

Sharp variance bounds for random greedy independent sets in split graphs and cographs

Arya Somu

Abstract

Inspect the vertices of a finite graph in a uniformly random order, accepting a vertex if it has no previously accepted neighbor. We determine the largest variance of the accepted cardinality among split graphs of each order, and separately among cographs. For $n \geq 3$, the unique extremizer is the join of a clique of order $n - k$ and an independent set of order k , where $k = \lfloor (3n + 2)/4 \rfloor$. Its variance is $k(n - k)(k - 1)^2/n^2$, asymptotic to $27n^2/256$. The proof uses a sharp lower bound $\mathbb{E}X_G \geq 1 + \alpha(G)(\alpha(G) - 1)/n$, with equality classified in both graph classes. We also obtain a stability statement for split graphs. The extremal conclusion does not extend to all graphs: an explicit five-part construction exceeds it for every order $48t$ with $t \geq 7$, and has limiting normalized variance $839833/7907424 > 27/256$.

Proposition draft. The arguments are self-contained and accompanied by exact-arithmetic checks. A targeted literature search did not locate these particular extremal statements; this is not a certification of novelty or a substitute for external mathematical review.

1 Problem and main results

All graphs are finite, simple, and undirected. Write $n = |V(G)|$, $\alpha(G)$ for the independence number, and X_G for the number of vertices accepted by the random-order greedy independent-set algorithm. Its output is maximal, but need not be maximum. All randomness here is in the order; the graph is fixed. Put $\mu(G) = \mathbb{E}X_G$ and $v(G) = \text{Var}(X_G)$.

A *split graph* admits a partition $V(G) = I \sqcup C$ with I independent and C a clique. The *join* $G \vee H$ of disjoint graphs adds every edge between their vertex sets. A *cograph* is obtained from single vertices by repeatedly taking disjoint unions and joins. This recursive description is standard [2]. Define

$$S_{n,k} = K_{n-k} \vee \overline{K}_k, \quad F_n(k) = \frac{k(n-k)(k-1)^2}{n^2}, \quad k_n = \left\lfloor \frac{3n+2}{4} \right\rfloor.$$

An empty clique is permitted, so $S_{n,n}$ is edgeless.

Theorem 1. *Let $n \geq 3$. Among all split graphs on n vertices, and separately among all cographs on n vertices,*

$$\max_G v(G) = F_n(k_n).$$

In either class equality holds precisely for $G \cong S_{n,k_n}$. In particular,

$$\frac{1}{n^2} \max_G v(G) \longrightarrow \frac{27}{256}.$$

For $n = 1, 2$, every graph has variance zero.

Theorem 2. *For every integer $t \geq 7$, there is a graph H_t on $48t$ vertices with*

$$v(H_t) > F_{48t}(k_{48t}).$$

The graph consists of five vertex classes along a path, of sizes

$$t, \quad 6t, \quad 34t, \quad 6t, \quad t.$$

The second and fourth classes are cliques; the other three are independent. Consecutive classes are completely joined, and there are no other interclass edges. Its variance is

$$v(H_t) = \frac{20155992t^2 - 2071572t + 55821}{82369},$$

and hence

$$\lim_{t \rightarrow \infty} \frac{v(H_t)}{(48t)^2} = \frac{839833}{7907424} = 0.1062081658 \dots > \frac{27}{256} = 0.10546875.$$

The model is also called blocking random sequential adsorption on a graph; see Penrose and Sudbury [5]. Gamarnik and Goldberg [3] give variance and concentration results under bounded-degree hypotheses. Krivelevich, Mészáros, Michaeli, and Shikhelman [4] study local limits and determine the mean-minimizing tree of each order. A recent preprint of Shaikh [6] proves the sharp bound $v(G) \leq |E(G)|((n-2)/n)^2$ for triangle-free graphs, including the variance-maximizing tree. The present question concerns exact maxima by order in two classes that allow dense graphs and triangles. The counterexample in Theorem 2 concerns extension of our class-specific conclusion; it does not contradict any of those results.

2 A sharp lower bound on the mean

For a vertex u , let $N[u]$ denote its closed neighborhood. Conditional on u being first, it is accepted, its neighbors are rejected, and the relative order on the remaining vertices is uniform. Thus, with $\mu(\emptyset) = 0$,

$$\mu(G) = 1 + \frac{1}{n} \sum_{u \in V(G)} \mu(G - N[u]). \quad (1)$$

Lemma 3. *Suppose G is a split graph with a specified partition $I \sqcup C$, $|I| = k \geq 1$, $|C| = \ell$, and $n = k + \ell$. Then*

$$\mu(G) \geq 1 + \frac{k(k-1)}{n}. \quad (2)$$

For $k \geq 2$, put $r_i = |C \setminus N(i)|$ for $i \in I$. More precisely,

$$n\mu(G) \geq n + k(k-1) + \sum_{i \in I} \frac{r_i(r_i+1)}{k-1+r_i}. \quad (3)$$

Equality in (2) holds if and only if G is the complete join of this I and C .

Proof. We induct on k , allowing every value of ℓ . For $k = 1$, $\mu(G) \geq 1$. If G has a nonedge xy , the positive-probability event that x, y are first and second accepts both, so $\mu(G) > 1$. Thus equality holds exactly when G is complete, as required.

Let $k \geq 2$. If the first vertex is $c \in C$, the other clique vertices are blocked and every nonneighbor of c in I is accepted. The count is exactly $1 + k - d_I(c)$, where $d_I(c) = |N(c) \cap I|$. If the first vertex

is $i \in I$, the residual graph has the split partition $I \setminus \{i\}$ and $C \setminus N(i)$, of sizes $k - 1$ and r_i . It is nonempty. The induction hypothesis makes its expected accepted count at least

$$1 + \frac{(k-1)(k-2)}{k-1+r_i}.$$

Writing $e(I, C)$ for the number of cross-edges, first-choice conditioning gives

$$n\mu(G) \geq \ell(k+1) - e(I, C) + 2k + (k-1)(k-2) \sum_{i \in I} \frac{1}{k-1+r_i}.$$

Because $\ell k - e(I, C) = \sum_i r_i$, subtracting $n + k(k-1)$ leaves

$$\sum_{i \in I} \left(r_i - (k-2) + \frac{(k-1)(k-2)}{k-1+r_i} \right) = \sum_{i \in I} \frac{r_i(r_i+1)}{k-1+r_i}.$$

This proves (3), and it is strict in (2) if any $r_i > 0$. When all $r_i = 0$, the first vertex determines the result: $X_G = k$ with probability k/n and $X_G = 1$ with probability ℓ/n . This law gives equality. $\square$

Every split graph admits such a partition with $k = \alpha(G)$. Indeed, given any split partition, an independent set contains at most one clique vertex. If I is not maximum, some clique vertex has no neighbor in I and can be moved into I , increasing its size by one to the maximum. Thus Lemma 3 applies with $k = \alpha(G)$.

Lemma 4. *If G is a nonempty cograph, $n = |V(G)|$, and $k = \alpha(G)$, then*

$$\mu(G) \geq 1 + \frac{k(k-1)}{n}.$$

Equality holds precisely when $G \cong S_{n,k}$.

Proof. Induct on a recursive union/join construction. The single-vertex case is immediate. For two nonempty constituent graphs, write their orders as p, q , their independence numbers as a, b , and their means as μ_1, μ_2 .

For a disjoint union, $\mu(G) = \mu_1 + \mu_2$ and $k = a + b$. By induction it suffices to prove

$$2 + \frac{a(a-1)}{p} + \frac{b(b-1)}{q} \geq 1 + \frac{(a+b)(a+b-1)}{p+q}.$$

Set $r = a/p$ and $s = b/q$. The difference between the two sides is

$$1 - \frac{qr + ps}{p+q} + \frac{pq}{p+q}(r-s)^2 \geq 0, \tag{4}$$

since $0 < r, s \leq 1$. Equality requires $r = s = 1$, so both constituent graphs, and hence their union, are edgeless.

For a join, the first accepted vertex blocks the entire other constituent. Conditional on which constituent contains the first vertex, the internal relative order in that constituent is uniform. Consequently

$$\mu(G) = \frac{p\mu_1 + q\mu_2}{p+q} \geq 1 + \frac{a(a-1) + b(b-1)}{p+q} \geq 1 + \frac{k(k-1)}{p+q},$$

where $k = \max(a, b)$. For equality, each constituent must attain its inductive bound, and the smaller independence number must be one. A graph with independence number one is a clique. By induction the other constituent is a complete join of a clique and an independent set, so their join is $S_{p+q,k}$. Conversely, the two-valued law in Lemma 3 verifies equality for every $S_{n,k}$. $\square$

3 From the mean to the exact variance maximum

If $1 \leq X \leq k$ almost surely and $\mu = \mathbb{E}X$, expansion of $(X - 1)(k - X)$ gives the identity

$$\text{Var}(X) = (\mu - 1)(k - \mu) - \mathbb{E}[(X - 1)(k - X)]. \quad (5)$$

In particular $\text{Var}(X) \leq (\mu - 1)(k - \mu) \leq (k - 1)^2/4$. This is the bounded-variable variance inequality commonly associated with Bhatia and Davis [1]; the identity supplies the whole argument needed here.

Lemma 5. *For every integer $n \geq 3$, the unique maximum of $F_n(k)$ over $k \in \{1, \dots, n\}$ occurs at $k = k_n$. Moreover, $k_n > n/2$, and*

$$\frac{(k - 1)^2}{4} < F_n(k_n) \quad (1 \leq k \leq n/2).$$

Proof. Put $f(k) = n^2 F_n(k)$. Direct expansion yields

$$f(k + 1) - f(k) = -k(4k^2 - 3nk + n).$$

The quadratic has its smaller root in $(0, 1)$ when $n > 2$, so this difference changes from positive to negative at most once for integer $k \geq 1$. To locate the change, write $n = 4q + r$, $0 \leq r \leq 3$, and set $k_0 = \lfloor (3n + 2)/4 \rfloor$. The values of $D(k) = f(k + 1) - f(k)$ are

r	k_0	$D(k_0 - 1)$	$D(k_0)$
0	$3q$	$4(2q - 1)(3q - 1)$	$-12q^2$
1	$3q + 1$	$3q(5q - 1)$	$-(3q + 1)(7q + 2)$
2	$3q + 2$	$2q(3q + 1)$	$-2(3q + 2)(5q + 3)$
3	$3q + 2$	$(3q + 1)(11q + 2)$	$-(q + 1)(3q + 2)$

For the values $n \geq 3$ covered by each row, the third column is positive and the fourth negative. This proves the unique maximum. The formula directly gives $k_n > n/2$.

For the remaining assertion, $k = 1$ is immediate. For $k \geq 2$, fix k and regard

$$g(b) = \frac{kb(b - 1)^2}{(k + b)^2} \quad (b \geq k).$$

It is strictly increasing, because

$$\frac{g'(b)}{g(b)} = \frac{1}{b} + \frac{2}{b - 1} - \frac{2}{k + b} > 0.$$

For $b = n - k \geq k$, this gives

$$F_n(n - k) = g(n - k) \geq g(k) = \frac{(k - 1)^2}{4},$$

strictly if $k < n/2$. If $k = n/2$, the unique maximizing index is larger than k , so the claimed strict inequality still follows. $\square$

Proof of Theorem 1. Put $k = \alpha(G)$ and $L = 1 + k(k - 1)/n$. In either class, the preceding mean lemmas give $\mu(G) \geq L$. If $k \leq n/2$, (5) and Lemma 5 give $v(G) < F_n(k_n)$.

Suppose $k > n/2$. Then $L > (k+1)/2$, and the function $(x-1)(k-x)$ is strictly decreasing for $x \geq L$. Hence

$$v(G) \leq (\mu(G) - 1)(k - \mu(G)) \leq (L - 1)(k - L) = F_n(k) \leq F_n(k_n).$$

If equality holds, Lemma 5 forces $k = k_n$, and strict decrease forces $\mu(G) = L$. The equality cases in the mean lemmas give $G \cong S_{n,k_n}$. Conversely, the two-valued law for S_{n,k_n} gives variance $F_n(k_n)$. Finally, $k_n/n \rightarrow 3/4$, so $F_n(k_n)/n^2 \rightarrow (3/4)^3(1/4) = 27/256$. $\square$

Corollary 6. *Within either graph class, the smallest expected fraction of a maximum independent set returned by random greedy is exactly*

$$\min_{|V(G)|=n} \frac{\mu(G)}{\alpha(G)} = \min_{1 \leq k \leq n} \left(\frac{1}{k} + \frac{k-1}{n} \right).$$

The minimizing integers k are exactly those satisfying

$$k(k-1) \leq n \leq k(k+1).$$

In particular the minimum is $2/\sqrt{n} + O(1/n)$.

Proof. Divide the mean bound by k ; $S_{n,k}$ attains equality for each k . Consecutive values of the displayed expression differ by $1/n - 1/(k(k+1))$, proving the integer criterion. Taking $k = \sqrt{n} + O(1)$ proves the asymptotic assertion. $\square$

4 Stability for split graphs

Proposition 7. *Let G be a split graph of order n with a split partition whose independent side has maximum size $k \geq 2$. Put*

$$\delta = \mu(G) - \left(1 + \frac{k(k-1)}{n} \right), \quad R = \sum_{i \in I} |C \setminus N(i)|.$$

Then

$$\delta \geq \frac{R^2}{kn(n-1)}. \quad (6)$$

If $k \geq n/2$, then

$$F_n(k) - v(G) \geq (k-1) \left(\frac{2k}{n} - 1 \right) \delta + \delta^2. \quad (7)$$

Consequently, if split graphs G_n satisfy $v(G_n)/n^2 \rightarrow 27/256$, then their maximum independent sides have size $(3/4 + o(1))n$, and they differ from S_{n,k_n} in $o(n^2)$ edge additions or deletions, after a suitable labeling.

Proof. Lemma 3 and $k-1 + r_i \leq n-1$ give

$$\delta \geq \frac{1}{n(n-1)} \sum_i r_i^2 \geq \frac{R^2}{kn(n-1)},$$

where the last step is Cauchy–Schwarz. Substitute $\mu = L + \delta$ into (5). Exact expansion gives

$$F_n(k) - v(G) = (k-1) \left(\frac{2k}{n} - 1 \right) \delta + \delta^2 + \mathbb{E}[(X_G - 1)(k - X_G)],$$

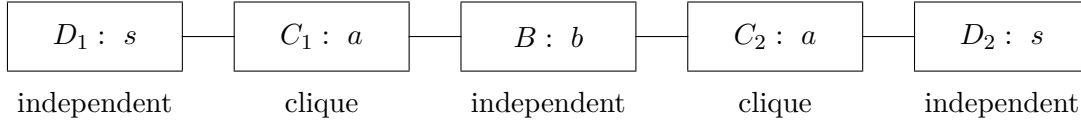
which proves (7).

For the asymptotic assertion, indices with $k \leq n/2$ have $v(G_n)/n^2 \leq 1/16$, so eventually $k > n/2$. Uniformly in k , $F_n(k)/n^2 = (k/n)^3(1 - k/n) + O(1/n)$. On $[1/2, 1]$, the function $x^3(1 - x)$ has its unique maximum $27/256$ at $x = 3/4$. The variance assumption therefore forces $k/n \rightarrow 3/4$. Equation (7) then implies $\delta = o(n)$, and (6) implies $R = o(n^2)$. Adding the R missing cross-edges produces $S_{n,k}$. Changing the side membership of $|k - k_n| = o(n)$ vertices changes at most $n|k - k_n| = o(n^2)$ additional edges and produces S_{n,k_n} . $\square$

5 A five-part counterexample outside the two classes

The following observation makes exact large examples manageable. Replace the vertices of a small template graph by nonempty vertex classes, each a clique or an independent set, and completely join precisely the classes corresponding to adjacent template vertices. The first vertex in an independent class, if accepted, guarantees that every vertex in that class will eventually be accepted: all its neighboring classes have become blocked. A selected clique class contributes exactly one vertex. Thus one may remove an accepted class and its neighbor classes at once, assigning it reward equal to its size for an independent class and one for a clique class. At each step the next class is chosen with probability proportional to its remaining size. This follows from the uniform relative order on the surviving vertices. In particular this reduction is exact, not an approximation by independent clocks.

For Theorem 2, name the classes D_1, C_1, B, C_2, D_2 , in path order. More generally let $|D_1| = |D_2| = s$, $|C_1| = |C_2| = a$, and $|B| = b$, with $N = 2s + 2a + b$. The two C classes are cliques. The other classes are independent.



Each displayed line represents all edges between its endpoint classes.

Lemma 8. *The accepted count in this graph has the following law (coincident outcomes are combined):*

count	probability
2	$\frac{2a^2}{N(a+s)}$
$s+1$	$\frac{2as}{(a+s)(a+b+s)}$
$b+2s$	$1 - \frac{2a^2}{N(a+s)} - \frac{2as}{(a+s)(a+b+s)}$

Proof. If the first class is B , both clique classes are blocked, giving $b+2s$. If it is C_1 (or symmetrically C_2), it contributes one and blocks D_1, B ; the surviving C_2, D_2 pair contributes one with probability $a/(a+s)$ or s with probability $s/(a+s)$. If the first class is D_1 (or D_2), it contributes s and blocks C_1 . The remaining graph is a clique of size a completely joined to an independent set of size $b+s$, so it contributes one with probability $a/(a+b+s)$ or $b+s$ otherwise. This lists all outcomes. Their first two probabilities are

$$\frac{2a}{N} \frac{a}{a+s}, \quad \frac{2a}{N} \frac{s}{a+s} + \frac{2s}{N} \frac{a}{a+b+s}.$$

The second simplifies as stated because $(a+s) + (a+b+s) = N$. $\square$

Proof of Theorem 2. Take $(a, b, s) = (6t, 34t, t)$. Lemma 8 gives

$$\frac{x}{\mathbb{P}(X_{H_t} = x)} \mid \begin{array}{ccc} 2 & t+1 & 36t \\ 3/14 & 12/287 & 61/82. \end{array}$$

The probabilities sum to one. Consequently

$$\mu(H_t) = \frac{7698t + 135}{287}, \quad v(H_t) = \frac{20155992t^2 - 2071572t + 55821}{82369}.$$

Since $k_{48t} = 36t$, the maximum from Theorem 1 is

$$F_{48t}(36t) = \frac{3}{16}(36t - 1)^2.$$

Subtracting gives the exact gap

$$v(H_t) - F_{48t}(36t) = \frac{3(748400t^2 - 5117816t + 215343)}{1317904}. \quad (8)$$

For $t = 7 + u$, $u \geq 0$, the polynomial in parentheses is

$$748400u^2 + 5359784u + 1062231 > 0.$$

This proves the finite-order claim. Dividing the variance by $(48t)^2$ and taking the limit yields the stated constant. Its excess over $27/256$ is exactly

$$\frac{839833}{7907424} - \frac{27}{256} = \frac{46775}{63259392} > 0.$$

□

For example H_7 has 336 vertices, variance $973198425/82369 = 11815.1055\dots$, and exceeds the split/cograph maximum $189003/16 = 11812.6875$ by $3186693/1317904$.

This family also identifies the failure of the mean argument outside our classes. Its independence number is $36t$: the independent classes together have that size, whereas an independent set meeting a clique class has size at most $t + 1$ or consists of two clique vertices. Hence

$$\mu(H_t) - \left(1 + \frac{\alpha(H_t)(\alpha(H_t) - 1)}{48t}\right) = \frac{253 - 204t}{1148} < 0 \quad (t \geq 2).$$

There is no conflict with the earlier lemmas. Taking one vertex from each of four consecutive classes produces an induced four-vertex path, excluding cographs. Taking two vertices from each clique class produces two disjoint edges as an induced subgraph, excluding split graphs.

6 The unrestricted asymptotic problem

Proposition 9. *Let $M_n = \max_{|V(G)|=n} v(G)$, with no restriction on G . The limit*

$$c_* = \lim_{n \rightarrow \infty} \frac{M_n}{n^2} = \sup_{m \geq 1} \frac{M_m}{m^2}$$

exists, and

$$\frac{839833}{7907424} \leq c_* \leq \frac{1}{4}.$$

Proof. Replace each vertex of a fixed m -vertex graph G by an independent class of size r , retaining complete joins across each original edge. The first arrival times of the classes are independent, identically distributed continuous random variables. Their order is therefore uniform. The class reduction above gives $X_{G[\overline{K}_r]} \stackrel{d}{=} rX_G$, so its variance is $r^2v(G)$.

For arbitrary n , let $r = \lfloor n/m \rfloor$, and pad this graph with $n - rm$ isolated vertices. They add a deterministic count. Therefore $M_n \geq r^2v(G)$, and

$$\liminf_{n \rightarrow \infty} \frac{M_n}{n^2} \geq \frac{v(G)}{m^2}.$$

Taking the supremum over all fixed G proves that the liminf is at least the displayed supremum; the reverse limsup bound follows from its definition. The lower bound follows from H_t . The upper bound follows from $1 \leq X_G \leq n$ and (5), which gives $v(G) \leq (n-1)^2/4$. $\square$

The value of c_* is not determined here. Nor do we claim that the five-part construction is optimal, or that 336 is the smallest order at which the split/cograph maximum can be exceeded. Those are separate extremal questions. Another direction is to characterize graph classes on which the sharp mean bound in Section 2 remains valid.

7 Exact verification and limitations

The accompanying Python program uses rational arithmetic. It verifies the mean and variance conclusions, including equality classification, for the split graphs and cographs in the complete Graph Atlas through seven vertices: 257 nonempty split graphs and 287 nonempty cographs. It also checks all 1252 nonempty atlas graphs, enumerates every vertex permutation of every labeled graph through four vertices to independently check the first-choice recursion, and verifies the five-part probability law against a separate class-state recursion over a grid of integer class sizes. The finite-difference maximizer is checked for $3 \leq n \leq 1000$.

These finite checks support implementation correctness but are not used as proofs of the theorems. In fact every atlas graph through seven vertices satisfies the proposed unrestricted variance bound, while Theorem 2 disproves that extension. The proofs, rather than the finite search, establish the scope of the result.

References

- [1] R. Bhatia and C. Davis, A better bound on the variance, *American Mathematical Monthly* **107** (2000), 353–357. [doi:10.1080/00029890.2000.12005203](https://doi.org/10.1080/00029890.2000.12005203).
- [2] D. G. Corneil, H. Lerchs, and L. Stewart Burlingham, Complement reducible graphs, *Discrete Applied Mathematics* **3** (1981), 163–174. [doi:10.1016/0166-218X\(81\)90013-5](https://doi.org/10.1016/0166-218X(81)90013-5).
- [3] D. Gamarnik and D. A. Goldberg, Randomized greedy algorithms for independent sets and matchings in regular graphs: exact results and finite girth corrections, *Combinatorics, Probability and Computing* **19** (2010), 61–85. [arXiv:0807.1277](https://arxiv.org/abs/0807.1277).
- [4] M. Krivelevich, T. Mészáros, P. Michaeli, and C. Shikhelman, Greedy maximal independent sets via local limits, *Random Structures & Algorithms* **64** (2024), 986–1015. [doi:10.1002/rsa.21200](https://doi.org/10.1002/rsa.21200).
- [5] M. D. Penrose and A. Sudbury, Exact and approximate results for deposition and annihilation processes on graphs, *Annals of Applied Probability* **15** (2005), 853–889. [doi:10.1214/105051604000000765](https://doi.org/10.1214/105051604000000765).
- [6] M. Shaikh, Variance of random greedy independent sets in triangle-free graphs, preprint (2026). [arXiv:2609.14826](https://arxiv.org/abs/2609.14826).