square$

This allows us to give a lower bound on α_G in optimal digraphs G.

Lemma 10. If G is an optimal digraph on n vertices, then $\alpha_G > n/4$.

Proof. If G has no non-edges, then $\alpha_G = \infty > n/4$. We may assume G has a non-edge. Suppose $\alpha = \alpha_G \leq n/4$ and let $e = uv$ be a non-edge of length α. Let $G' = G + e$, which is also a 2-free circular interval digraph by Lemma 7. Define $c = |N^+(v) \cap N^-(u)|$. Then

$$\tilde{P}_3(G') = \tilde{P}_3(G) - (\alpha - 1 + c) + (\delta^-(u) + \delta^+(v) - 2c).$$

Since $\tilde{P}_3(G') \leq \tilde{P}_3(G)$ because G is optimal, we have

$$\alpha - 1 + 3c \geq \delta^-(u) + \delta^+(v). \tag{3}$$

Suppose $c = 0$. Then since $\delta^+(v), \delta^-(u) \geq \alpha - 1$ by Lemma 9, we have $\alpha = 1$ and $\delta^-(u) = \delta^+(v) = 0$. Letting w be the vertex immediately following v in the circular order, we see that $G + \{uv, vw\}$ has more induced 2-edge paths than G, contradicting its optimality. Thus $c > 0$. Now $N^+(v), N^-(u) \setminus N^+(v)$, $\{w : u, w, v \text{ are in clockwise order}, w \neq u, v\}$, and $\{u, v\}$ form a partition of $V(G)$, so

$$\delta^+(v) + \delta^-(u) - c + (\alpha - 1) + 2 = n. \tag{4}$$

We observe that Lemmas 8 and 9 imply

$$\delta^+(v) + \delta^-(u) \leq 2\beta \leq n - 1. \tag{5}$$

Taking the combination (3)+(4)−2·(5) and simplifying gives $4\alpha \geq n$, so $\alpha = n/4$ and we have equality in both (3) and (5). The equality in (5) implies $\beta_G \geq (n-1)/2$, so $\beta_G > \alpha = n/4$. It follows that $\xi(G') < \xi(G)$. Yet the equality in (3) tells us $\tilde{P}_3(G) = \tilde{P}_3(G')$, contradicting the optimality of G. This proves Lemma 10. $\square$

Lemma 11. If G is an optimal digraph and uv is a shortest non-edge in G, then $N^+(v) \cap N^-(u) \neq \emptyset$.

Proof. Suppose not, and let uv be a non-edge of length α_G in an optimal digraph G on n vertices. Then by Lemma 10 and the fact that $n \geq 4$, $\alpha_G > n/4 \geq 1$. Let $G' = G + uv$ and $\alpha = \alpha_G$. Then

$$\tilde{P}_3(G') = \tilde{P}_3(G) + \delta^+(v) + \delta^-(u) - (\alpha - 1).$$

Since $\tilde{P}_3(G') \leq \tilde{P}_3(G)$ by optimality, $\delta^+(v) + \delta^-(u) \leq \alpha - 1$. But by Lemma 9, $\delta^+(v), \delta^-(u) \geq \alpha - 1$. Then $2\alpha - 2 \leq \alpha - 1$, or $\alpha \leq 1$, a contradiction. This proves Lemma 11. □

We can now prove that in an optimal digraph G, $\alpha_G + \beta_G$ is approximately $3|V(G)|/4$. Let

$$\epsilon_G = \begin{cases} 0 & \beta_G > \alpha_G \\ 1 & \beta_G \leq \alpha_G. \end{cases}$$

Lemma 12. If G is an optimal digraph on n vertices, then

$$\frac{3n}{4} - \frac{1}{2} - \frac{\epsilon_G}{4} < \alpha_G + \beta_G < \frac{3n}{4} + \frac{1}{2} + \frac{\epsilon_G}{4}.$$

Additionally, if some vertex is incident with a longest edge, but with no shortest non-edge, then $\alpha_G + \beta_G < 3n/4 + \epsilon_G/4$, and if some vertex is incident with a shortest non-edge but with no longest edge, then $\alpha_G + \beta_G > 3n/4 - \epsilon_G/4$.

Proof. Let G be an optimal digraph on n vertices with $\alpha = \alpha_G$, $\beta = \beta_G$, and $\epsilon = \epsilon_G$.

Step 1. $\alpha + \beta < 3n/4 + 1/2 + \epsilon/4$, and if some vertex is incident with a longest edge but no shortest non-edge, then $\alpha + \beta < 3n/4 + \epsilon/4$.

If G has no edges, then $\beta = 0$, and since $\alpha \leq \beta + 1$ by Lemma 9, we have $\alpha + \beta \leq 1 \leq 3n/4$ (since $n \geq 4$), as required. Thus $E(G) \neq \emptyset$. Let uv be a longest edge in G and $G' = G \setminus uv$. For notational convenience, set $\delta^+(v) = a$, $\delta^-(u) = b$, and $|N^+(v) \cap N^-(u)| = c$. The number of induced 3-vertex paths in G' is

$$\tilde{P}_3(G') = \tilde{P}_3(G) + (b - c) + (a - c) + (\beta - 1 + c).$$

Since G is optimal, $\tilde{P}_3(G') \leq \tilde{P}_3(G)$, and strict inequality holds if $\beta > \alpha$ (because then $\xi(G') < \xi(G)$), it follows that:

$$3c < a + b - \beta + 1 + \epsilon. \tag{6}$$

Suppose that no vertex is non-adjacent to both u and v. Then counting the vertices, we have $(c+\beta-1)+(a-c)+(b-c)+2 = n$, or $c = a+b+\beta+1-n$. Substituting for c in (6) gives $2(a + b + 1) + 3(\beta - n) < \epsilon - \beta$, or

$$4\beta + 2 + 2(a + b) < 3n + \epsilon.$$

Since $a, b \geq \alpha - 1$ by Lemma 9, it follows that $4\beta + 2 + (4\alpha - 4) < 3n + \epsilon$, or

$$\alpha + \beta < 3n/4 + 1/2 + \epsilon/4.$$

Note that if $a \geq \alpha$ or $b \geq \alpha$ (one of the endpoints of uv is not incident with a shortest non-edge), we have $4\beta + 2 + (4\alpha - 2) < 3n + \epsilon$, or

$$\alpha + \beta < 3n/4 + \epsilon/4.$$

Thus we may assume there is a vertex y non-adjacent to u and v. In this case, $a + b + (\beta - 1) + 3 \leq n$, so $2\alpha + \beta \leq n$. We know $\alpha \geq (n + 1)/4$ by Lemma 10 (since $\alpha \in \mathbb{Z}$), proving

$$\alpha + \beta < 3n/4 - 1/4 < 3n/4 + \epsilon/4.$$

This proves Step 1. $\square$

Step 2. $\alpha + \beta > 3n/4 - 1/2 - \epsilon/4$, and if some vertex is incident with a shortest non-edge but no longest edge, then $\alpha + \beta > 3n/4 - \epsilon/4$.

If G has no non-edges, then $\alpha = \infty$, yet $\alpha \leq \beta + 1 < n/2 + 1$ by Lemmas 9 and 8, a contradiction. Then let uv be a shortest non-edge in G, and $G' = G + uv$. For notational convenience, set $\delta^+(v) = a$, $\delta^-(u) = b$, and $|N^+(v) \cap N^-(u)| = c$. The number of induced 3-vertex paths in G' is

$$\tilde{P}_3(G') = \tilde{P}_3(G) + (b - c) + (a - c) - (c + \alpha - 1).$$

Since G is optimal, $\tilde{P}_3(G') \leq \tilde{P}_3(G)$, and strict inequality holds if $\beta > \alpha$ (because then $\xi(G') < \xi(G)$), it follows that:

$$a + b - \alpha + 1 < 3c + \epsilon. \tag{7}$$

We know $\alpha + 1 + a + b - c = n$ by Lemma 11. We solve for $c = \alpha + a + b + 1 - n$, and substitute into equation (7). This gives $a + b - \alpha + 1 - \epsilon < 3(\alpha + a + b + 1 - n)$, or

$$3n < 4\alpha + 2(a + b) + 2 + \epsilon.$$

Since $a, b \leq \beta$ by Lemma 9, $3n < 4(\alpha + \beta) + 2 + \epsilon$, or

$$\alpha + \beta > 3n/4 - 1/2 - \epsilon/4.$$

If $a < \beta$ or $b < \beta$ (one of the endpoints of uv is not incident with a longest edge), we instead have $3n < 4\alpha + 2\beta + 2(\beta - 1) + \epsilon$, or

$$\alpha + \beta > 3n/4 - \epsilon/4,$$

as desired. This proves Step 2, and completes the proof of Lemma 12. $\square$

We define $\gamma_G = 4(\alpha_G + \beta_G) - 3n$, where $|V(G)| = n$.

Lemma 13. Let G be an optimal digraph on n vertices with $\beta_G > \alpha_G$. Then $-1 \leq \gamma_G \leq 1$. Furthermore, $\gamma_G = -1$ if some vertex is incident with a longest edge and with no shortest non-edge, and $\gamma_G = 1$ if some vertex is incident with a shortest non-edge and with no longest edge.

Proof. Since $\gamma_G = 4(\alpha_G + \beta_G) - 3n$ and $\epsilon_G = 0$, Lemma 12 implies

$$-2 = 4(3n/4 - 1/2) - 3n < \gamma_G < 4(3n/4 + 1/2) - 3n = 2.$$

Since γ_G is an integer, this is equivalent to $-1 \leq \gamma_G \leq 1$. Furthermore, if some vertex is incident with a longest edge and no shortest non-edge, Lemma 12 proves $\gamma_G < 0$. Since $\gamma_G \geq -1$, this implies $\gamma_G = -1$. Similarly, if some vertex is incident with a shortest non-edge and no longest edge, we have $\gamma_G > 0$, which combined with $\gamma_G \leq 1$ implies $\gamma_G = 1$. This proves Lemma 13. $\square$

We now prove several more facts about optimal digraphs. We say a set of vertices X in a digraph G is *stable* if $uv \notin E(G)$ for all $u, v \in X$; that is, $G|X$ has no edges.

Lemma 14. If G is an optimal digraph, it has no stable set of size at least 3.

Proof. Suppose not and let $|V(G)| = n$. Take a stable set $\{u, v, w\}$ so that $d(u, v)$ is minimum, and let $k = d(u, v)$. Then by Lemma 10, $\alpha_G > n/4 \geq 1$, so every vertex has at least one out-neighbor. It follows that $k \geq 2$. Let $G' = G + uv$. Then G' is a circular interval digraph, since $d(u, v)$'s minimality implies uw and wv are edges for all w between u and v in the circular order.

Since $|N^+(v) \cap N^-(u)| = 0$, it follows that

$$\tilde{P}_3(G') = \tilde{P}_3(G) + \delta^+(v) + \delta^-(u) - (k-1).$$

Since $\tilde{P}_3(G') \leq \tilde{P}_3(G)$ by optimality, $\delta^+(v) + \delta^-(u) \leq k-1$.

Let y be the furthest out-neighbor of v in G. Then v is non-adjacent to the next vertex in the circular order (call it a), and the non-edge va is part of a stable set of size three, namely $\{v, a, u\}$ (if u were adjacent to a then there could not be a vertex w to which both u and v were non-adjacent). This means the length of va is at least k by choice of uv, so $\delta^+(v) \geq k-1$. An analogous argument shows $d^-(u) \geq k-1$. Since $\delta^+(v) + \delta^-(u) \leq k-1$, it follows that $k = 1$, a contradiction. This proves Lemma 14. $\square$

We say a pair of vertices uv in an optimal digraph G is *extreme* if uv is a longest edge or a shortest non-edge in G.

Lemma 15. Let G be an optimal digraph with $\beta_G > \alpha_G$, and u, v, w vertices appearing in clockwise order. Then not all of uv, vw, wu are extreme pairs. Additionally, if some two of them are extreme, then either all three pairs are edges, or two are edges and the third is a shortest non-edge.

Proof. Let G be an optimal digraph on n vertices with $\beta_G > \alpha_G$. We begin by proving that no vertex is in two shortest non-edges. Assume not, and let v be a vertex with uv and vw non-edges of length α_G. Hence u, v, w appear in clockwise order. Then by Lemma 14, u and w must be adjacent. If $uw \in E(G)$, G is not a circular interval graph, a contradiction. Thus $wu \in E(G)$, and say it has length L. Let $G' = G + vw$. This is a circular interval digraph by Lemma

7. Then

$$\tilde{P}_3(G') = \tilde{P}_3(G) - (\alpha_G - 1 + \delta^+(w) - L) + L + (\alpha_G - 1) - (\delta^+(w) - L).$$

Since $\tilde{P}_3(G') \leq \tilde{P}_3(G)$ by optimality, $3L - 2\delta^+(w) \leq 0$, or

$$L \leq \frac{2\delta^+(w)}{3} \leq \frac{2\beta_G}{3}.$$

Now since $2\alpha_G + L = n$, we have

$$n \leq 2\alpha_G + \frac{2\beta_G}{3}. \tag{8}$$

We note that since $\beta_G > \alpha_G$ and v is incident with two shortest non-edges, v cannot be incident with a longest edge. Then Lemma 13 gives $\gamma_G = 1$, or

$$\alpha_G + \beta_G = \frac{3n+1}{4}. \tag{9}$$

Combining equations (8) and (9), we have

$$n \leq \frac{2(\alpha_G + \beta_G)}{3} + \frac{4\alpha_G}{3} = \frac{3n+1}{6} + \frac{4\alpha_G}{3}.$$

This implies

$$\alpha_G \geq \frac{3n-1}{8}.$$

Then since $\alpha_G + \beta_G = (3n+1)/4$, $\beta_G \leq (3n+3)/8$. Yet $\beta_G > \alpha_G$, and both are integers. This implies there are two integers in the range $[(3n-1)/8, (3n+3)/8]$, a contradiction. This proves no vertex is incident with two shortest non-edges.

We now show no triple of vertices forms three longest edges. Let u, v, w be in clockwise order, and assume uv, vw, wu are all edges of length β_G. Then $3\beta_G = n$. Since $\beta_G > \alpha_G$, this implies $n > 2\beta_G + \alpha_G$. Now, Lemma 12 gives $\alpha_G + \beta_G > 3n/4 - 1/2$ (since $\epsilon_G = 0$), so $n > \beta_G + 3n/4 - 1/2$, or $\beta_G < n/4 + 1/2$. Then $\alpha_G < \beta_G < (n+2)/4$ implies $\alpha_G < (n+1)/4$, a contradiction to Lemma 10.

This proves that for u, v, w in clockwise order, not all of uv, vw, wu are extreme pairs. Additionally, it proves that if two are extreme pairs, at least one must be a longest edge. To complete the proof of the theorem, we need to show that if two of uv, vw, wv are longest edges, the third pair must be an edge and that if there is a shortest non-edge among uv, vw, wv, the other two pairs must be edges.

We first prove there do not exist $u, v, w \in V(G)$ so that uv, vw are edges of length β_G and wu is a non-edge. Suppose such u, v, w exist. It follows that u, v, w are in clockwise order. Since all three pairs are not extreme, wu has length $L > \alpha_G$. Then $2\beta_G + L = n$, or $2\beta_G + \alpha_G < n$. Since v is incident with two longest edges, it cannot be incident with a shortest non-edge, and Lemma 13 implies $\gamma_G = -1$, or $\alpha_G + \beta_G = (3n-1)/4$. We then have $(3n-1)/4 + \beta_G < n$, or $\beta_G < (n+1)/4$. Since $\alpha_G < \beta_G$, this contradicts Lemma 10.

Finally, suppose there are $u, v, w \in V(G)$ so that uv, vw, and wu consist of a shortest non-edge, a longest edge, and a non-edge of length $L > \alpha_G$. Then $\alpha_G + \beta_G + L = n$ implies $2\alpha_G + \beta_G < n$. Since $\alpha_G + \beta_G > (3n-2)/4$ by Lemma 12, we see that $\alpha_G + (3n-1)/4 < n$, or $\alpha_G < (n+1)/4$. Again, this contradicts Lemma 10.

This proves Lemma 15. $\square$

Given an optimal digraph G, let $S = v_1, f_1, v_2, f_2, v_3, f_3, \ldots, f_k, v_{k+1}$ be a sequence where the v_i are vertices of G, and the $f_i = v_i v_{i+1}$ are extreme pairs of G. We say S is an *alternating sequence* if it satisfies the conditions:

i. $f_i \neq f_j$ for $i \neq j$.

ii. For $1 \leq i \leq k-1$, if f_i is an edge, then f_{i+1} is a non-edge.

iii. For $1 \leq j \leq k-1$, if f_j is a non-edge, then f_{j+1} is an edge.

In other words, S is an alternating sequence of longest edges and shortest non-edges. Define X_S to be the set of longest edges in S and Y_S to be the set of shortest non-edges. We say a sequence S is an *augmenting sequence* if it is a maximal alternating sequence.

Lemma 16. Let G be an optimal digraph on n vertices with $\beta_G > \alpha_G$. If $\alpha_G + \beta_G \leq 3n/4$, then every shortest non-edge has a longest edge incident with each of its endpoints. If $\alpha_G + \beta_G \geq 3n/4$, every longest edge has a shortest non-edge incident with each of its endpoints. Consequently, every augmenting sequence in G has at least 3 extreme pairs.

Proof. Let G be an optimal digraph on n vertices with $\beta_G > \alpha_G$. If some longest edge uv is not incident with a shortest non-edge at both u and v, since $\beta_G > \alpha_G$, Lemma 12 gives $\alpha_G + \beta_G > 3n/4$. Similarly, if some shortest non-edge uv is not incident with a longest edge at both u and v, Lemma 12 gives $\alpha_G + \beta_G < 3n/4$. Clearly, these cannot hold simultaneously. This proves Lemma 16. $\square$

Lemma 17. Let G be an optimal digraph with $\beta_G > \alpha_G$. For every augmenting sequence $S = v_1, f_1, v_2, f_2, v_3, f_3, \ldots, f_k, v_{k+1}$ in G, $v_i \neq v_j$ for $i \neq j$, except possibly $v_{k+1} = v_1$.

Proof. Suppose not, and let $v_h = v_j$ with $1 \leq h, j \leq k+1$, and h different from j. Suppose $h, j > 1$. Then one of f_{h-1}, f_{j-1} is a longest edge and one is a shortest non-edge (since there cannot be two longest edges ending at a given vertex). However, this contradicts that G is a circular interval digraph with $\beta_G > \alpha_G$. Analogously, if $h, j < k+1$, one of f_{h+1}, f_{j+1} is a longest edge, and the other is a shortest non-edge, and the same contradiction is reached. This proves that if $v_h = v_j$, then $\{h, j\} = \{1, k+1\}$. $\square$

Theorem 18. *If G is an optimal digraph, then $\beta_G \leq \alpha_G$. Furthermore, either $\alpha_G = \beta_G$ or $\alpha_G = \beta_G + 1$.*

Proof. The second statement in Theorem 18 follows from the first since $\alpha_G \leq \beta_G + 1$ by Lemma 9, so it remains prove the first. Suppose G is an optimal digraph on n vertices with $\beta = \beta_G > \alpha_G = \alpha$. Also let $\gamma = \gamma_G$. Let S be an augmenting sequence in G, and set $X = X_S$ and $Y = Y_S$. Let $S = v_1, f_1, v_2, f_2, \ldots, f_k, v_{k+1}$. Define $G' = (G + X) \setminus Y$, and note this is also a 2-free circular interval digraph by Lemma 7.

Fix an extreme pair $uv \in X \cup Y$. Define $G_{uv} = G + uv$ if $uv \in Y$ and $G_{uv} = G \setminus uv$ if $uv \in X$. For vertices w different from u, v, define $p(w) = 1$ if $G|\{u, v, w\}$ is a directed path (and $p(w) = 0$ otherwise), and $q(w) = 1$ if $G_{uv}|\{u, v, w\}$ is a directed path (and $q(w) = 0$ otherwise). Finally, let

$$R(uv) = \sum_{w \neq u,v} q(w) - p(w).$$

We now define

$$R = \sum_{uv \in X \cup Y} R(uv).$$

By Lemma 15, no triple of vertices in G contains three extreme pairs. Let T_1 be the number of triples of vertices $\{u, v, w\}$ such that u, v, w are in clockwise order in G and two of uv, vw, wu are in X. Let T_2 be the number of $\{u, v, w\}$ such that u, v, w are in clockwise order in G, one of uv, vw, wu is in X, and one is in Y.

Finally, for a vertex v, define $s^+(v) = |N^+(v)| - (\alpha - 1)$ and $s^-(v) = |N^-(v)| -$

$(\alpha - 1)$. Also define $t^+(v) = \beta - |N^+(v)|$ and $t^-(v) = \beta - |N^-(v)|$. Then $s^+(v), s^-(v), t^+(v)$, and $t^-(v)$ are non-negative by Lemma 9.

Step 1. $\tilde{P}_3(G') - \tilde{P}_3(G) = R - 2T_1 + 2T_2$.

This follows from the definitions, using Lemma 15 to characterize those pairs with two extreme pairs in $X \cup Y$. □

Step 2. For $uv \in X$, $R(uv) = \gamma + 2s^+(v) + 2s^-(u) - 2$.

There are $\beta - 1 + |N^+(v) \cap N^-(u)|$ vertices w where $q(w) = 1$ and $p(w) = 0$. Since $\beta > \alpha$, Lemma 12 implies $\alpha + \beta > 3n/4 - 1/2$, and combined with Lemma 10, this gives $2\alpha + \beta \geq n$. By Lemma 15, this implies no vertex is non-adjacent to both u and v. We can now count that there are $n - (\beta - 1) - 2 - |N^+(v) \cap N^-(u)|$ vertices w where $q(w) = 0$ but $p(w) = 1$. Recalling that $R(uv) = \sum_{w \neq u,v} q(w) - p(w)$, we have

$$R(uv) = \beta - 1 + |N^+(v) \cap N^-(u)| - (n - \beta - 1 - |N^+(v) \cap N^-(u)|),$$

or

$$R(uv) = 2\beta + 2|N^+(v) \cap N^-(u)| - n.$$

We know $|N^+(v) \cap N^-(u)| = |N^+(v)| + |N^-(u)| + (\beta - 1) + 2 - n$, which we can rewrite as $2\alpha + \beta - n - 1 + (|N^+(v)| - (\alpha - 1)) + (|N^-(u)| - (\alpha - 1))$. Then using the definitions of $s^+(v)$ and $s^-(u)$, we have

$$R(uv) = 2\beta + 2(2\alpha + \beta - n - 1) + 2s^+(v) + 2s^-(u) - n,$$

which simplifies to

$$R(uv) = 4(\alpha + \beta) - 3n - 2 + 2s^+(v) + 2s^-(u).$$

From the definition of $\gamma = 4(\alpha + \beta) - 3n$, this proves Step 2. □

Step 3. For $uv \in Y$, $R(uv) = 2t^+(v) + 2t^-(u) - \gamma - 2$.

There are $n - (\alpha + 1) - |N^+(v) \cap N^-(u)|$ vertices w where $q(w) = 1$ but $p(w) = 0$, and there are $\alpha - 1 + |N^+(v) \cap N^-(u)|$ vertices w where $q(w) = 0$ and $p(w) = 1$. Recalling that $R(uv) = \sum_{w \neq u,v} q(w) - p(w)$, we have

$$R(uv) = n - \alpha - 1 - |N^+(v) \cap N^-(u)| - (\alpha - 1 + |N^+(v) \cap N^-(u)|),$$

or

$$R(uv) = n - 2\alpha - 2|N^+(v) \cap N^-(u)|.$$

We know $|N^+(v) \cap N^-(u)| = |N^+(v)| + |N^-(u)| + (\alpha - 1) + 2 - n$, which we can rewrite as $\alpha + 2\beta - n + 1 - (\beta - |N^+(v)|) - (\beta - |N^-(u)|)$. Then using the definitions of $t^+(v), t^-(u)$, we have

$$R(uv) = n - 2\alpha - 2(\alpha + 2\beta - n + 1) + 2t^+(v) + 2t^-(u),$$

which simplifies to

$$R(uv) = 3n - 4(\alpha + \beta) + 2t^+(v) + 2t^-(u) - 2.$$

From the definition of $\gamma = 4(\alpha + \beta) - 3n$, this proves Step 3. $\square$

Step 4. $\tilde{P}_3(G') - \tilde{P}_3(G) \geq 0$.

For $1 \leq i \leq k$, if $v_i v_{i+1} \in X$, then v_i has out-degree β and $t^+(v_i) = 0$; similarly, if $v_i v_{i+1} \in Y$, then $s^+(v_i) = 0$. For $2 \leq i \leq k+1$, if $v_{i-1}v_i \in X$, then v_i has in-degree β and $t^-(v_i) = 0$; analogously, if $v_{i-1}v_i \in Y$, then $s^-(v_i) = 0$. For $2 \leq i \leq k$, we note that the definition of an augmenting sequence implies that one of $v_{i-1}v_i, v_i v_{i+1}$ is in X and the other in Y. Then by Steps 2 and 3, for

$3 \le i \le k$

$$R(v_{i-1}v_i) = \begin{cases} \gamma - 2 & \text{if } v_{i-1}v_i \in X \\ \\ -\gamma - 2 & \text{if } v_{i-1}v_i \in Y. \end{cases} \tag{10}$$

We see $T_1 = 0$ unless $v_1 = v_{k+1}$, k is odd, and v_1v_2, v_kv_{k+1} are both in X, and in that case $T_1 = 1$. Also $T_2 = k - 1$ unless $v_1 = v_{k+1}$ and k is even, and in that case $T_2 = k$.

First, suppose $v_1 = v_{k+1}$ and k is odd. Then v_1v_2 and v_kv_1 must be in X by Lemma 15. By Step 1, $\tilde{P}_3(G') - \tilde{P}_3(G) = R - 2T_1 + 2T_2$, which in conjunction with equation (10) and the earlier argument giving $s^+(v_2) = s^-(v_k) = 0$ implies

$$\tilde{P}_3(G') - \tilde{P}_3(G) = (\gamma + 2s^-(v_1) - 2) + (\gamma + 2s^+(v_{k+1}) - 2) - 2(k-2) - \gamma - 2T_1 + 2T_2.$$

Using that $v_1 = v_{k+1}$ and substituting $T_1 = 1$ and $T_2 = k - 1$, we have

$$\tilde{P}_3(G') - \tilde{P}_3(G) = \gamma + 2(s^-(v_1) + s^+(v_1) - k - 1 + (k-1)) = \gamma + 2s^-(v_1) + 2s^+(v_1) - 4.$$

We know $|N^-(v_1)| = |N^+(v_1)| = \beta$, so $s^-(v_1) = s^+(v_1) = \beta - \alpha + 1$. Since $\beta > \alpha$, $s^-(v_1), s^+(v_1) \ge 2$. Substituting in the above inequality, we have

$$\tilde{P}_3(G') - \tilde{P}_3(G) \ge \gamma + 4.$$

Since $\gamma \ge -1$ by Lemma 13, $\tilde{P}_3(G') > \tilde{P}_3(G)$, as required.

Now suppose that $v_1 = v_{k+1}$ and k is even. Since $v_1 = v_{k+1}$, equation (10) holds for $2 \le i \le k + 1$. Also, by Step 1,

$$\tilde{P}_3(G') - \tilde{P}_3(G) = R - 2T_1 + 2T_2 = -2k - 2T_1 + 2T_2. \tag{11}$$

Substitution of $T_1 = 0$, $T_2 = k$ in equation (11) yields $\tilde{P}_3(G') - \tilde{P}_3(G) = 0$, as required.

Thus we may assume $v_1 \neq v_{k+1}$. We note Lemma 16 implies that if $v_1 \neq v_{k+1}$, then v_1v_2, v_kv_{k+1} are either both in X or both in Y, and k is odd. Let $\mu = -1$ if $v_1v_2 \in X$ and $\mu = 1$ otherwise.

By Step 1, $\tilde{P}_3(G') - \tilde{P}_3(G) = R - 2T_1 + 2T_2$, which in conjunction with equation (10) implies

$$\tilde{P}_3(G') - \tilde{P}_3(G) = R(v_1v_2) + R(v_kv_{k+1}) - 2(k-2) + \mu\gamma - 2T_1 + 2T_2, \quad (12)$$

where

$$R(v_1v_2) = \begin{cases} \gamma + 2s^-(v_1) - 2 & \text{if } v_1v_2 \in X \\ 2t^-(v_1) - \gamma - 2 & \text{if } v_1v_2 \in Y, \end{cases}$$

and

$$R(v_kv_{k+1}) = \begin{cases} \gamma + 2s^+(v_{k+1}) - 2 & \text{if } v_kv_{k+1} \in X \\ 2t^+(v_{k+1}) - \gamma - 2 & \text{if } v_kv_{k+1} \in Y. \end{cases}$$

Then equation (12) can be simplified to

$$\tilde{P}_3(G') - \tilde{P}_3(G) = \begin{cases} \gamma + 2s^-(v_1) + 2s^+(v_{k+1}) - 2k - 2T_1 + 2T_2 & \text{if } v_1v_2 \in X \\ -\gamma + 2t^-(v_1) + 2t^+(v_{k+1}) - 2k - 2T_1 + 2T_2 & \text{if } v_1v_2 \in Y. \end{cases}$$

Recalling that $T_1 = 0$ and $T_2 = k - 1$, we have

$$\tilde{P}_3(G') - \tilde{P}_3(G) = \begin{cases} \gamma + 2s^-(v_1) + 2s^+(v_{k+1}) - 2 & \text{if } v_1v_2 \in X \\ -\gamma + 2t^-(v_1) + 2t^+(v_{k+1}) - 2 & \text{if } v_1v_2 \in Y. \end{cases}$$

First, suppose $v_1v_2 \in X$. Then $s^-(v_1), s^+(v_{k+1}) \geq 1$, since otherwise S is not maximal. Since $\gamma \geq -1$ by Lemma 13, this proves $\tilde{P}_3(G') - \tilde{P}_3(G) > 0$. On the other hand, suppose $v_1v_2 \in Y$. Then $t^-(v_1), t^+(v_{k+1}) \geq 1$ by the maximality of S. Now $\gamma \leq 1$ by Lemma 13, and again $\tilde{P}_3(G') - \tilde{P}_3(G) > 0$. This completes the proof of Step 4. $\square$

We observe that $\xi(G') < \xi(G)$ follows immediately from the definition of $\xi(G)$ and the fact $\beta < \alpha$. Yet we have now contradicted the optimality of G. This proves Theorem 18. $\square$

Finally, we prove a lemma relating the number of induced 3-vertex paths in a general circular interval digraph with longest edge of length β to the number in G_β. We need two further definitions.

Let H_β be the subgraph of G_β with the same vertex set, and $E(H_\beta) = \{uv : d(u,v) = \beta\}$. Also, for $X \subseteq E(H_\beta)$, let $t(X)$ be the number of vertices of H_β which are incident with exactly one edge in X.

Lemma 19. Let $n \geq 4$, and let β be an integer satisfying $-2 \leq 8\beta - 3n \leq 2$. Then for all $X \subseteq E(H_\beta)$,

$$|X|(8\beta - 3n) + t(X) + n(n - 2\beta - 1)(2\beta - n/2 + 1) \leq n^3/16.$$

Proof. Let $\delta = 8\beta - 3n$. Then $-2 \leq \delta \leq 2$, and (eliminating β) we must show that

$$|X|\delta + t(X) + n(n - \delta - 4)(n + \delta + 4)/16 \leq n^3/16,$$

that is,

$$|X|\delta + t(X) \leq n(\delta + 4)^2/16 \tag{13}$$

for all $X \subseteq E(H_\beta)$.

Let $t = t(X)$, and $Y = E(H_\beta) \setminus X$. In G_β, every vertex is incident with two edges of length β. Since $X \cup Y = E(H_\beta)$, and t counts vertices which are incident with exactly one edge in X, we have that $2|Y| \geq t$, $2|X| \geq t$, and $|X| + |Y| = n$.

Case 1. $\delta = 0$.

Since $\delta = 0$, equation (13) becomes $t \leq n$, which is clear since G has n vertices.

Case 2. $\delta = 1$.

Substituting into inequality (13), we must show that $|X| + t \leq 25n/16$. Since $2|Y| \geq t$ and $2|X| \geq t$, it follows that $6|Y| + 2|X| \geq 4t$. Using $|X| + |Y| = n$ to eliminate $|Y|$ gives $6(n - |X|) + 2|X| \geq 4t$, that is, $|X| + t \leq 3n/2 < 25n/16$, as required.

Case 3. $\delta = 2$.

In this case, equation (13) becomes $2|X| + t \leq 9n/4$. But since $2|Y| \geq t$ and $|Y| = n - |X|$, we have $2(n - |X|) \geq t$, or $2|X| + t \leq 2t \leq 2n < 9n/4$, as required.

Case 4. $\delta = -1$.

When $\delta = -1$, we need to show $t - |X| \leq 9n/16$ to prove the inequality in (13). If $|X| \leq n/2$ then $t \leq 2|X| \leq |X| + n/2$. If $|X| > n/2$, then $t \leq 2|Y| = 2(n - |X|) \leq n/2 + |X|$. In both cases, $t \leq |X| + n/2 < |X| + 9n/16$, as required.

Case 5. $\delta = -2$.

Finally, when $\delta = -2$, proving (13) requires $t - 2|X| \leq n/4$. But $2|X| \geq t$, so this is trivial. This proves Lemma 19. $\square$

Lemma 20. Let $G = G_\beta \setminus X$, where $X \subseteq E(H_\beta)$, and $8\beta - 3n \geq 2$. Then $\tilde{P}_3(G) = \tilde{P}_3(G_\beta) + |X|(8\beta - 3n) + t(X)$.

Proof. For each edge uv in X, the number of induced 3-vertex paths using both of u, v which are in G and not G_β is $\beta - 1 + (3\beta - n - 1)$, plus one for each vertex w so that uw or wv is in X. The number of induced 3-vertex paths using u and v which are in G_β and not G is $2(n - 2\beta - 1)$. Summing over all uv in X, we see that $\tilde{P}_3(G) = \tilde{P}_3(G_\beta) + |X|(8\beta - 3n) + t(X)$, by definition of $t(X)$. This proves Lemma 20. $\square$

Proof of Theorem 5. Let G be a digraph on n vertices. If $n \leq 2$, then $\tilde{P}_3(G) = 0 \leq n^3/16$, and if $n = 3$, then $\tilde{P}_3(G) \leq 1 \leq 27/16$. So we may assume $n \geq 4$, and that G is optimal. It follows from Theorem 18 that every optimal digraph G with maximum edge length β can be written as $G_\beta \setminus X$ for some set $X \subseteq H_\beta$. We now show that every choice of X gives $\tilde{P}_3(G) \leq n^3/16$. Let $\alpha = \alpha_G$ and $\beta = \beta_G$. By Lemma 18, either $\alpha = \beta$, or $\alpha = \beta + 1$.

Suppose $\alpha = \beta + 1$. Then $X = \emptyset$, and $G = G_\beta$. A straightforward calculation gives that

$$\tilde{P}_3(G_\beta) = n(n - 2\beta - 1)(2\beta - n/2 + 1). \tag{14}$$

Let $x = 2\beta + 1$. Then we need to show $n(n - x)(x - n/2) \leq n^3/16$, or $x(3n/2 - x) \leq 9n^2/16$. Now, Lemma 12 implies that $3n/4 - 1/2 \leq x \leq 3n/4 + 1/2$. We see that $x(3n/2 - x)$ is maximized when $x = 3n/4$, where it is equal to $9n^2/16$. This proves that when $\alpha = \beta + 1$, $\tilde{P}_3(G) \leq n^3/16$.

Thus we may assume $\alpha = \beta$. Lemma 12 now gives $3n/8 - 1/4 \leq \beta \leq 3n/8 + 1/4$, or $3n - 2 \leq 8\beta \leq 3n + 2$. Theorem 5 then follows directly from equation (14), together with Lemmas 19 and 20. $\square$

4 Four-Vertex Paths in 3-free Digraphs

The main result of this section is:

Theorem 21. *If G is a 3-free digraph on n vertices, then $P_4(G) \leq \frac{4}{75}n^4$.*

We first establish some notation and two key lemmas.

Let G be a 3-free digraph on n vertices. We will use the term *square* to refer to a subgraph of G which is a directed cycle of length four. If $X \subseteq V(G)$ with $|X| = 4$, let $t(X)$ be the number of 4-vertex directed paths with vertex set X. We observe that since G is 3-free, $t(X) \in \{0, 1, 4\}$ for every such X. This motivates the following definitions. Let R be the number of four-tuples of distinct vertices (a, b, c, d) such that $t(\{a, b, c, d\}) = 1$. Let S be the number of four-tuples of distinct vertices (a, b, c, d) such that $G|\{a, b, c, d\}$ is a square (equivalently, $t(\{a, b, c, d\}) = 4$). Then S is 24 times the number of squares. Define N to be the set of four-tuples of vertices not counted by either R or S, so $|N| = n^4 - R - S$. For distinct vertices u, v, let $M(u, v)$ be the set of all vertices x such that (u, x, v) is an induced 3-vertex path. Set $m(u, v) = |M(u, v)|$, the number of induced directed 3-vertex paths starting at u and ending at v. Finally, define $T = \tilde{P}_3(G)$.

Lemma 22. In a 3-free digraph G, $S \leq \frac{3n}{2}T$.

Proof. We will write $P \sqsubset G$ to mean P is a (directed) path of G, and $\sum_P, \sum_\Gamma$ to mean the sum over all induced 3-vertex paths in G and the sum over all squares in G, respectively. For each square $\Gamma = a$-b-c-d-a in G, define

$$\omega(\Gamma) = \frac{1}{m(c, a)} + \frac{1}{m(d, b)} + \frac{1}{m(a, c)} + \frac{1}{m(b, d)}.$$

Now, since $m(a,c)+m(c,a)+m(b,d)+m(d,b) \le n$ (each path has a middle vertex, and no vertex can serve as the middle of two of the paths counted since G has no directed cycle of length at most three), $\omega(\Gamma) \ge 16/n$ for all Γ. Since there are $S/24$ squares, it follows that

$$\sum_{\Gamma} \omega(\Gamma) \ge \frac{16}{n}\left(\frac{S}{24}\right) = \frac{2}{3n}S.$$

For an induced 3-vertex path $P = u$-w-v in G, let

$$\omega(P) = \frac{1}{m(v,u)} \left|\{\text{squares } \Gamma : P \sqsubset \Gamma\}\right|.$$

We claim that $\omega(P) = 1$ for all P. The squares containing P are of the form u-w-v-x-u where (v,x,u) is also an induced 3-vertex path. Since G is 3-free, every 4-cycle is induced, so every choice of $x \in M(v,u)$ gives a square, proving $\omega(P) = m(v,u) \cdot \frac{1}{m(v,u)} = 1$. Then $\sum_P \omega(P) = \sum_P 1 = T$ by definition.

Finally, we show $\sum_P \omega(P) = \sum_\Gamma \omega(\Gamma)$. Below, let P be u-w-v. Then

$$\begin{aligned}\sum_P \omega(P) &= \sum_P \frac{1}{m(v,u)} \left|\{\text{squares } \Gamma : P \sqsubset \Gamma\}\right| \\ &= \sum_P \sum_{\Gamma \sqsupset P} \frac{1}{m(u,v)} = \sum_\Gamma \sum_{P \sqsubset \Gamma} \frac{1}{m(u,v)} = \sum_\Gamma \omega(\Gamma).\end{aligned}$$

We now have $T = \sum_P \omega(P) = \sum_\Gamma \omega(\Gamma) \ge \frac{2}{3n}S$, or $S \le \frac{3n}{2}T$. This proves Lemma 22. □

Lemma 23. If G is a 3-free digraph, then $|N| \ge \frac{2}{3}S$.

Proof. Let $\Gamma = a$-b-c-d-a be a square in G. Define

$$\omega(\Gamma) = 2\left(\frac{(m(b,d)+m(d,b))^2}{m(a,c)m(c,a)} + \frac{(m(a,c)+m(c,a))^2}{m(b,d)m(d,b)}\right).$$

Again, $m(a,c)+m(c,a)+m(d,b)+m(b,d) \le n$, and by Cauchy-Schwarz,

$\omega(\Gamma) \geq 16$ (since we know that $m(u, v) > 0$ for each relevant u, v). Since there are $S/24$ squares, we have

$$\frac{2}{3}S \leq \sum_{\Gamma} \omega(\Gamma). \tag{15}$$

Given a four-tuple of vertices $\pi = (p, q, r, s)$ and a square Γ, we say they are *associated*, and write $\pi \sim \Gamma$, if there exist vertices u, v such that $\Gamma = p\text{-}u\text{-}q\text{-}v\text{-}p$ and $r, s \in M(u, v) \cup M(v, u)$. Note that for a square $\Gamma = a\text{-}b\text{-}c\text{-}d\text{-}a$, the four-tuples associated with it are precisely those of the forms (a, c, x, y) or (c, a, x, y) where $x, y \in M(d, b) \cup M(b, d)$, and (b, d, x, y) or (d, b, x, y) with $x, y \in M(a, c) \cup M(c, a)$.

Now, for a four-tuple of vertices $\pi = (p, q, r, s)$, define $\omega(\pi)$ as follows:

$$\omega(\pi) = \frac{|\{\Gamma : \Gamma \sim \pi\}|}{m(p, q)m(q, p)}.$$

Note that $\omega(\pi) \leq 1$, since the number of squares associated with π is at most $m(p, q)m(q, p)$ by definition. Then

$$\sum_{\pi \in N} \omega(\pi) \leq \sum_{\pi \in N} 1 \leq |N|. \tag{16}$$

Next, if $\Gamma = a\text{-}b\text{-}c\text{-}d\text{-}a$ is a square in G, we show that $\pi \sim \Gamma$ implies $\pi \in N$. Without loss of generality, we may let $\pi = (a, c, x, y)$. We need to show that there is no 4-vertex path with vertex set $\{a, c, x, y\}$. This is clear if a, c, x, y are not all distinct, so we assume they are distinct. Since b is adjacent to every vertex in $M(b, d)$ and from every vertex in $M(d, b)$, there is no edge from $M(b, d)$ to $M(d, b)$, since otherwise there would be a directed triangle. Similarly, there is no edge from a vertex in $M(d, b)$ to a vertex in $M(b, d)$. Consequently, if X is a set of four vertices so that $G|X$ has a 4-vertex path as a subgraph and $X \subseteq M(b, d) \cup M(b, d)$, then $X \subseteq M(b, d)$ or $X \subseteq M(b, d)$. So not both of a, c are in X. This proves that every π associated with Γ belongs to N.

This observation allows us to relate $\sum_\Gamma \omega(\Gamma)$ to $\sum_{\pi \in N} \omega(\pi)$. Assuming $\pi = (p, q, r, s)$ for the purposes of writing $\omega(\pi)$,

$$\sum_\Gamma \omega(\Gamma) = \sum_\Gamma \sum_{\pi \sim \Gamma} \frac{1}{m(p,q)m(q,p)} = \sum_{\pi \in N} \sum_{\Gamma \sim \pi} \frac{1}{m(p,q)m(q,p)} = \sum_{\pi \in N} \omega(\pi).$$

Combining this with (15) and (16), we have

$$\frac{2S}{3} \leq \sum_\Gamma \omega(\Gamma) = \sum_{\pi \in N} \omega(\pi) \leq |N|.$$

This proves Lemma 23. $\square$

Proof of Theorem 21: Note that $n^4 = R + S + |N|$ by definition. We can also express the number of 4-vertex paths $P_4(G)$ in terms of these parameters, as $24P_4(G) = 4S + R$. Combining these equalities, we write

$$24P_4(G) = n^4 + 3S - |N|. \tag{17}$$

To prove an upper bound for $P_4(G)$, it then suffices to bound S from above and $|N|$ from below. From Lemmas 22 and 23, we have $S \leq \frac{3n}{2}T$ and $|N| \geq \frac{2}{3}S$. Combining these with (17), we see that:

$$24P_4(G) \leq n^4 + \frac{7}{3}S \leq n^4 + \frac{7}{2}nT.$$

But $T \leq \frac{2}{25}n^3$ by Theorem 4, and so $24P_4(G) \leq (1+7/25)n^4$, or $P_4(G) \leq \frac{4}{75}n^4$, as desired. $\square$

It follows immediately from Theorem 21 that every 3-free digraph on n vertices has a vertex of out-degree at most $\sqrt[3]{4/75}n \approx .3764n$. Note that if Conjecture

3 holds, we could replace Bondy's bound on P_3 by $n^3/15$, and the proof of Theorem 21 would then give $P_4(G) < \frac{1}{19.45}n^4 \approx .0514n^4$, implying the existence of a vertex with out-degree at most $\sqrt[3]{\frac{1}{19.45}}n \approx .37184n$.

References

[1] L. Caccetta and R. Häggkvist, "On minimal digraphs with given girth," *Proceedings of the Ninth Southeastern Conference on Combinatorics, Graph Theory and Computing* (Florida Atlantic University, Boca Raton, Florida, 1978), Congressus Numerantium XXI, Utilitas Math, 181-187, 1978.

[2] P. Hamburger and P. Haxell and A. Kostochka, "On Directed Triangles in Digraphs," Electronic Journal of Combinatorics, 14(1), 2007.

[3] J. Shen, "Paths in 3-free Digraphs," personal communication, 2006.