

Approximating the Maximum Weighted Independent Set Problem Empirically

Department of Computer Science & Software Engineering, College of Engineering

Student Researcher: Sue Sue, Advisor: Dr. Daniel Frishberg



CAL POLY
College of Engineering

Motivation

Why Graphs Matter

- Real-world networks (social media, internet, biology) follow scale-free patterns with “superstar” hubs
- Barabási-Albert (B-A) graphs model realistic preferential attachment vs. Erdős-Rényi (E-R) random connections

Why Independent Sets Matter

- Collections of non-interfering nodes - crucial for scheduling, clustering, distributed algorithms
- Closely related to clique problem - crucial for social networks
- Maximum Independent Set is NP-hard, requiring efficient approximation algorithms

Research Questions

We investigated three greedy algorithms for the Independent Set problem: *nearest-neighbor chain* (NNC) and *priority-queue* (PQ)-based algorithms, and a “naive” greedy algorithm.

Our questions are:

- Is NNC better than PQ and naive greedy for finding weighted independent sets in E-R and/or B-A graphs?
- How close to optimal are greedy algorithms in B-A graphs?
- Can we recover *planted* independent sets with a fast, simple algorithm?

Experimental Design

Graph Types:

- Erdős-Rényi (E-R): Random graphs where each edge appears independently with probability p
- Barabási-Albert (B-A) with star/complete initialization: Scale-free networks using preferential attachment where new nodes connect to high-degree nodes)

Weight Settings: Unweighted and Weighted graphs
(Node) Weights: uniformly chosen in range 1-10,000
Sizes: 500-5000 nodes (increments of 500)
Average Degree: $2/N$, $10/N$, $1/N$, $\log(N)/N$, $1/4$
Sample Size: 10 graphs per configuration

Three Algorithms Compared:

A *greedy algorithm* finds an independent set by repeatedly selecting a node and eliminating its neighbors. Different greedy algorithms use different rules for selecting nodes:

- Priority Queue (PQ):** selects the node with the lowest *degree* (fewest neighbors)
- Nearest Neighbor Chain (NNC):** selects a node with low degree, but not the lowest
- Naive Greedy:** selects any node

Key Results

RQ1:

NNC emerges as the best choice among the algorithms we evaluated for the **weighted** independent set problem, across both graph types. Performance analysis shows that PQ and NNC outperform Naive, with PQ achieving only slightly better quality than NNC while requiring 2-3x longer runtime. **NNC delivers 98-99% of PQ's solution quality.** While NNC's theoretical running time is slower than Naive, it's only by a constant factor, maintaining the **same asymptotic complexity** as Naive. This suggests that NNC may achieve the best trade-off between solution quality and computational cost.

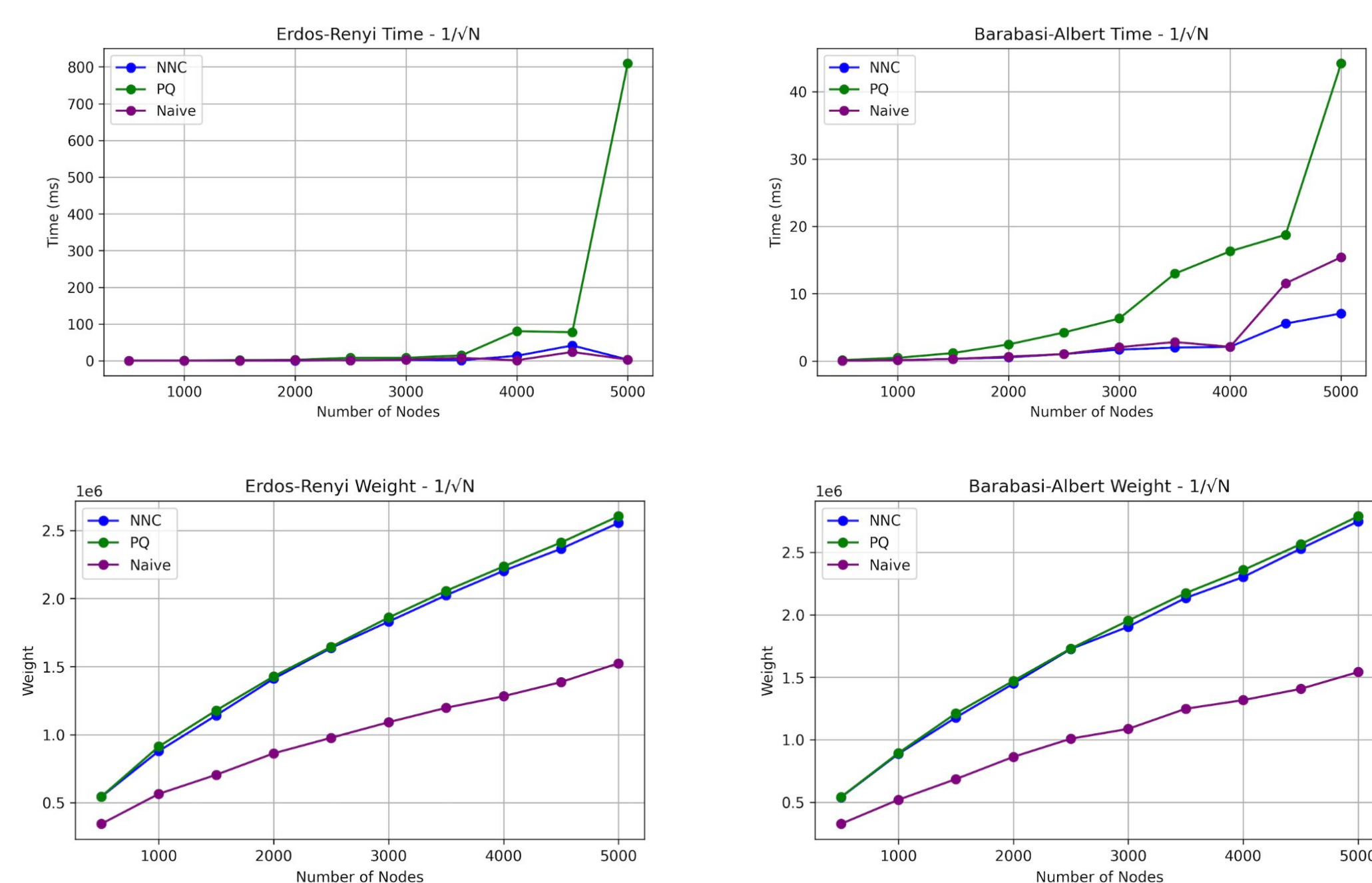


Fig.1 Performance comparisons of PQ, NNC, and Naive on weighted E-R and B-A graphs with complete initialization of medium density of $1/\sqrt{n}$

RQ2:

Our empirical analysis shows that in unweighted E-R graphs, the naive greedy algorithm achieves ratios around 1.0 (measured as algorithm output size divided by known theoretical bounds, which are half-optimal [1,4]), which agrees with the known theory [4].

More importantly, our analysis provides evidence that greedy algorithms **find similar-sized sets in B-A graphs** to those found in E-R graphs, with all three algorithms showing consistent behavior patterns across both graph types. This suggests that the particular random graph model used may not change the underlying property of random graphs that makes them amenable to greedy algorithms.

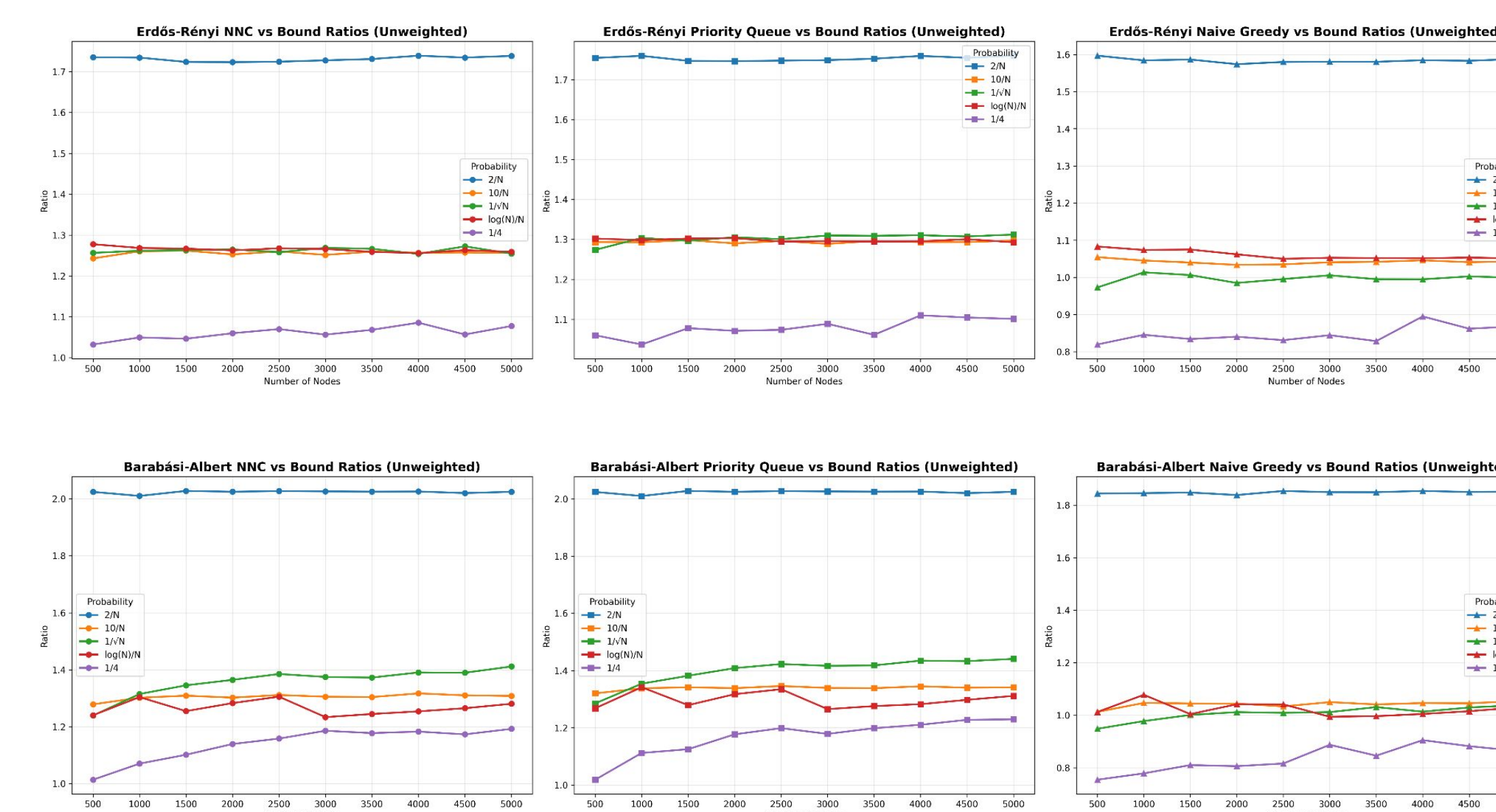


Fig. 2 Interesting bound results in the unweighted graphs of B-A and E-R

RQ3:

Our planted independent set experiments in E-R graphs confirm that **degree-based greedy algorithms** (PQ and NNC) significantly outperform naive greedy in recovering planted structures. However, recovery performance was strongly dependent on the planted set size as larger sets ($N/4$) showed substantially better recovery than smaller sets, supporting the theoretical threshold of $C/(N \log N)$ for effective recovery [3].

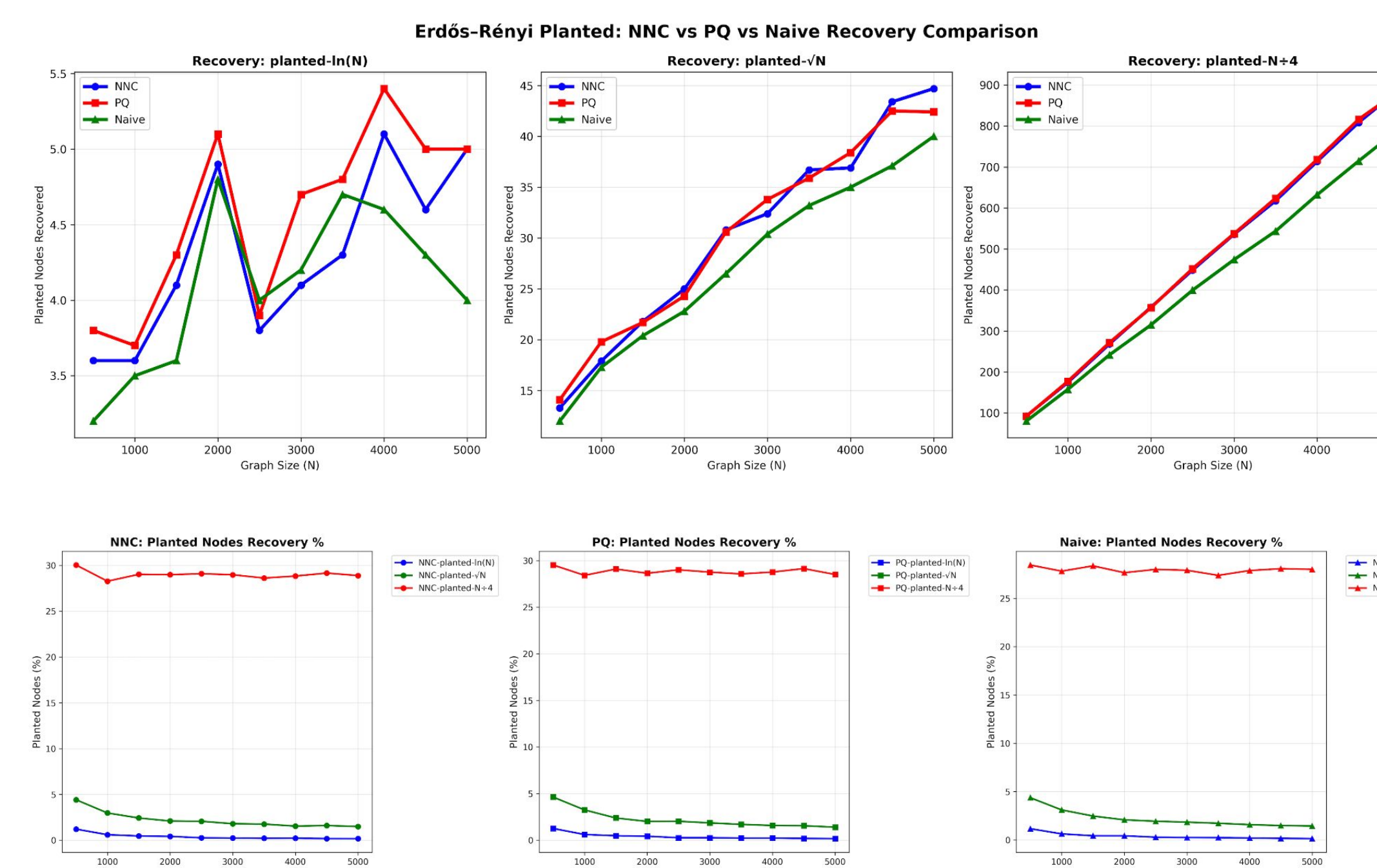


Fig. 3 Planted sets recovery comparisons of PQ, NNC, and Naive on unweighted E-R graphs

Discussion

Our findings provide practical guidance: for **RQ1**, NNC delivers 98-99% of PQ's quality with better theoretical runtime, making it better for weighted problems. However, mixed empirical runtime results require validation with larger graphs.

For **RQ2**, our results suggest that half-optimality performance may extend from E-R to B-A graphs, as all algorithms show similar behavior patterns across both graph types. However, important caveats apply: we do not know the actual optimal independent set sizes in B-A graphs (although some bounds are known [2]), providing only a partial answer to RQ2.

Additionally, our observations of ratios above 1.0 (especially in the $2/N$ case) contradict theory, suggesting issues with our bound formulas—or perhaps effects that occur only in moderate-sized graphs ($n=500-5000$).

For **RQ3**, the planted set experiments confirm that simple greedy approaches achieve competitive recovery performance, with effectiveness dependent on planted set size as theory predicts.

Next Steps

- Formal mathematical proofs to determine if observed performance differences are theoretically expected
- Parallel processing implementation to test larger networks (currently limited to 5,000 nodes)
- Scaling experiments to reveal true asymptotic behavior in real-world network sizes

Acknowledgements

The research was funded by Cal Poly's Computer Science and Software Engineering Department. The authors acknowledge substantial contributions by Luke Fanguna to the experimental design.

REFERENCES

- Grimmett, G.R. & McDiarmid, C.J.H. (1975). *On colouring random graphs*. Mathematical Proceedings of the Cambridge Philosophical Society, 77(2), 313-324.
- Kovalenko, K. (2020). *On the independence number and chromatic number of generalized preferential attachment models*. Discrete Applied Mathematics, 271, 78-91.
- Kučera, L. (1995). *Expected complexity of graph partitioning problems*. Discrete Applied Mathematics, 57(2-3), 193-212.
- Wein, A.S. (2020). *Optimal low-degree hardness of maximum independent set*. Retrieved from <https://arxiv.org/abs/2010.06563>.