

IMPROVED LOWER BOUNDS FOR THE MAXIMUM ORDER OF AN INDUCED ACYCLIC SUBGRAPH

SHAMIL ASGARLI, DONALD FALKENHAGEN, AND KAYA HOSHI

ABSTRACT. Computing the cardinality of a maximum induced acyclic vertex set in a digraph is NP-hard. Since finding an exact solution is computationally difficult, a fruitful approach is to establish high-quality lower bounds that are efficiently computable. We build on the Akbari–Ghodrati–Jabalameli–Saghafian (AGJS) bound for digraphs by adapting refinement techniques used by (a) Selkow and Harant–Mohr and (b) Angel–Campigotto–Laforest in their respective improvements of the Caro–Wei bound for undirected graphs. First, inspired by (a), we prove a neighborhood-based refinement of the AGJS bound that incorporates local degree data of each vertex. Second, inspired by (b), we compute the variance of the size of a feedback vertex set returned by a randomized algorithm. This result, combined with the Bhatia–Davis inequality, yields a tighter lower bound than the AGJS bound.

1. INTRODUCTION

Finding the independence number $\alpha(G)$ of an undirected graph (the size of a maximum set of pairwise nonadjacent vertices) is a foundational problem in extremal graph theory. Computing $\alpha(G)$ is one of the classic NP-hard problems. Caro [4] and Wei [9] independently discovered an elegant lower bound in terms of the degree sequence:

$$(1.1) \quad \alpha(G) \geq \sum_{v \in V(G)} \frac{1}{1 + d(v)}.$$

The standard proof is a direct application of the probabilistic method: consider a uniformly random permutation π of the vertices $V(G)$. Construct an independent set I by including every vertex v that appears before all of its neighbors in π . By linearity of expectation, the expected size of this set is precisely the sum on the right-hand side of (1.1). Since the maximum size is at least the expected size of a randomly generated instance, the bound (1.1) follows.

The Caro–Wei bound inspired more sophisticated bounds:

- **Neighborhood-based refinements:** Selkow [8] proposed strengthening the Caro–Wei bound by applying it again to the residual graph formed by vertices at distance at least two from the initial independent set I . Harant and Mohr [6] identified a subtle error in Selkow’s argument and provided a correct proof of the following bound:

$$\alpha(G) \geq \sum_{v \in V(G)} \frac{1}{1 + d(v)} \left(1 + \max \left\{ 0, \frac{d(v)}{d(v) + 1} - \sum_{u \in N(v)} \frac{1}{1 + d(u)} \right\} \right).$$

The improvement is based on the local degree distribution around each vertex.

- **Variance-based refinements:** Another approach examines the second moment of the size of a randomly generated set. Angel, Campigotto, and Laforest [2] analyzed a random process (the LL algorithm) for producing vertex covers. This is complementary to constructing

2020 *Mathematics Subject Classification.* Primary 05C69; Secondary 05C20, 05C85, 05D40.

Key words and phrases. maximum induced acyclic subgraph; acyclic set; feedback vertex set; Caro–Wei bound; probabilistic method; Bhatia–Davis inequality.

independent sets. By applying the Bhatia–Davis inequality [3] to the size of the random vertex cover S generated by this algorithm, they derived the bound:

$$\alpha(G) \geq \left(\sum_{v \in V(G)} \frac{1}{1 + d(v)} \right) + \frac{\text{Var}(|S|)}{\left(\sum_{v \in V(G)} \frac{1}{1 + d(v)} \right) - c},$$

where $\text{Var}(|S|)$ is the variance of the random cover size and c is the number of connected components of G .

Both refinements go beyond the first-moment method and use additional graph structure to tighten the bound (1.1).

Extension to digraphs. These concepts extend naturally to directed graphs, where the analog of an independent set is an *acyclic set* (a set of vertices that induces a subgraph with no directed cycles). The cardinality of the largest acyclic set is denoted by $\vec{\alpha}(D)$. Computing $\vec{\alpha}(D)$ is also NP-hard. Akbari, Ghodrati, Jabalameli, and Saghafian (AGJS) [1] established the following directed analog of the Caro–Wei bound:

$$(1.2) \quad \vec{\alpha}(D) \geq \sum_{v \in V(D)} \rho_D(v).$$

The term $\rho_D(v)$ arises from a probabilistic argument on a random vertex permutation. A vertex v is selected to be in the acyclic set if it appears before all of its out-neighbors (event A_v) *or* before all of its in-neighbors (event B_v). The probability of this union event is computed by:

$$(1.3) \quad \rho_D(v) := \mathbb{P}(A_v \cup B_v) = \mathbb{P}(A_v) + \mathbb{P}(B_v) - \mathbb{P}(A_v \cap B_v) = \frac{1}{1 + d_v^+} + \frac{1}{1 + d_v^-} - \frac{1}{1 + d_v},$$

where d_v^+ , d_v^- , and d_v are the out-degree, in-degree, and total degree of the vertex v , respectively (see Preliminaries for precise definitions). Crucially, the set of vertices satisfying the event $A_v \cup B_v$ forms a random acyclic set. We include a self-contained proof in Section 2 and use this construction as our starting point. The bound (1.2) is the first-order estimate for $\vec{\alpha}(D)$, playing the same role as the Caro–Wei bound does for $\alpha(G)$. For comparison, Gruber [5] proved the weaker bound:

$$\vec{\alpha}(D) \geq \sum_{v \in V(D)} \frac{1}{1 + d_v^+}.$$

Our contributions. To the best of our knowledge, the aforementioned refinements for the Caro–Wei bound have not been extended to the directed graph setting. This paper fills that gap. We obtain both the neighborhood-based and variance-based improvements, yielding two new lower bounds on $\vec{\alpha}(D)$ that strengthen the AGJS bound (1.2).

- **Neighborhood-based approach.** Following Selkow’s idea [8] (with the correction of Harant–Mohr [6]), we prove a neighborhood-based improvement for digraphs by applying the bound (1.2) to a random residual subgraph. See Theorem 2.2.
- **Variance-based approach.** Inspired by Angel–Campigotto–Laforest [2], we define the *DL algorithm*, a randomized procedure for generating a feedback vertex set S . We show that the probability a vertex is *not* in S equals $\rho_D(v)$, thus linking our algorithm to the bound (1.2). We then derive an exact formula for $\text{Var}(|S|)$ via a detailed covariance analysis and apply the Bhatia–Davis inequality [3] to obtain our second refinement. See Theorem 5.1.

Preliminaries. Before proceeding, we formally define the terminology used throughout this paper. We work with finite simple loopless digraphs: for any ordered pair of vertices (u, v) , at most one arc $u \rightarrow v$ is allowed.¹ In particular, both arcs $u \rightarrow v$ and $v \rightarrow u$ may be present simultaneously. Given a digraph D , we write $V(D)$ and $A(D)$ for the vertex and arc sets, respectively.

For a vertex v , its in-neighborhood, out-neighborhood, and neighborhood are denoted $N^-(v) := \{w \mid (w, v) \in A(D)\}$, $N^+(v) := \{w \mid (v, w) \in A(D)\}$, and $N(v) := N^-(v) \cup N^+(v)$, respectively. The *in-degree* d_v^- and *out-degree* d_v^+ are the cardinalities of these neighborhoods. We define the *total degree* of v , denoted d_v , as the size of its neighborhood: $d_v := |N(v)|$. For conciseness, we use the subscript notation d_v rather than $d(v)$. Note that this definition may differ from the sum $d_v^- + d_v^+$, since we allow directed 2-cycles (that is, a vertex u may belong to both $N^-(v)$ and $N^+(v)$). Defining d_v as the neighborhood size (rather than the number of incident arcs) ensures that the probabilistic formulas in this paper are valid for all simple digraphs. For a set of vertices $S \subseteq V(D)$, its neighborhood is:

$$N(S) := \bigcup_{x \in S} N(x).$$

The *weakly connected components* of D are the connected components of its underlying undirected graph; for brevity, we will simply call them *connected components*.

A vertex subset $T \subseteq V(D)$ is called an *acyclic set* if the induced subgraph $D[T]$ contains no directed cycle. The maximum size of such a set is denoted by $\vec{\alpha}(D)$. A *feedback vertex set* (FVS) is a set of vertices S such that $D - S$ is acyclic; the minimum size of such a set is denoted $\vec{\beta}(D)$. These two complementary parameters satisfy the identity

$$(1.4) \quad \vec{\alpha}(D) + \vec{\beta}(D) = |V(D)|.$$

Terminology. In some sources, “Maximum Acyclic Subgraph” refers to an *arc*-deletion problem; throughout, we study the *vertex*-induced version.

The *dichromatic number*, introduced by Neumann-Lara [7], is the minimum number of acyclic color classes needed to partition $V(D)$. While we do not study colorings here, the quantity $\vec{\alpha}(D)$ is an extremal parameter in that theory (that is, the maximum possible size of a color class), analogous to the independence number in traditional graph coloring.

Outline of the paper. The remainder of the paper is organized as follows. Section 2 develops the neighborhood-based lower bound for $\vec{\alpha}(D)$. Section 3 introduces our DL algorithm and proves its fundamental properties. Section 4 provides the variance decomposition and a complete covariance catalog for the DL algorithm. Section 5 assembles the variance-based lower bound for $\vec{\alpha}(D)$. Finally, Section 6 offers numerical experiments on certain types of random graphs to illustrate the behavior of our improvements.

2. A NEIGHBORHOOD-BASED REFINEMENT FOR DIGRAPHS

The main idea of the neighborhood-based improvement is to apply the AGJS bound (1.2) to a *residual subgraph*, namely, the subgraph that remains after removing an initial acyclic set and its neighbors. Since we consider degrees in a subgraph, we first need to understand how the quantity $\rho_D(v)$ from (1.3) behaves as degrees decrease.

Lemma 2.1 (Monotonicity). *Consider the function*

$$f(x, y, z) := \frac{1}{1+x} + \frac{1}{1+y} - \frac{1}{1+x+y-z}.$$

on the domain $x \geq 0, y \geq 0, 0 \leq z \leq \min(x, y)$. Suppose a, b, c, d, s, t are nonnegative integers such that $a \leq c, b \leq d$, and $s \leq t$. Then $f(a, b, s) \geq f(c, d, t)$.

¹We use the term *arc* for a directed edge; an arc (u, v) means the edge $u \rightarrow v$.

Proof. We compute the partial derivatives:

$$\begin{aligned}\frac{\partial f}{\partial x} &= -\frac{1}{(1+x)^2} + \frac{1}{(1+x+y-z)^2} \leq 0, \\ \frac{\partial f}{\partial y} &= -\frac{1}{(1+y)^2} + \frac{1}{(1+x+y-z)^2} \leq 0. \\ \frac{\partial f}{\partial z} &= -\frac{1}{(1+x+y-z)^2} \leq 0.\end{aligned}$$

Thus, f is nonincreasing in each variable. Since $a \leq c$, $b \leq d$, and $s \leq t$, we conclude that $f(a, b, s) \geq f(c, d, t)$. $\square$

The function in Lemma 2.1 matches $\rho_D(v)$ in (1.3). For a vertex v , we have $d_v = |N(v)| = d_v^+ + d_v^- - |N^+(v) \cap N^-(v)|$ since by convention we count a neighbor once, even if it is both an in- and an out-neighbor. Setting $x = d_v^+$, $y = d_v^-$, and $z = |N^+(v) \cap N^-(v)|$ gives $\rho_D(v) = f(x, y, z)$.

We now prove one of the main results of our paper.

Theorem 2.2 (Neighborhood-based refinement). *Let D be a digraph and let ρ_D be the function defined in (1.3). Then*

$$(2.1) \quad \vec{\alpha}(D) \geq \sum_{v \in V(D)} \rho_D(v) \left(1 + \max \left\{ 0, 1 - \rho_D(v) - \sum_{u \in N(v)} \rho_D(u) \right\} \right).$$

Proof. Let π be a uniformly random permutation of $V(D)$. For each $v \in V(D)$, define the events:

$$A_v := \text{all } N^+(v) \text{ appear after } v \text{ in } \pi, \quad B_v := \text{all } N^-(v) \text{ appear after } v \text{ in } \pi.$$

Let I be the random set of vertices for which the event $A_v \cup B_v$ occurs.

We claim that I is acyclic. We show that no directed cycle can have all its vertices in I . Let $C = (v_1 \rightarrow v_2 \rightarrow \cdots \rightarrow v_k \rightarrow v_1)$ be an arbitrary directed cycle in the graph D . Consider the vertex on this cycle, say v_j , that appears *last* in the random permutation π . We show that $v_j \notin I$.

For v_j to be included in I , the event $A_{v_j} \cup B_{v_j}$ must occur. However, both events are impossible:

- The event A_{v_j} requires all out-neighbors of v_j to appear *after* v_j in π . But the successor of v_j on the cycle, v_{j+1} (indices modulo k), is an out-neighbor that must appear *before* v_j in the permutation, since v_j is the last vertex of the cycle to appear in π . Thus, A_{v_j} cannot occur.
- The event B_{v_j} requires all in-neighbors of v_j to appear *after* v_j in π . But the predecessor of v_j on the cycle, v_{j-1} (indices modulo k), is an in-neighbor that must also appear *before* v_j in the permutation, since v_j is the last vertex of the cycle to appear in π . Thus, B_{v_j} cannot occur.

Thus, the union event $A_{v_j} \cup B_{v_j}$ cannot occur, meaning $v_j \notin I$. We have shown that every directed cycle C in the graph has at least one vertex that is not in I . Therefore, I does not induce any directed cycles, that is, I is acyclic.

Next, we apply linearity of expectation and formula (1.3) for $\mathbb{P}(A_v \cup B_v)$ to obtain:

$$(2.2) \quad \mathbb{E}[|I|] = \sum_{v \in V(D)} \mathbb{P}(A_v \cup B_v) = \sum_{v \in V(D)} \rho_D(v).$$

Now, define the random residual subgraph H by

$$V(H) = V(D) \setminus (I \cup N(I)).$$

For any π , the sets I and $V(H)$ are disjoint and there are no arcs between them. Thus, the union of an acyclic set from $D[I]$ and an acyclic set from H is acyclic in D . This yields the inequality:

$$\vec{\alpha}(D) \geq |I| + \vec{\alpha}(H).$$

Taking expectations, we have

$$(2.3) \quad \vec{\alpha}(D) \geq \mathbb{E}[|I|] + \mathbb{E}[\vec{\alpha}(H)] = \sum_{v \in V(D)} \rho_D(v) + \mathbb{E}[\vec{\alpha}(H)].$$

We now bound $\mathbb{E}[\vec{\alpha}(H)]$ from below by applying the bound (1.2) to the random subgraph H . Observe that $\rho_D(v)$ can be expressed as:

$$\rho_D(v) = \frac{1}{1 + d_v^+} + \frac{1}{1 + d_v^-} - \frac{1}{1 + d_v^+ + d_v^- - t_v}$$

where $t_v = |N^+(v) \cap N^-(v)|$. For any vertex $v \in V(H)$, its in- and out-degree in H do not exceed its in- and out-degree in D ; moreover, $|N_H^+(v) \cap N_H^-(v)| \leq |N_D^+(v) \cap N_D^-(v)|$. Therefore, by Lemma 2.1, we have $\rho_H(v) \geq \rho_D(v)$. Thus,

$$(2.4) \quad \mathbb{E}[\vec{\alpha}(H)] \geq \mathbb{E}\left[\sum_{v \in V(H)} \rho_H(v)\right] \geq \mathbb{E}\left[\sum_{v \in V(H)} \rho_D(v)\right] = \sum_{v \in V(D)} \rho_D(v) \cdot \mathbb{P}(v \in V(H)).$$

Combining (2.3) and (2.4) gives

$$(2.5) \quad \vec{\alpha}(D) \geq \sum_{v \in V(D)} \rho_D(v) \left(1 + \mathbb{P}(v \in V(H))\right).$$

The final step is to bound $\mathbb{P}(v \in V(H))$ from below. A vertex v is *not* in $V(H)$ if and only if $v \in I \cup N(I)$. Using the union bound,

$$\mathbb{P}(v \notin V(H)) \leq \mathbb{P}(v \in I) + \sum_{u \in N(v)} \mathbb{P}(u \in I) = \rho_D(v) + \sum_{u \in N(v)} \rho_D(u).$$

This implies

$$(2.6) \quad \mathbb{P}(v \in V(H)) = 1 - \mathbb{P}(v \notin V(H)) \geq 1 - \rho_D(v) - \sum_{u \in N(v)} \rho_D(u).$$

Combining the trivial inequality $\mathbb{P}(v \in V(H)) \geq 0$ with (2.6), we obtain:

$$(2.7) \quad \mathbb{P}(v \in V(H)) \geq \max \left\{ 0, 1 - \rho_D(v) - \sum_{u \in N(v)} \rho_D(u) \right\}.$$

Finally, substituting the bound (2.7) for $\mathbb{P}(v \in V(H))$ back into our main inequality (2.5) for $\vec{\alpha}(D)$ yields

$$\vec{\alpha}(D) \geq \sum_{v \in V(D)} \rho_D(v) \left(1 + \max\{0, 1 - \rho_D(v) - \sum_{u \in N(v)} \rho_D(u)\}\right),$$

which is precisely the inequality (2.1) claimed in the theorem. $\square$

3. A VARIANCE-BASED APPROACH VIA FEEDBACK VERTEX SETS

Our next goal is to develop a variance-based refinement in the spirit of Angel, Campigotto, and Laforest (ACL) [2]. This requires a random variable whose expectation recovers the AGJS bound (1.2) from Akbari–Ghodrati–Jabalameli–Saghafian [1]. Our strategy is to construct such a variable by analyzing the problem from a complementary perspective: instead of building a random *acyclic set*, we define a randomized procedure that generates a feedback vertex set (FVS).

To implement our strategy, this section introduces the DL algorithm.² The analysis is organized as follows. In §3.1, we formally define the algorithm and prove its correctness (that it always produces an FVS). In §3.2, we establish its key structural properties, showing that it produces optimal solutions and deriving a worst-case bound on its output size. Finally, in §3.3, we compute the expected FVS size and show that it recovers the AGJS bound (1.2).

3.1. The DL algorithm and its correctness. Our algorithm applies a simple local rule to the vertices based on a random permutation; we formalize this as a labeling. Throughout this section, we define a *labeling* as a bijection $L: V(D) \rightarrow \{1, \dots, n\}$ where $n = |V(D)|$ is the number of vertices. We imagine a labeling as arranging the vertices in a line, with labels increasing from left to right.

Definition 3.1 (Right neighbors for a labeling). Given a labeling $L: V(D) \rightarrow \{1, \dots, n\}$, a neighbor w of a vertex u is called a *right in-neighbor* if $w \rightarrow u$ is an arc and $L(u) < L(w)$. Similarly, w is a *right out-neighbor* if $u \rightarrow w$ is an arc and $L(u) < L(w)$.

We denote the set of all right in-neighbors and right out-neighbors of u by $R^-(u)$ and $R^+(u)$, respectively. Formally, these sets are given by:

$$R^-(u) := \{w \in N^-(u) \mid L(u) < L(w)\}$$

and

$$R^+(u) := \{w \in N^+(u) \mid L(u) < L(w)\}.$$

The DL algorithm uses the existence of both types of right neighbors to decide which vertices to place in its output set.

Definition 3.2 (DL algorithm). Let L be a uniformly random labeling of $V(D)$. The *DL algorithm* outputs a set $S \subseteq V(D)$ defined by the following rule for each vertex u :

$$u \in S \iff (\exists \text{ a right in-neighbor of } u) \text{ and } (\exists \text{ a right out-neighbor of } u).$$

Algorithm 1 DL: A randomized FVS procedure (input: digraph D)

- 1: Let $L: V(D) \rightarrow \{1, \dots, n\}$ be a uniform random labeling.
 - 2: $S \leftarrow \emptyset$.
 - 3: **for all** $u \in V(D)$
 - 4: **if** $(\exists w \in N^-(u) : L(u) < L(w))$ **and** $(\exists w \in N^+(u) : L(u) < L(w))$ **then**
 - 5: $S \leftarrow S \cup \{u\}$.
 - 6: **return** S .
-

First, we confirm that the procedure is correct: its output is always a feedback vertex set.

Lemma 3.3 (DL produces an FVS). *For any labeling L , the set S returned by the DL algorithm is a feedback vertex set.*

²The name “DL algorithm” is chosen as a counterpart to the “LL algorithm” from the undirected setting [2]; “DL” is an abbreviation for **D**irected **L**abeling.

Proof. We show directly that the subgraph $D - S$ is acyclic. It suffices to show that for any directed cycle γ in the original graph D , at least one of its vertices must belong to S .

Let γ be an arbitrary directed cycle in D . Consider the vertex on this cycle, say v , that has the *minimal* label under the given labeling L . Since v is part of a cycle, it must have an in-neighbor x and an out-neighbor y that are also on γ . By our choice of v as the vertex with the minimum label on the cycle, it must be that $L(v) < L(x)$ and $L(v) < L(y)$.

- Since $x \rightarrow v$ is an arc and $L(v) < L(x)$, it follows that x is a right in-neighbor of v .
- Since $v \rightarrow y$ is an arc and $L(v) < L(y)$, it follows that y is a right out-neighbor of v .

By the definition of the DL algorithm, any vertex that has both a right in-neighbor and a right out-neighbor is included in the set S . Therefore, $v \in S$.

Since we have shown that every directed cycle in D contains at least one vertex that must be in S , no cycle can exist within the complement $D - S$. Thus, $D - S$ is acyclic, and S is a feedback vertex set. $\square$

3.2. Structural properties of the DL algorithm. The DL algorithm can produce a minimum FVS. The proof of this property relies on the following elementary lemma regarding the structure of *minimal* feedback vertex sets, which we state here for completeness.

Lemma 3.4 (Criticality of minimal FVS). *If S^* is a minimal feedback vertex set (with respect to inclusion), then for every $v \in S^*$, the digraph $D - (S^* \setminus \{v\})$ contains a directed cycle that passes through v .*

Proof. Let $v \in S^*$. By the minimality of S^* , the set $S' := S^* \setminus \{v\}$ is not a feedback vertex set. This means the subgraph $D - S'$ must contain at least one directed cycle. Any such cycle must necessarily pass through the vertex v , because all other vertices of S^* are still in S' , and any cycle not involving v would have already existed in the acyclic graph $D - S^*$. $\square$

With these preliminaries, we can now establish the fundamental properties of the DL algorithm. We begin by showing that the algorithm is powerful enough, in principle, to find an optimal solution.

Proposition 3.5 (Optimality of the DL algorithm). *For any digraph D , there exists a labeling L^* for which the DL algorithm returns a minimum feedback vertex set.*

Proof. The proof is constructive: given a minimum feedback vertex set S^* , we construct a specific labeling L^* and show that the DL algorithm with this labeling returns exactly the set S^* . To define the labeling L^* , we partition the vertices $V(D)$ into two sets, S^* and its complement $V(D) \setminus S^*$, and assign labels as follows:

- Vertices $v \in S^*$ are assigned the smallest available labels, that is,

$$L^*(v) \in \{1, 2, \dots, |S^*|\}.$$

- The remaining vertices $u \in V(D) \setminus S^*$ induce an acyclic subgraph. We assign them the labels from $\{|S^*| + 1, \dots, n\}$ according to a topological sort, which guarantees that $L^*(x) < L^*(y)$ for any arc $x \rightarrow y$ induced by $V(D) \setminus S^*$.

We analyze the output S of the DL algorithm with the labeling L^* by considering the two cases.

Case 1: A vertex $u \notin S^*$.

We must show that DL does not include u in its output. It suffices to show that u has no right in-neighbor with respect to the labeling L^* . Consider any in-neighbor $w \in N^-(u)$. There are two possibilities for w :

- If $w \in S^*$, then by our construction $L^*(w) \leq |S^*| < L^*(u)$. Thus, w is not a right in-neighbor of u .
- If $w \notin S^*$, then the arc $w \rightarrow u$ is induced by the acyclic set $V(D) \setminus S^*$. By the property of the topological sort, we must have $L^*(w) < L^*(u)$. Thus, w is not a right in-neighbor of u .

In either case, u has no right in-neighbors. Therefore, the algorithm does not add u to S .

Case 2: A vertex $v \in S^*$.

We must show that DL adds v to its output. By Lemma 3.4, since S^* is a minimal FVS, there exists a directed cycle in $D - (S^* \setminus \{v\})$ that passes through v . Let $u \rightarrow v \rightarrow w$ be arcs on such a cycle. The vertices u and w are not in $S^* \setminus \{v\}$. Since u and w are neighbors of v , we infer that $u \neq v$ and $w \neq v$, which means $u, w \notin S^*$. Accordingly, our labeling construction guarantees the following ordering:

$$L^*(v) \leq |S^*| \quad \text{and} \quad L^*(u), L^*(w) > |S^*|.$$

This directly implies that $L^*(v) < L^*(u)$ and $L^*(v) < L^*(w)$. Therefore, u is a right in-neighbor of v ; similarly, w is a right out-neighbor of v . As a result, the algorithm adds v to S .

Since the output S contains all vertices in S^* and none from its complement, we have $S = S^*$. Thus, the DL algorithm can indeed produce a minimum feedback vertex set. $\square$

Next, we establish a deterministic upper bound on the size of the output set S , which holds for any random labeling and depends on the component structure of the graph.

Proposition 3.6 (Worst-case bound on FVS size). *Let D be a digraph with n vertices and c connected components. For any labeling, the output S of the DL algorithm satisfies $|S| \leq n - c$.*

Proof. We establish a lower bound on $|V(D) \setminus S|$ by inspecting the highest-labeled vertex in each component. Consider any connected component W . We show that at least one vertex of W does not belong to S . Let $v_{\max}$ be the vertex in W with the largest label assigned by L . By definition, $v_{\max}$ has no neighbors with a larger label, so it has no right neighbors at all. Therefore, $v_{\max} \notin S$. Since at least c vertices are not in S , we deduce that $|S| \leq n - c$. $\square$

3.3. Recovering the AGJS bound. We now show that the probability a vertex is excluded by the DL algorithm is precisely the term $\rho_D(v)$ from the AGJS bound (1.2).

Proposition 3.7 (DL probability identity). *For any vertex $v \in V(D)$,*

$$\mathbb{P}(v \notin S) = \frac{1}{1 + d_v^+} + \frac{1}{1 + d_v^-} - \frac{1}{1 + d_v} = \rho_D(v).$$

Proof. By the definition of the DL algorithm, a vertex v is *not* included in S if and only if it has no right in-neighbors or no right out-neighbors.

Let E_v^- be the event that v has no right in-neighbors (i.e., all vertices in $N^-(v)$ appear before v in the random labeling L). Let E_v^+ be the event that v has no right out-neighbors (i.e., all vertices in $N^+(v)$ appear before v in L).

We wish to compute $\mathbb{P}(v \notin S) = \mathbb{P}(E_v^- \cup E_v^+)$. By the principle of inclusion-exclusion, this is:

$$\mathbb{P}(v \notin S) = \mathbb{P}(E_v^-) + \mathbb{P}(E_v^+) - \mathbb{P}(E_v^- \cap E_v^+).$$

Each of these probabilities depends only on the relative ordering of v and its neighbors. In a uniform random permutation:

- $\mathbb{P}(E_v^-)$ is the probability that v is the last element among the $d_v^- + 1$ vertices in $N^-(v) \cup \{v\}$. Thus, $\mathbb{P}(E_v^-) = \frac{1}{1 + d_v^-}$.
- $\mathbb{P}(E_v^+)$ is the probability that v is the last element among the $d_v^+ + 1$ vertices in $N^+(v) \cup \{v\}$. Thus, $\mathbb{P}(E_v^+) = \frac{1}{1 + d_v^+}$.
- $\mathbb{P}(E_v^- \cap E_v^+)$ is the probability that v is the last element among all $d_v + 1$ vertices in $N(v) \cup \{v\}$. Thus, $\mathbb{P}(E_v^- \cap E_v^+) = \frac{1}{1 + d_v}$.

Substituting these values into the inclusion-exclusion formula gives the identity:

$$\mathbb{P}(v \notin S) = \frac{1}{1 + d_v^+} + \frac{1}{1 + d_v^-} - \frac{1}{1 + d_v} = \rho_D(v). \quad \square$$

Corollary 3.8 (Expected FVS size). *The expected size of the set S returned by the DL algorithm is*

$$\mathbb{E}[|S|] = n - \sum_{v \in V(D)} \rho_D(v).$$

Proof. By linearity of expectation and Proposition 3.7, we have:

$$\mathbb{E}[|S|] = \sum_{v \in V(D)} \mathbb{P}(v \in S) = \sum_{v \in V(D)} (1 - \mathbb{P}(v \notin S)) = n - \sum_{v \in V(D)} \rho_D(v). \quad \square$$

This result provides an alternative derivation of the AGJS bound (1.2). The expected FVS size from Corollary 3.8 gives an upper bound on the minimum FVS size, $\vec{\beta}(D)$. Using the identity $\vec{\alpha}(D) = n - \vec{\beta}(D)$, this translates directly to the familiar lower bound on the acyclic number:

$$\vec{\alpha}(D) = n - \vec{\beta}(D) \geq n - \mathbb{E}[|S|] = \sum_{v \in V(D)} \rho_D(v).$$

Thus, the first-order bound (1.2) is recovered, which confirms that our FVS-based approach is correctly calibrated. This consistency sets the stage for the variance-based second-order refinement developed in the next section.

4. THE VARIANCE OF THE DL ALGORITHM

Recall that the output of the DL algorithm is a random feedback vertex set S . The previous section computed a formula for the expectation of the random variable $|S|$. To develop our variance-based refinement, we move beyond the first moment of $|S|$ and derive an explicit formula for its variance. We begin by expressing $|S|$ as a sum of indicator random variables. To achieve this, for each vertex v , let $\mathbb{I}_v$ be the indicator for the event $v \in S$. For two vertices u and v , we use $\mathbb{P}(u, v \in S)$ to mean $\mathbb{P}(u \in S \text{ and } v \in S)$. The variance of $|S| = \sum_v \mathbb{I}_v$ then decomposes into a sum of variances and covariances:

$$\begin{aligned} \text{Var}(|S|) &= \sum_v \text{Var}(\mathbb{I}_v) + \sum_{u \neq v} \text{Cov}(\mathbb{I}_u, \mathbb{I}_v) \\ &= \sum_v \mathbb{P}(v \in S) \mathbb{P}(v \notin S) + \sum_{u \neq v} \left(\mathbb{P}(u, v \in S) - \mathbb{P}(u \in S) \mathbb{P}(v \in S) \right) \\ (4.1) \quad &= \sum_v \rho_D(v) (1 - \rho_D(v)) + \sum_{u \neq v} \left(\mathbb{P}(u, v \in S) - (1 - \rho_D(u))(1 - \rho_D(v)) \right). \end{aligned}$$

In the last step, we used the identity $\mathbb{P}(v \in S) = 1 - \rho_D(v)$ from Proposition 3.7. We already have an explicit formula for $\rho_D(v)$ from (1.3). It remains to derive a formula for the joint probability $\mathbb{P}(u, v \in S)$ for any pair of distinct vertices, which is the focus of the present section.

This section is organized as follows. We begin in §4.1 by outlining our strategy, which involves analyzing a complementary event using the Principle of Inclusion-Exclusion. To evaluate the terms in this expansion, we establish a general ordering lemma in §4.2. We then apply this framework to three scenarios: nonadjacent vertices in §4.3, adjacent vertices with one arc in §4.4, and adjacent vertices with two arcs in §4.5. Finally, in Theorem 4.5, we synthesize the analysis into a final theorem: a complete recipe for the covariance.

4.1. A strategy via inclusion-exclusion. A direct calculation of $\mathbb{P}(u, v \in S)$ seems difficult. The condition for inclusion requires the *existence* of both a right in-neighbor and a right out-neighbor; this structure is not easy to work with. It is more tractable to analyze the complementary probability, $\mathbb{P}(u \notin S \text{ or } v \notin S)$. The condition for a vertex w to be excluded (that *all* its in-neighbors

or *all* its out-neighbors are to the left of w) is defined by universal quantifiers. This structure is ideally suited for a direct application of the Principle of Inclusion-Exclusion (PIE).

To formalize this strategy, we consider the following events for a vertex x based on its ordering relative to its neighbors:

$$\begin{aligned} E_x^- &= \{\text{all in-neighbors of } x \text{ appear before } x\}, \\ E_x^+ &= \{\text{all out-neighbors of } x \text{ appear before } x\}. \end{aligned}$$

As a reminder, saying that all in-/out-neighbors appear before x is the same as saying x has no right in-/out-neighbors. As discussed in the proof of Proposition 3.7, the event $x \notin S$ is equivalent to the union $E_x^- \cup E_x^+$. For two distinct vertices u, v , we are interested in the probability:

$$\Psi(u, v) := \mathbb{P}(u \notin S \text{ or } v \notin S) = \mathbb{P}((E_u^- \cup E_u^+) \cup (E_v^- \cup E_v^+)).$$

Let $A_1 := E_u^-$, $A_2 := E_u^+$, $A_3 := E_v^-$, and $A_4 := E_v^+$. Since $\Psi(u, v) = \mathbb{P}(A_1 \cup A_2 \cup A_3 \cup A_4)$, we can apply PIE to this union probability to express:

$$(4.2) \quad \Psi(u, v) = \Sigma_1 - \Sigma_2 + \Sigma_3 - \Sigma_4,$$

where Σ_k is the sum of probabilities of all k -fold intersections of these events:

$$\begin{aligned} \Sigma_1 &= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) + \mathbb{P}(A_4) \\ \Sigma_2 &= \mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_1 \cap A_3) + \cdots + \mathbb{P}(A_3 \cap A_4) \quad (6 \text{ terms}) \\ \Sigma_3 &= \mathbb{P}(A_1 \cap A_2 \cap A_3) + \mathbb{P}(A_1 \cap A_2 \cap A_4) + \mathbb{P}(A_1 \cap A_3 \cap A_4) + \mathbb{P}(A_2 \cap A_3 \cap A_4) \\ \Sigma_4 &= \mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4). \end{aligned}$$

The desired joint probability is then $\mathbb{P}(u, v \in S) = 1 - \Psi(u, v)$. The evaluation of each term is sensitive to the relationship between u and v . We therefore bifurcate our analysis into two scenarios: the nonadjacent case and the adjacent cases. To facilitate these calculations, we denote neighborhood overlaps as:

$$\begin{aligned} n_{uv}^{--} &:= |N^-(u) \cap N^-(v)|, & n_{uv}^{++} &:= |N^+(u) \cap N^+(v)|, & n_{uv}^{-+} &:= |N^-(u) \cap N^+(v)|, \\ n_{uv}^{+-} &:= |N^+(u) \cap N^-(v)|, & n_{uv}^{\pm-} &:= |N(u) \cap N^-(v)|, & n_{uv}^{-\pm} &:= |N^-(u) \cap N(v)|, \\ n_{uv}^{\pm+} &:= |N(u) \cap N^+(v)|, & n_{uv}^{+\pm} &:= |N^+(u) \cap N(v)|, & n_{uv}^{\pm\pm} &:= |N(u) \cap N(v)|. \end{aligned}$$

4.2. A general ordering lemma. Before proceeding to the case analysis, we present a lemma for computing the probability of joint ordering events.

Lemma 4.1. *Let X, Y be two sets of vertices, and let v, w be two distinct vertices not in $X \cup Y$. In a uniform random permutation of the vertices in $Z := X \cup Y \cup \{v, w\}$, the probability that all vertices of X appear before v and all vertices of Y appear before w is given by*

$$g(X, Y) = \frac{1}{|X| + |Y| - |X \cap Y| + 2} \left(\frac{1}{|X| + 1} + \frac{1}{|Y| + 1} \right).$$

Proof. Let us write $a \prec b$ to denote that vertex a appears before vertex b in a uniform random permutation of the vertices in Z . Let $x := |X|$, $y := |Y|$, $t := |X \cap Y|$, and let $N := |Z| = x + y - t + 2$ be the total number of vertices. Denote by $X \prec v$ the event that all vertices of X appear before v . We want to find the probability of the event $\mathcal{E}$, defined as $\{X \prec v \text{ and } Y \prec w\}$.

We condition on the last element, z , in the random permutation of Z . For $\mathcal{E}$ to occur, z cannot be in X (since $X \prec v$) or in Y (since $Y \prec w$). By hypothesis, $v, w \notin X \cup Y$, so the only candidates for the last element are v and w .

Case 1: The last element is v .

Any element of Z is equally likely to be last, so $\mathbb{P}(z = v) = 1/N$. Conditional on v being the last element, the condition $X \prec v$ is automatically satisfied. The event $\mathcal{E}$ thus reduces to the single

condition $Y \prec w$. This depends only on the relative ordering of the $y + 1$ vertices in $Y \cup \{w\}$; the event $Y \prec w$ holds if and only if w is the last among them, which has probability $1/(y + 1)$.

Case 2: The last element is w .

By a symmetric argument, $\mathbb{P}(z = w) = 1/N$. Conditional on w being last, the condition $Y \prec w$ is satisfied. The event $\mathcal{E}$ now requires only $X \prec v$. This occurs if and only if v is the last among the $x + 1$ vertices in $X \cup \{v\}$, which has probability $1/(x + 1)$.

Since these two cases are disjoint and are the only ways for $\mathcal{E}$ to occur, we sum their probabilities:

$$\begin{aligned} \mathbb{P}(\mathcal{E}) &= \mathbb{P}(z = v)\mathbb{P}(\mathcal{E} \mid z = v) + \mathbb{P}(z = w)\mathbb{P}(\mathcal{E} \mid z = w) \\ &= \frac{1}{N} \cdot \frac{1}{y + 1} + \frac{1}{N} \cdot \frac{1}{x + 1} = \frac{1}{x + y - t + 2} \left(\frac{1}{x + 1} + \frac{1}{y + 1} \right). \end{aligned} \quad \square$$

4.3. Case 1: Nonadjacent vertices. We begin with the simpler case where there is no arc between u and v . The lack of a directed edge ensures that $u \notin N(v)$ and $v \notin N(u)$. Consequently, for any choice of neighborhoods $X \subseteq N(u)$ and $Y \subseteq N(v)$, the conditions of Lemma 4.1 are met.

The four singleton probabilities for Σ_1 are given by the probability that a vertex is the last among its local neighborhood:

$$\begin{aligned} \mathbb{P}(E_u^-) &= \frac{1}{d_u^- + 1}, & \mathbb{P}(E_v^-) &= \frac{1}{d_v^- + 1}, \\ \mathbb{P}(E_u^+) &= \frac{1}{d_u^+ + 1}, & \mathbb{P}(E_v^+) &= \frac{1}{d_v^+ + 1}. \end{aligned}$$

The six pairwise intersection probabilities for Σ_2 consist of two self-pairs and four mixed-pairs. The self-pairs are:

$$\mathbb{P}(E_u^- \cap E_u^+) = \frac{1}{d_u + 1}, \quad \mathbb{P}(E_v^- \cap E_v^+) = \frac{1}{d_v + 1}.$$

The four mixed-pairs are all direct applications of Lemma 4.1:

$$\begin{aligned} \mathbb{P}(E_u^- \cap E_v^-) &= g(N^-(u), N^-(v)), & \mathbb{P}(E_u^- \cap E_v^+) &= g(N^-(u), N^+(v)), \\ \mathbb{P}(E_u^+ \cap E_v^-) &= g(N^+(u), N^-(v)), & \mathbb{P}(E_u^+ \cap E_v^+) &= g(N^+(u), N^+(v)). \end{aligned}$$

Similarly, the four triple intersection probabilities for Σ_3 are given by:

$$\begin{aligned} \mathbb{P}(E_u^- \cap E_u^+ \cap E_v^-) &= g(N(u), N^-(v)), & \mathbb{P}(E_u^- \cap E_u^+ \cap E_v^+) &= g(N(u), N^+(v)), \\ \mathbb{P}(E_u^- \cap E_v^- \cap E_v^+) &= g(N^-(u), N(v)), & \mathbb{P}(E_u^+ \cap E_v^- \cap E_v^+) &= g(N^+(u), N(v)). \end{aligned}$$

Finally, the single quadruple probability for Σ_4 is:

$$\mathbb{P}(E_u^- \cap E_u^+ \cap E_v^- \cap E_v^+) = g(N(u), N(v)).$$

Substituting these 15 quantities into the PIE formula (4.2) yields a closed-form expression for $\mathbb{P}(u, v \in S)$ in the nonadjacent case.

4.4. Case 2: Adjacent vertices with one arc. Suppose we have $u \rightarrow v$ but $v \nrightarrow u$ (the case $v \rightarrow u$ and $u \nrightarrow v$ follows by symmetry). This implies $v \in N^+(u)$ and $u \in N^-(v)$, creating a dependency that invalidates the direct application of Lemma 4.1 for some terms and makes certain joint events impossible. In particular, the arc $u \rightarrow v$ forces one pairwise intersection to be impossible.

Claim 4.2 (An impossible intersection). If $u \rightarrow v$ is an arc, the joint event $E_u^+ \cap E_v^-$ cannot occur, so $\mathbb{P}(E_u^+ \cap E_v^-) = 0$.

Proof. The event E_u^+ requires all out-neighbors of u to appear before u . Since $v \in N^+(u)$, this implies the relative order $v \prec u$. The event E_v^- requires all in-neighbors of v to appear before v . Since $u \in N^-(v)$, this implies $u \prec v$. These required orderings are mutually exclusive, so the joint event has probability zero. $\square$

Claim 4.2 has a cascading effect on the PIE expansion, as the term $\mathbb{P}(E_u^+ \cap E_v^-)$ and any higher-order term (in Σ_3 and Σ_4) containing this intersection will vanish. We now evaluate the remaining nonzero terms again under the assumption that $u \rightarrow v$ and $v \nrightarrow u$. The singleton probabilities for Σ_1 and the self-pair probabilities in Σ_2 are unaffected by the arc. We still have:

$$\mathbb{P}(E_u^- \cap E_u^+) = \frac{1}{d_u + 1}, \quad \mathbb{P}(E_v^- \cap E_v^+) = \frac{1}{d_v + 1}.$$

Of the mixed-pairs, the term $\mathbb{P}(E_u^- \cap E_v^+)$ can still be calculated with Lemma 4.1:

$$\mathbb{P}(E_u^- \cap E_v^+) = g(N^-(u), N^+(v)).$$

The remaining two pairwise intersections, however, require a modified analysis.

Claim 4.3 (Custom pairwise intersections). Assume $u \rightarrow v$. The joint probabilities are:

$$\begin{aligned} \mathbb{P}(E_u^- \cap E_v^-) &= \frac{1}{(d_u^- + d_v^- - n_{uv}^{--} + 1)(d_u^- + 1)}, \\ \mathbb{P}(E_u^+ \cap E_v^+) &= \frac{1}{(d_u^+ + d_v^+ - n_{uv}^{++} + 1)(d_v^+ + 1)}. \end{aligned}$$

Proof. We prove the result for $\mathbb{P}(E_u^- \cap E_v^-)$; the proof for $\mathbb{P}(E_u^+ \cap E_v^+)$ is symmetric.

Let $X = N^-(u)$ and $Y = N^-(v)$. We are interested in the event $\mathcal{E}$ where all of X appear before u (event $X \prec u$) and all of Y appear before v (event $Y \prec v$). Since there is an arc $u \rightarrow v$, we have $u \in N^-(v) = Y$. The condition $Y \prec v$ therefore forces the relative ordering $u \prec v$.

The outcome of the event $\mathcal{E}$ depends only on the relative ordering of the vertices in the set $Z := X \cup Y \cup \{v\}$. Note that $u \in Y$, so $u \in Z$. The size of this set is:

$$|Z| = |X \cup Y \cup \{v\}| = |X \cup Y| + 1 = |X| + |Y| - |X \cap Y| + 1 = d_u^- + d_v^- - n_{uv}^{--} + 1.$$

For the event $\mathcal{E}$ to occur, two conditions must be met:

- (1) The vertex v must appear after all vertices in Y . Since $u \in Y$, v must appear after u .
- (2) The vertex u must appear after all vertices in X .

Combining these two conditions implies that v must appear after all of $X \cup Y$, that is, v must be the last element in the random permutation of all vertices in Z .

We formalize the calculation using conditional probability. Let E_1 be the event that v is the last element in a permutation of Z . The probability of this event is:

$$\mathbb{P}(E_1) = \frac{1}{|Z|} = \frac{1}{d_u^- + d_v^- - n_{uv}^{--} + 1}.$$

Now, conditioned on v being last, the requirement $Y \prec v$ is automatically satisfied. The event $\mathcal{E}$ now only requires $X \prec u$. This remaining condition only depends on the relative ordering of the vertices in $X \cup \{u\}$. The probability that u is the last among these $|X| + 1$ vertices is:

$$\mathbb{P}(\mathcal{E} \mid E_1) = \mathbb{P}(X \prec u) = \frac{1}{|X| + 1} = \frac{1}{d_u^- + 1}.$$

Since the event $\mathcal{E}$ can only occur if v is the last element in the permutation of Z , the joint probability is the product:

$$\begin{aligned} \mathbb{P}(E_u^- \cap E_v^-) &= \mathbb{P}(\mathcal{E}) = \mathbb{P}(E_1) \cdot \mathbb{P}(\mathcal{E} \mid E_1) \\ &= \frac{1}{|Z|} \cdot \frac{1}{|X| + 1} = \frac{1}{(d_u^- + d_v^- - n_{uv}^{--} + 1)(d_u^- + 1)}. \end{aligned}$$

A symmetric argument holds for $\mathbb{P}(E_u^+ \cap E_v^+)$, where the dependency $v \in N^+(u)$ forces u to be the last element in the relevant permutation. $\square$

Finally, we consider the higher-order terms. The impossible intersection from Claim 4.2 implies that certain events have zero probability:

$$\mathbb{P}(E_u^- \cap E_u^+ \cap E_v^-) = 0, \quad \mathbb{P}(E_u^+ \cap E_v^- \cap E_v^+) = 0, \quad \mathbb{P}(E_u^- \cap E_u^+ \cap E_v^- \cap E_v^+) = 0.$$

The remaining two triple terms are $\mathbb{P}(E_u^- \cap E_v^- \cap E_v^+)$ and $\mathbb{P}(E_u^- \cap E_u^+ \cap E_v^+)$, which require a custom argument.

Claim 4.4 (Custom triple intersections). Assume $u \rightarrow v$. The two nonzero triple intersection probabilities are given below.

$$\begin{aligned} \mathbb{P}(E_u^- \cap E_u^+ \cap E_v^+) &= \frac{1}{(d_u + d_v^+ - n_{uv}^{\pm+} + 1)(d_v^+ + 1)} \\ \mathbb{P}(E_u^- \cap E_v^- \cap E_v^+) &= \frac{1}{(d_u^- + d_v - n_{uv}^{\pm-} + 1)(d_u^- + 1)} \end{aligned}$$

Proof. We explain the argument for the first identity; the proof for the second identity is similar. Let $X = N(u)$ and $Y = N^+(v)$. We compute $\mathbb{P}(\mathcal{E})$ where $\mathcal{E}$ is the event $\{X \prec u \text{ and } Y \prec v\}$. Since $u \rightarrow v$, the membership $v \in X$ forces the ordering $v \prec u$. Combined, these conditions imply that u must be the last element in any permutation of the relevant set $Z := X \cup Y \cup \{u\}$.

The probability of the event $\mathcal{E}$ is therefore the probability that u is last in Z , multiplied by the conditional probability that $Y \prec v$ (given u is last):

$$\begin{aligned} \mathbb{P}(\mathcal{E}) &= \mathbb{P}(u \text{ is last in } Z) \cdot \mathbb{P}(Y \prec v \mid u \text{ is last}) \\ &= \frac{1}{|X \cup Y| + 1} \cdot \frac{1}{|Y| + 1} \\ &= \frac{1}{(d_u + d_v^+ - n_{uv}^{\pm+} + 1)(d_v^+ + 1)}. \end{aligned} \quad \square$$

4.5. Case 3: Adjacent vertices with two arcs. We now analyze the case where a directed 2-cycle exists between u and v , meaning both arcs $u \rightarrow v$ and $v \rightarrow u$ are present.³

We proceed to the analysis of the terms in the PIE expansion (4.2). The four singleton probabilities and the two specific probabilities $\mathbb{P}(E_u^- \cap E_u^+)$ and $\mathbb{P}(E_v^- \cap E_v^+)$ are unaffected, as their calculations are local to a single vertex.

However, we show that all the other terms in the PIE expansion vanish. Since the directed edges $u \rightarrow v$ and $v \rightarrow u$ both exist, the four primary events imply an ordering relation between u and v :

$$\begin{aligned} E_u^- \text{ requires } v \prec u, & & E_v^- \text{ requires } u \prec v, \\ E_u^+ \text{ requires } v \prec u, & & E_v^+ \text{ requires } u \prec v. \end{aligned}$$

Since $u \prec v$ and $v \prec u$ are mutually exclusive, any event that requires both to be true is impossible:

$$\mathbb{P}(E_u^- \cap E_v^-) = \mathbb{P}(E_u^- \cap E_v^+) = \mathbb{P}(E_u^+ \cap E_v^-) = \mathbb{P}(E_u^+ \cap E_v^+) = 0.$$

Furthermore, every triple and quadruple intersection is a sub-event of at least one of the impossible mixed pairs identified above. For example, the event $E_u^- \cap E_u^+ \cap E_v^-$ requires the event $E_u^+ \cap E_v^-$ to hold. Therefore, all triple and quadruple intersection probabilities are 0.

We now consolidate our findings into a single theorem. This result provides a complete recipe for calculating the covariance by cataloging the 15 intersection probabilities required for the PIE expansion of $\Psi(u, v)$.

³The presence of such 2-cycles is the key distinction between a simple digraph and an *oriented graph*, which forbids them. As oriented graphs are a special case of the digraphs considered here, all our results apply to them as well.

Theorem 4.5 (Covariance formula). *For any two distinct vertices $u, v \in V(D)$, the covariance of their indicator variables $\mathbb{I}_u$ and $\mathbb{I}_v$ (for the events $u \in S$ and $v \in S$) is given by:*

$$\text{Cov}(\mathbb{I}_u, \mathbb{I}_v) = \mathbb{P}(u, v \in S) - (1 - \rho_D(u))(1 - \rho_D(v)),$$

where $\mathbb{P}(u, v \in S) = 1 - \Psi(u, v)$. The term $\Psi(u, v)$ is computed using the PIE formula (4.2), with the required probabilities given in Table 1.

TABLE 1. Complete catalog of PIE terms for all cases of vertex pairs.

Term	Case 1: Nonadjacent	Case 2: $u \rightarrow v$	Case 3: $u \rightleftarrows v$
Σ_1 Terms			
$\mathbb{P}(E_u^-)$	$\frac{1}{d_u^- + 1}$	$\frac{1}{d_u^- + 1}$	$\frac{1}{d_u^- + 1}$
$\mathbb{P}(E_u^+)$	$\frac{1}{d_u^+ + 1}$	$\frac{1}{d_u^+ + 1}$	$\frac{1}{d_u^+ + 1}$
$\mathbb{P}(E_v^-)$	$\frac{1}{d_v^- + 1}$	$\frac{1}{d_v^- + 1}$	$\frac{1}{d_v^- + 1}$
$\mathbb{P}(E_v^+)$	$\frac{1}{d_v^+ + 1}$	$\frac{1}{d_v^+ + 1}$	$\frac{1}{d_v^+ + 1}$
Σ_2 Terms			
$\mathbb{P}(E_u^- \cap E_u^+)$	$\frac{1}{d_u + 1}$	$\frac{1}{d_u + 1}$	$\frac{1}{d_u + 1}$
$\mathbb{P}(E_v^- \cap E_v^+)$	$\frac{1}{d_v + 1}$	$\frac{1}{d_v + 1}$	$\frac{1}{d_v + 1}$
$\mathbb{P}(E_u^- \cap E_v^-)$	$\frac{1}{d_u^- + d_v^- - n_{uv}^- + 2} \left(\frac{1}{d_u^- + 1} + \frac{1}{d_v^- + 1} \right)$	$\frac{1}{(d_u^- + d_v^- - n_{uv}^- + 1)(d_u^- + 1)}$	0
$\mathbb{P}(E_u^- \cap E_v^+)$	$\frac{1}{d_u^- + d_v^+ - n_{uv}^+ + 2} \left(\frac{1}{d_u^- + 1} + \frac{1}{d_v^+ + 1} \right)$	$\frac{1}{d_u^- + d_v^+ - n_{uv}^+ + 2} \left(\frac{1}{d_u^- + 1} + \frac{1}{d_v^+ + 1} \right)$	0
$\mathbb{P}(E_u^+ \cap E_v^-)$	$\frac{1}{d_u^+ + d_v^- - n_{uv}^- + 2} \left(\frac{1}{d_u^+ + 1} + \frac{1}{d_v^- + 1} \right)$	0	0
$\mathbb{P}(E_u^+ \cap E_v^+)$	$\frac{1}{d_u^+ + d_v^+ - n_{uv}^+ + 2} \left(\frac{1}{d_u^+ + 1} + \frac{1}{d_v^+ + 1} \right)$	$\frac{1}{(d_u^+ + d_v^+ - n_{uv}^+ + 1)(d_v^+ + 1)}$	0
Σ_3 Terms			
$\mathbb{P}(E_u^- \cap E_u^+ \cap E_v^-)$	$\frac{1}{d_u + d_v - n_{uv}^\pm + 2} \left(\frac{1}{d_u + 1} + \frac{1}{d_v + 1} \right)$	0	0
$\mathbb{P}(E_u^- \cap E_u^+ \cap E_v^+)$	$\frac{1}{d_u + d_v - n_{uv}^\pm + 2} \left(\frac{1}{d_u + 1} + \frac{1}{d_v + 1} \right)$	$\frac{1}{(d_u + d_v - n_{uv}^\pm + 1)(d_v + 1)}$	0
$\mathbb{P}(E_u^- \cap E_v^- \cap E_v^+)$	$\frac{1}{d_u^- + d_v - n_{uv}^\pm + 2} \left(\frac{1}{d_u^- + 1} + \frac{1}{d_v + 1} \right)$	$\frac{1}{(d_u^- + d_v - n_{uv}^\pm + 1)(d_u^- + 1)}$	0
$\mathbb{P}(E_u^+ \cap E_v^- \cap E_v^+)$	$\frac{1}{d_u^+ + d_v - n_{uv}^\pm + 2} \left(\frac{1}{d_u^+ + 1} + \frac{1}{d_v + 1} \right)$	0	0
Σ_4 Term			
$\mathbb{P}(E_u^- \cap E_u^+ \cap E_v^- \cap E_v^+)$	$\frac{1}{d_u + d_v - n_{uv}^\pm + 2} \left(\frac{1}{d_u + 1} + \frac{1}{d_v + 1} \right)$	0	0

5. THE VARIANCE-BASED LOWER BOUND

We have now assembled all the necessary components for our second main result: a variance-based refinement of the AGJS bound (1.2). The previous sections provided a randomized algorithm for generating a feedback vertex set, S , and established a formula for calculating the variance of its size, $\text{Var}(|S|)$. In this section, we connect these pieces using Bhatia–Davis inequality [3] to produce a refined lower bound on $\vec{\alpha}(D)$.

Theorem 5.1. *Let D be a digraph with n vertices and c connected components. Let S be the random feedback vertex set produced by the DL algorithm. Then*

$$(5.1) \quad \vec{\alpha}(D) \geq \sum_{v \in V(D)} \rho_D(v) + \frac{\text{Var}(|S|)}{\left(\sum_{v \in V(D)} \rho_D(v)\right) - c}.$$

If the denominator is zero, the second term is taken to be zero.

Proof. We apply the Bhatia–Davis inequality, which bounds the variance of a random variable by its expectation and range. Let $|S|$ be the size of the random FVS generated by the DL algorithm. From our previous analysis, we have three key ingredients:

- The minimum FVS size provides a lower bound, so $|S| \geq \vec{\beta}(D)$.
- Proposition 3.6 gives a deterministic upper bound, $|S| \leq n - c$.
- Proposition 3.7 and Corollary 3.8 give the expected size, $\mu := \mathbb{E}[|S|] = n - \sum_v \rho_D(v)$.

For a random variable X with range $[m, M]$, the Bhatia–Davis inequality [3] states that $\text{Var}(X) \leq (M - \mathbb{E}[X])(\mathbb{E}[X] - m)$. Applying this to $X = |S|$ with $m = \vec{\beta}(D)$ and $M = n - c$, we get:

$$\text{Var}(|S|) \leq (M - \mu)(\mu - \vec{\beta}(D)).$$

If the term $M - \mu$ is positive, we can rearrange this inequality to get an improved upper bound on the minimum FVS size, $\vec{\beta}(D)$:

$$\vec{\beta}(D) \leq \mu - \frac{\text{Var}(|S|)}{M - \mu}.$$

We now use the identity $\vec{\alpha}(D) = n - \vec{\beta}(D)$ to translate this into a lower bound on $\vec{\alpha}(D)$. Substituting the expressions for μ and M , we obtain the desired inequality:

$$\begin{aligned} \vec{\alpha}(D) &= n - \vec{\beta}(D) \geq n - \left(\mu - \frac{\text{Var}(|S|)}{M - \mu} \right) = (n - \mu) + \frac{\text{Var}(|S|)}{M - \mu} \\ &= \left(n - \left(n - \sum_{v \in V(D)} \rho_D(v) \right) \right) + \frac{\text{Var}(|S|)}{(n - c) - (n - \sum_v \rho_D(v))} \\ &= \sum_{v \in V(D)} \rho_D(v) + \frac{\text{Var}(|S|)}{\sum_v \rho_D(v) - c}. \end{aligned}$$

If the term $M - \mu = 0$, then the denominator $\sum_v \rho_D(v) - c$ vanishes; in this case, the correction term is taken to be zero. This completes the proof. $\square$

Next, we address the case when the denominator $\sum_v \rho_D(v) - c$ is zero.

Lemma 5.2. *Let D be a digraph with c connected components. Then $\sum_v \rho_D(v) - c = 0$ if and only if every connected component of D is a complete symmetric digraph (namely, a graph which has an arc between every ordered pair of vertices).*

Proof. Let S denote the random FVS produced as an output of the DL algorithm. We have $\mathbb{E}[|S|] = n - \sum_{v \in V(D)} \rho_D(v)$ from Corollary 3.8. Therefore, the equality $\sum_v \rho_D(v) - c = 0$ holds if and only if $\mathbb{E}[|S|] = n - c$. By Proposition 3.6, the maximum possible value of $|S|$ is also $n - c$. Since the average value satisfies $\mathbb{E}[|S|] = n - c$, the random variable $|S|$ must be constant, namely $n - c$. It follows that in *every* ordering of the vertices, the DL algorithm produces a FVS of size exactly $n - c$.

From the proof of Proposition 3.6, each connected component of D has at least one vertex not in S . We have exactly c elements not in S , which forces exactly one element from each connected component not to be in S for every ordering of the vertices. This latter condition forces each connected component to be a complete symmetric digraph. Indeed, if a connected component were

not a complete symmetric digraph, there would exist two vertices u and v such that (i) u and v are nonadjacent, (ii) $u \rightarrow v$ but $v \nrightarrow u$, or (iii) $v \rightarrow u$ but $u \nrightarrow v$. In all of these cases, the vertex order in which u and v appear as the last two vertices will produce a feedback vertex set not containing u and v . This is because, in such an ordering, neither u nor v has both a right in-neighbor *and* a right out-neighbor; so, they do not belong to the feedback vertex set produced by the DL algorithm. $\square$

Remark 5.3 (Interpreting the bound). The denominator of the refinement term in equation 5.1 is always nonnegative. Indeed, $\sum_{v \in V(D)} \rho_D(v) - c$ is the gap between the expected number of vertices not in S and the minimum possible number of vertices not in S . Our bound therefore provides an improvement over the AGJS bound (1.2) precisely when both the variance $\text{Var}(|S|)$ and this denominator are strictly positive. The denominator is zero if and only if D is a complete symmetric digraph. In this case, the maximum output size from the DL algorithm is equal to its expected size. This rare scenario is characterized by Lemma 5.2. In these exceptional cases, we set the correction term to zero and our bound coincides with the AGJS bound (1.2).

In the important case of connected graphs, the bound simplifies as follows:

Corollary 5.4. *If a digraph D is connected, then*

$$\vec{\alpha}(D) \geq \sum_{v \in V(D)} \left(\frac{1}{1+d_v^+} + \frac{1}{1+d_v^-} - \frac{1}{1+d_v} \right) + \frac{\text{Var}(|S|)}{\left(\sum_{v \in V(D)} \left(\frac{1}{1+d_v^+} + \frac{1}{1+d_v^-} - \frac{1}{1+d_v} \right) \right) - 1}.$$

Proof. This follows directly by substituting the value $c = 1$ for a connected graph into the denominator term in (5.1) and substituting the explicit formula for $\rho_D(v)$ from (1.3). $\square$

6. NUMERICAL COMPARISONS ON RANDOM GRAPH MODELS

We established two distinct refinements of the AGJS bound (1.2) for the maximum size of an induced acyclic subgraph in a digraph: a neighborhood-based method (Theorem 2.2) and a variance-based approach (Theorem 5.1). Because their formulas are complex, an analytic comparison is unwieldy; so, we analyze each bound separately and then benchmark them on random models.

6.1. Neighborhood-based bound. From (2.1), the improvement for the neighborhood-based bound is:

$$\Delta_{\text{neigh}} := \sum_{v \in V(D)} \rho_D(v) \cdot \max \left\{ 0, 1 - \rho_D(v) - \sum_{u \in N(v)} \rho_D(u) \right\}$$

where $\rho_D(v)$ is given by the formula (1.3):

$$\rho_D(v) = \frac{1}{1+d_v^+} + \frac{1}{1+d_v^-} - \frac{1}{1+d_v}.$$

A vertex v contributes to the improvement Δ_{neigh} only if $1 - \rho_D(v) - \sum_{u \in N(v)} \rho_D(u) > 0$. In that case, its contribution is $C_v = \rho_D(v)(1 - \rho_D(v) - \sum_{u \in N(v)} \rho_D(u))$. This is maximized (for fixed v) when $\sum_{u \in N(v)} \rho_D(u)$ is small, in which case $C_v \approx \rho_D(v)(1 - \rho_D(v))$, whose maximum occurs at $\rho_D(v) = \frac{1}{2}$. Vertices with $\rho_D(v) \approx \frac{1}{2}$ typically have relatively low in- and out-degrees. We conclude that the vertices that contribute to an improved neighborhood-based bound are those that (a) have relatively low in- and out-degrees, and (b) have neighbors with relatively high in- and out-degrees.

6.2. **Variance-based bound.** From (5.1), the improvement for the variance-based bound is:

$$\Delta_{\text{var}} := \frac{\text{Var}(|S|)}{\sum_{v \in V(D)} \rho_D(v) - c},$$

where S is the DL output (Section 3.1) and c is the number of connected components.

From this formula, we conclude that a meaningful improvement of the variance-based bound over the AGJS bound (1.2) is dependent on a sufficiently large $\text{Var}(|S|)$. Recall that the DL algorithm produces a feedback vertex set for each of the $n!$ distinct vertex orderings using the rule:

Let π be a random vertex ordering from digraph D , and let v_i be the vertex in the i -th position within π . Include v_i in S if and only if v_i has at least one right in-neighbor and at least one right out-neighbor.

Recall $\text{Var}(|S|)$ measures the spread of $|S|$ over all $n!$ vertex orderings. To identify the type of digraph where $\text{Var}(|S|)$ would be meaningful, we need to consider what graph properties would result in significant differences in FVS sizes as we vary the vertex orderings.

Consider a graph with two classes: high-degree and low-degree vertices. High-degree vertices are likely to enter S regardless of position, while low-degree vertices are sensitive to position. Orderings that place many low-degree vertices early tend to yield larger $|S|$, and the reverse yields smaller $|S|$, inflating $\text{Var}(|S|)$.

The heuristic above suggests that heterogeneous degree structure should favor both refinements over (1.2). We now test this on random models. We also compare the two refinements to each other.

6.3. **An Erdős–Rényi random digraph model.** In this model, each ordered pair (u, v) is included as an arc independently with probability p (both directions allowed). Degrees are nearly uniform in expectation, so we do not expect substantial gains from either refinement. Computer simulations support this prediction. As shown in Table 2, the two new lower bounds had no meaningful improvement over the AGJS bound (1.2) in the random Erdős–Rényi simple digraphs.

$p = 0.05$		$p = 0.50$		$p = 0.95$	
Δ_{neigh}	Δ_{var}	Δ_{neigh}	Δ_{var}	Δ_{neigh}	Δ_{var}
0.0	0.31	0.0	0.50	0.0	0.91

TABLE 2. Erdős–Rényi $G(n, p)$ digraphs with $n = 100$; for each p we generate 1000 graphs, and the entries report the average improvements Δ_{neigh} and Δ_{var} .

6.4. **A two-type random digraph model.** We next use a graph model with two types of vertices: those with relatively high degrees and those with relatively low degrees.

Definition 6.1 (Two-type random digraph). A random digraph D is generated from the following parameters:

- the number of vertices n ;
- the proportion of low-degree vertices $p_{\text{low}} \in [0, 1]$;
- arc probabilities q_1 (high–high), q_2 (high–low), q_3 (low–low), where $q_i \in [0, 1]$.

The generation process is as follows:

- (1) The vertex set V is partitioned into two disjoint subsets: V_{low} of size $\lfloor p_{\text{low}} \cdot n \rfloor$ and V_{high} of size $n - |V_{\text{low}}|$.

- (2) For each ordered pair of distinct vertices (u, v) , there is the potential of zero, one, or two (directionally different) arcs between them. The probability that $u \rightarrow v$ exists is given by:

$$p_{\text{edge}} = \begin{cases} q_1 & \text{if } u, v \in V_{\text{high}} \\ q_2 & \text{if one vertex is in } V_{\text{high}} \text{ and the other is in } V_{\text{low}} \\ q_3 & \text{if } u, v \in V_{\text{low}} \end{cases}$$

Based on the computer experiments, we observed that the relative strength of the two bounds is sensitive to the value of q_3 . For instance, consider a scenario with a large proportion of low-degree type vertices ($p_{\text{low}} = 0.90$) and high connectivity involving the high-degree vertices ($q_1 = 0.70, q_2 = 0.50$). We observed that:

- When the low-degree vertices are very sparsely connected (e.g., $q_3 = 0.01$), the neighborhood-based refinement appears to provide a stronger correction ($\Delta_{\text{neigh}} > \Delta_{\text{var}}$).
- However, by only slightly increasing the connectivity between low-degree vertices (e.g., to $q_3 = 0.05$), the variance-based bound becomes superior ($\Delta_{\text{var}} > \Delta_{\text{neigh}}$).

We could vary other variables (say q_1) while fixing the rest. The variation of q_3 was the most sensitive, which is why we focus only on this scenario.

Tables 3, 4, and 5 report the average value of improvements in a random simulation of the two-type digraph model run with the indicated parameter values.

$q_3 \rightarrow$	0.005			0.01		
$n \downarrow$	AGJS	Δ_{neigh}	Δ_{var}	AGJS	Δ_{neigh}	Δ_{var}
100	20.89	8.78	3.84	19.93	5.82	3.53
150	21.72	10.60	4.58	20.40	6.88	4.20
200	21.94	11.39	5.02	20.67	7.32	4.52

TABLE 3. Two-type random digraphs with $p_{\text{low}} = 0.90$, $q_1 = 0.70$, $q_2 = 0.50$. Columns list $q_3 \in \{0.005, 0.01\}$. Averages based on 100 graphs for each (n, q_3) .

$q_3 \rightarrow$	0.02			0.025		
$n \downarrow$	AGJS	Δ_{neigh}	Δ_{var}	AGJS	Δ_{neigh}	Δ_{var}
100	17.92	2.48	3.07	16.95	1.53	2.84
150	18.12	2.52	3.50	17.14	1.49	3.20
200	18.31	2.53	3.74	17.32	1.37	3.39

TABLE 4. Two-type random digraphs with $p_{\text{low}} = 0.90$, $q_1 = 0.70$, $q_2 = 0.50$. Columns list $q_3 \in \{0.02, 0.025\}$. Averages based on 100 graphs for each (n, q_3) .

$q_3 \rightarrow$	0.05			0.10		
$n \downarrow$	AGJS	Δ_{neigh}	Δ_{var}	AGJS	Δ_{neigh}	Δ_{var}
100	13.47	0.15	2.01	9.39	0.00	1.15
150	13.53	0.08	2.15	9.41	0.00	1.20
200	13.61	0.04	2.23	9.49	0.00	1.22

TABLE 5. Two-type random digraphs with $p_{\text{low}} = 0.90$, $q_1 = 0.70$, $q_2 = 0.50$. Columns list $q_3 \in \{0.05, 0.10\}$. Averages based on 100 graphs for each (n, q_3) .

6.5. A bipartite random digraph model. Finally, we study a bipartite random digraph model with three parameters:

- The number of vertices, n .
- The proportion of vertices in the first bipartite set, denoted by a where $a \in [0, 1]$.
- The probability p that two vertices, one in each of the two partitions, are connected by an arc in a given direction.

Let D be the resulting bipartite digraph with $V(D) = A \cup B$, where A and B are the two bipartite sets. We set $n_1 = |A| = \lceil n \cdot a \rceil$ and $n_2 = |B| = n - n_1$. For each ordered pair (u, v) with $u \in A$ and $v \in B$, include the arc $u \rightarrow v$ with probability p . Independently, for (v, u) with $v \in B$ and $u \in A$, include the arc $v \rightarrow u$ with probability p . Both directions may occur since 2-cycles are permitted.

We observed the largest improvement when the value of a was small. This is consistent with our earlier observations, because the smaller the value of a , the more imbalanced the bipartite graph is. When $a = 0.5$, the bipartite graph would have uniform degrees, where we expect no improvement between our bounds and (1.2). Tables 6, 7, and 8 show the value of these improvements in a random computer simulation of the bipartite digraph model run with the indicated parameter values.

$a \rightarrow$	0.05			0.10		
$n \downarrow$	AGJS	Δ_{neigh}	Δ_{var}	AGJS	Δ_{neigh}	Δ_{var}
100	30.82	16.94	6.09	16.12	10.01	4.89
150	30.59	20.62	8.04	16.61	11.28	5.66
200	33.55	24.30	9.56	16.87	11.97	6.09

TABLE 6. Bipartite random digraphs with edge probability $p = 0.65$. Columns list $a \in \{0.05, 0.10\}$. Averages based on 100 graphs for each (n, a) .

$a \rightarrow$	0.15			0.20		
$n \downarrow$	AGJS	Δ_{neigh}	Δ_{var}	AGJS	Δ_{neigh}	Δ_{var}
100	10.65	5.99	3.77	7.88	3.62	2.95
150	10.65	6.31	4.07	7.99	3.88	3.18
200	10.98	6.81	4.36	8.06	4.02	3.29

TABLE 7. Bipartite random digraphs with edge probability $p = 0.65$. Columns list $a \in \{0.15, 0.20\}$. Averages based on 100 graphs for each (n, a) .

$a \rightarrow$	0.25			0.30		
$n \downarrow$	AGJS	Δ_{neigh}	Δ_{var}	AGJS	Δ_{neigh}	Δ_{var}
100	6.25	2.06	2.32	5.21	0.97	1.81
150	6.23	2.12	2.42	5.26	1.05	1.90
200	6.35	2.28	2.53	5.28	1.10	1.95

TABLE 8. Bipartite random digraphs with edge probability $p = 0.65$. Columns list $a \in \{0.25, 0.30\}$. Averages based on 100 graphs for each (n, a) .

REFERENCES

1. S. Akbari, A. H. Ghodrati, A. Jabalameli, and M. Saghafian, *Chromatic number and dichromatic polynomial of digraphs*, arXiv preprint (2017), 1–13, Available at <https://arxiv.org/abs/1711.06293>.
2. E. Angel, R. Campigotto, and C. Laforest, *A new lower bound on the independence number of graphs*, Discrete Appl. Math. **161** (2013), no. 6, 847–852.

3. R. Bhatia and C. Davis, *A Better Bound on the Variance*, The American Mathematical Monthly **107** (2000), no. 4, 353–357.
4. Y. Caro, *New results on the independence number*, Tech. report, Tel-Aviv University, 1979.
5. Hermann Gruber, *Bounding the feedback vertex number of digraphs in terms of vertex degrees*, Discrete Appl. Math. **159** (2011), no. 8, 872–875. MR 2782647
6. J. Harant and S. Mohr, *On Selkow’s bound on the independence number of graphs*, Discuss. Math. Graph Theory (2019), no. 3, 655–657.
7. Víctor Neumann-Lara, *The dichromatic number of a digraph*, Journal of Combinatorial Theory, Series B **33** (1982), no. 3, 265–270.
8. S. M. Selkow, *A probabilistic lower bound on the independence number of graphs*, Discrete Mathematics **132** (1994), no. 1-3, 363–365.
9. V. K. Wei, *A lower bound on the stability number of a simple graph*, Tech. Report 81-11217-9, Bell Laboratories, Murray Hill, NJ, 1981.

DEPT. OF MATHEMATICS & COMPUTER SCIENCE, SANTA CLARA UNIVERSITY, SANTA CLARA, CA 95053.
Email address: `sasgarli@scu.edu`

MONTE SERENO, CA 95030
Email address: `donaldfalkenhagen@gmail.com`

SCHOOL OF ENGINEERING, SANTA CLARA UNIVERSITY, SANTA CLARA, CA 95053.
Email address: `khoshi@scu.edu`