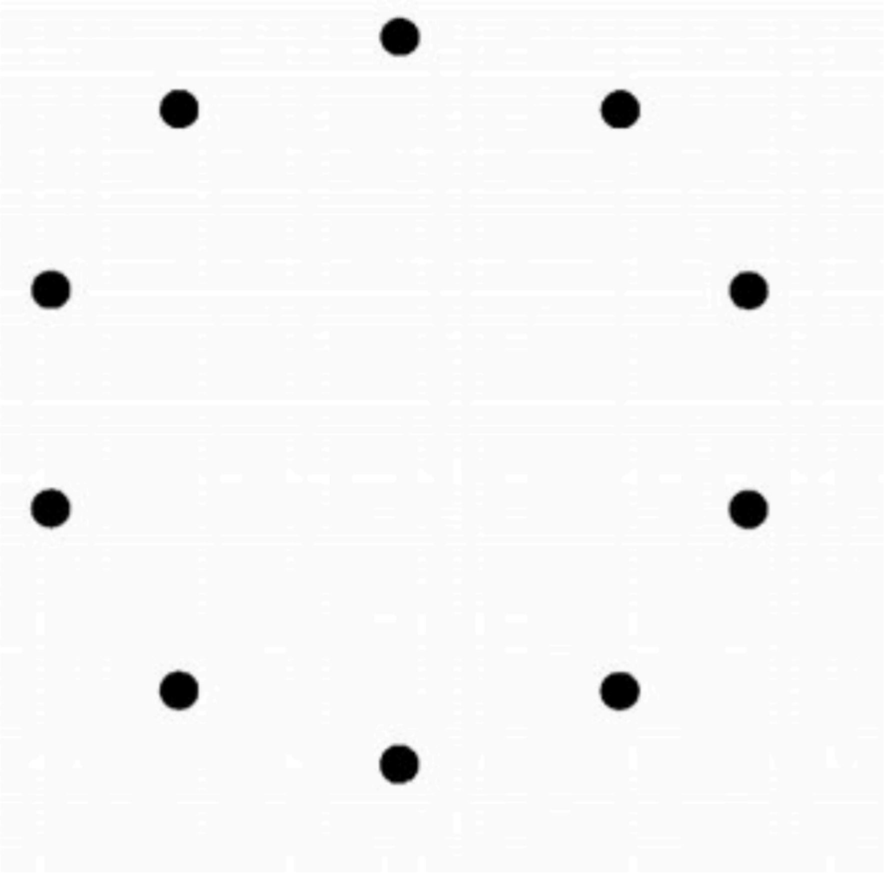
Robust Estimation for Erdös-Rényi Graphs

Arjan Chakravarthy, George Chemmala, Heon Lee

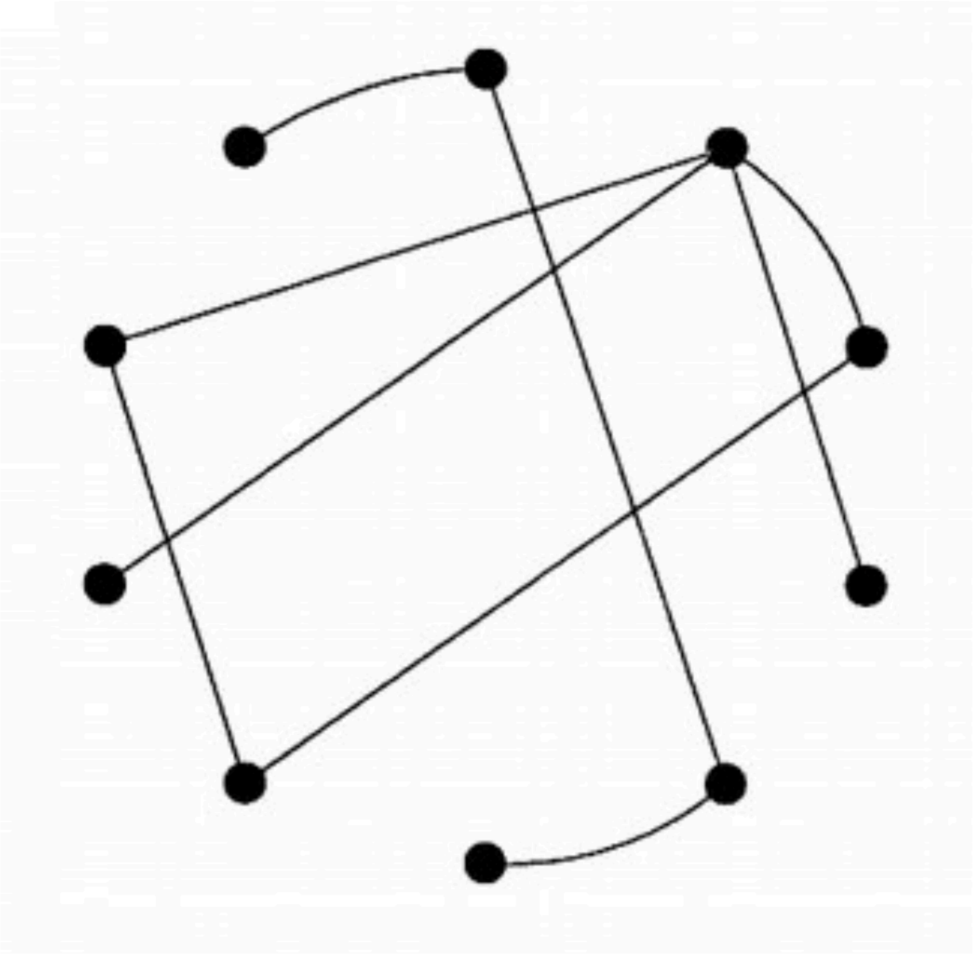
Erdös-Rényi Random Graph

- Parameters: n, p
- n nodes



Erdös-Rényi Random Graph

- Parameters: n, p
- n nodes
- Each edge independently has probability p of existing



Motivation

- Natural / baseline model for real world phenomena
  - social networks
  - epidemiology

Can we estimate p given that an ε-fraction of the nodes have been corrupted?

Adversarial Model

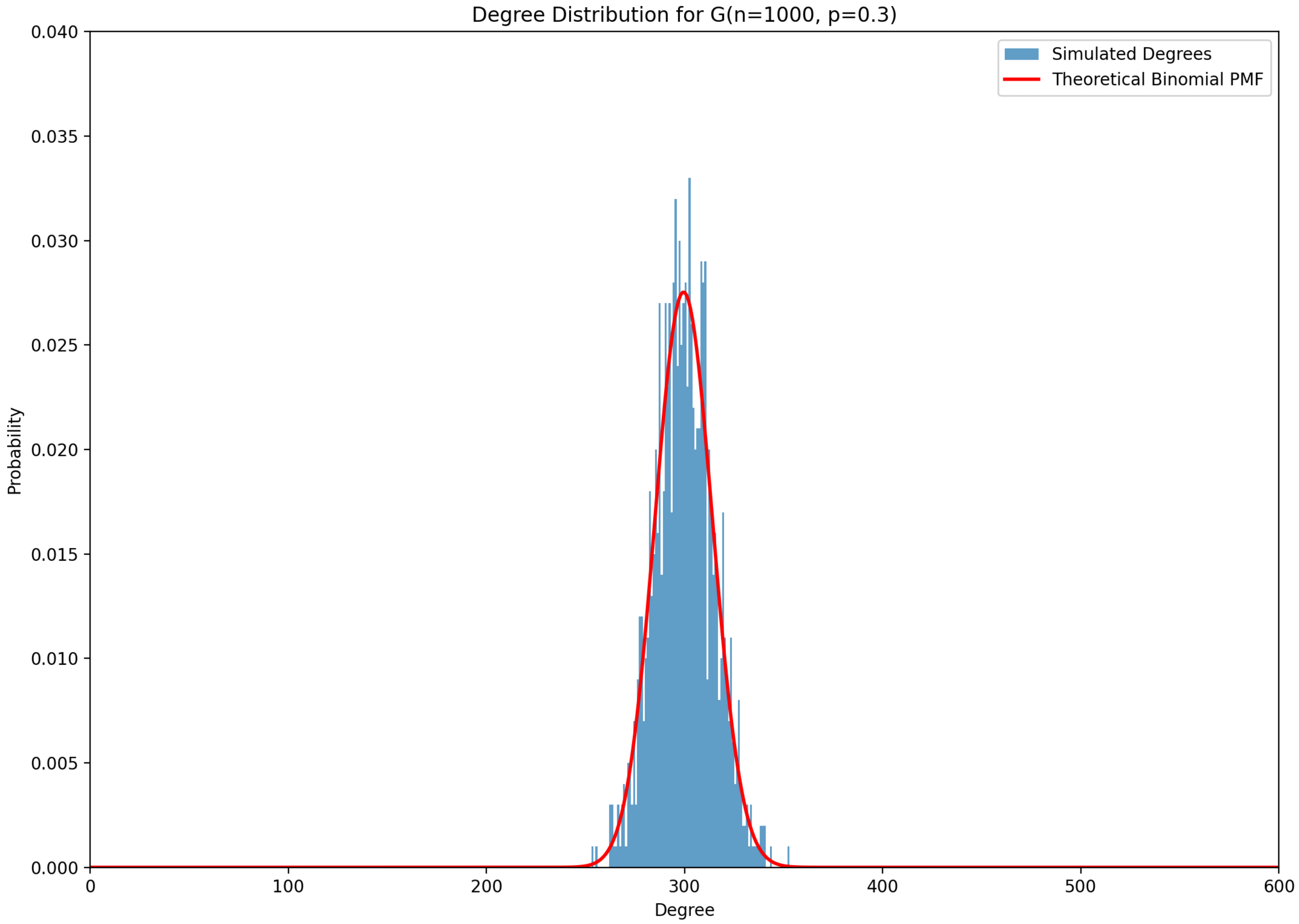
- ε-**omniscient** adversary:
  - observes the realization of graph $G \sim G(n, p)$
  - choose $B \subset V$ and rewire its edges
- ε-**oblivious** adversary:
  - choose $B \subset V$ and distribution on B without observing the realization of G
- (q, ε)-omniscient / oblivious adversary

Existing Algorithms

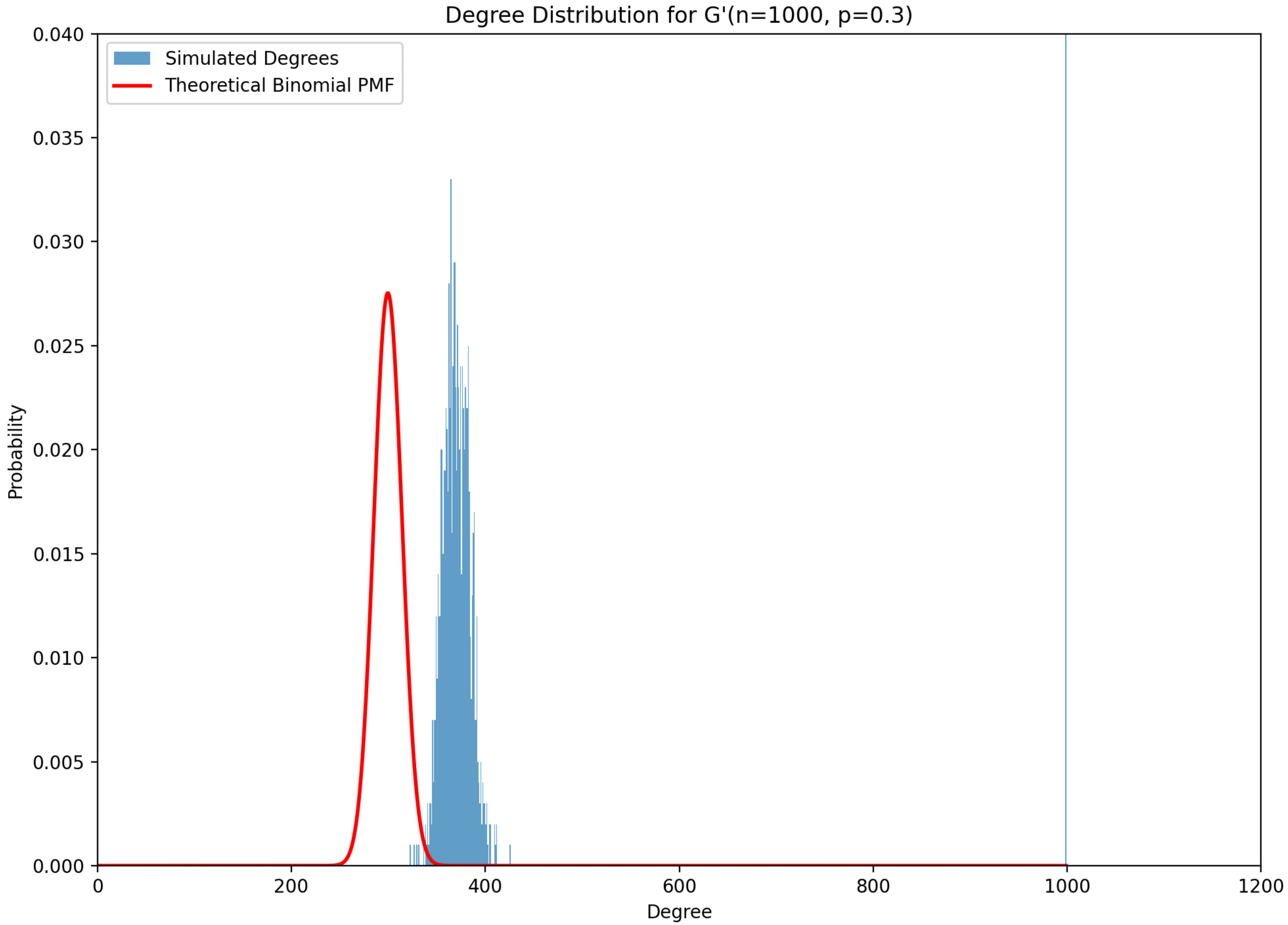
Methods	Error	Runtime	Authors
Mean/Median	$O(\varepsilon)$	$O(n)$	Acharya et al.
Prune then Mean/Median	$O(\varepsilon^2)$	$O(n\log(\epsilon n))$	Acharya et al.
Spectral Method	$O\left(\frac{\sqrt{p(1-p)\log n}}{n}+\frac{\varepsilon\sqrt{p(1-p)\log(1/\varepsilon)}}{\sqrt{n}}+\frac{\varepsilon}{n}\log n\right)$	$O\left(\varepsilon n^3\frac{\log n}{\gamma}\right)$	Acharya et al.
Median – Mean	?	?	Us
Variance Method	?	?	Us

Mean / Median

Degrees of Unperturbed Graph



Degrees of Perturbed Graph

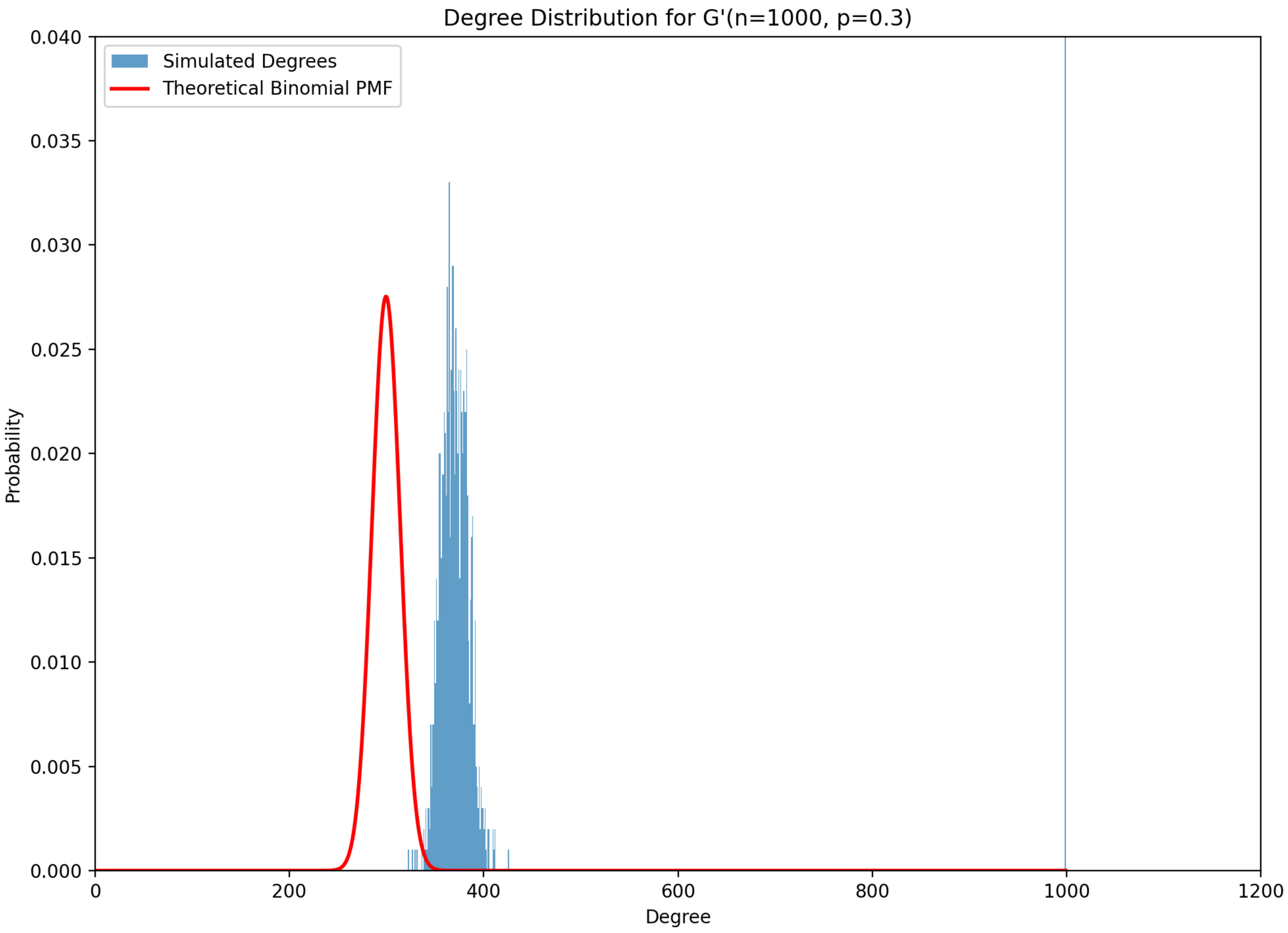


Errors and Time Complexity

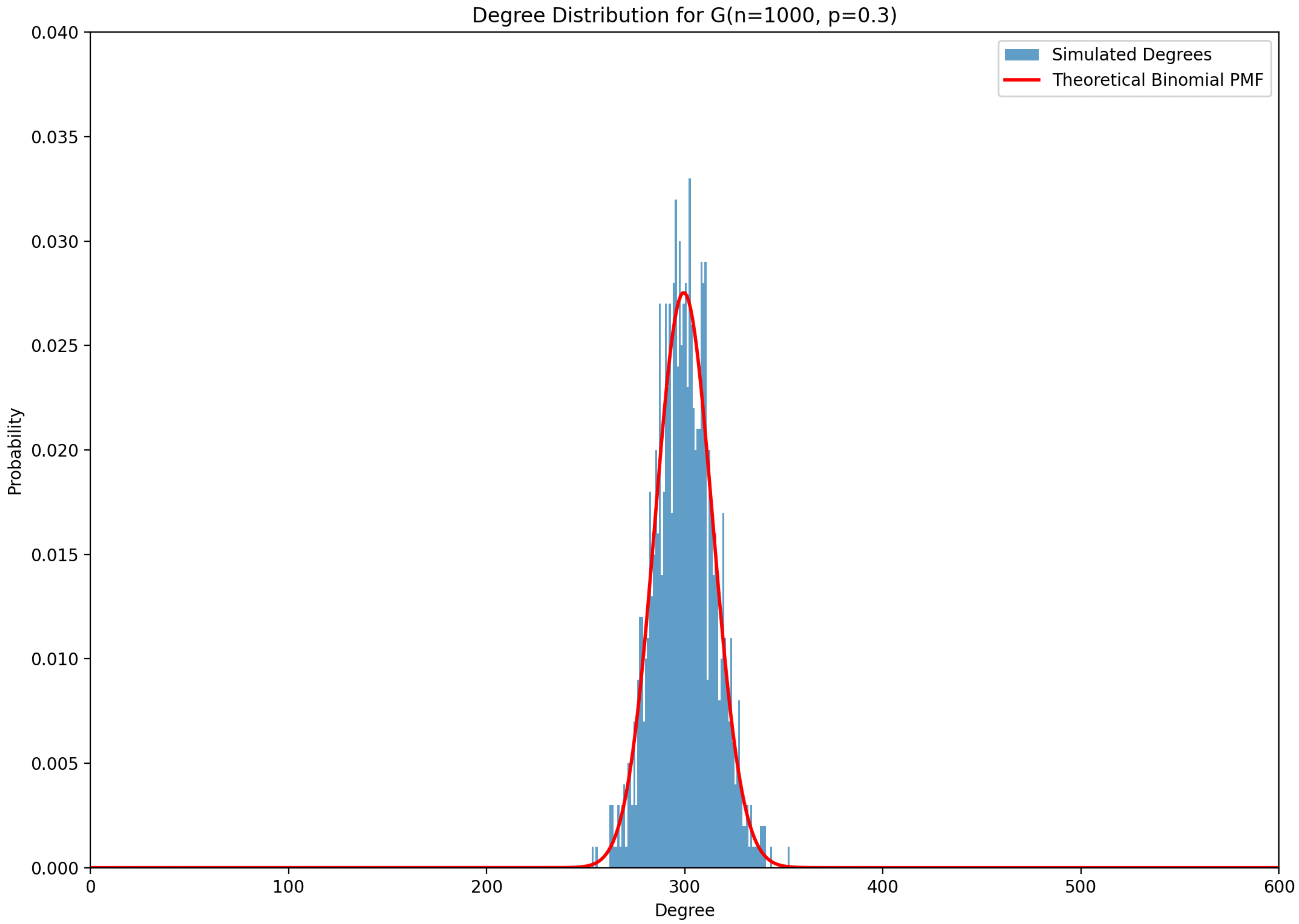
- Error: $O(\varepsilon)$
- Time complexity: $O(n)$

Prune Then Mean / Median

Degrees of Perturbed Graph



Degrees of Unperturbed Graph



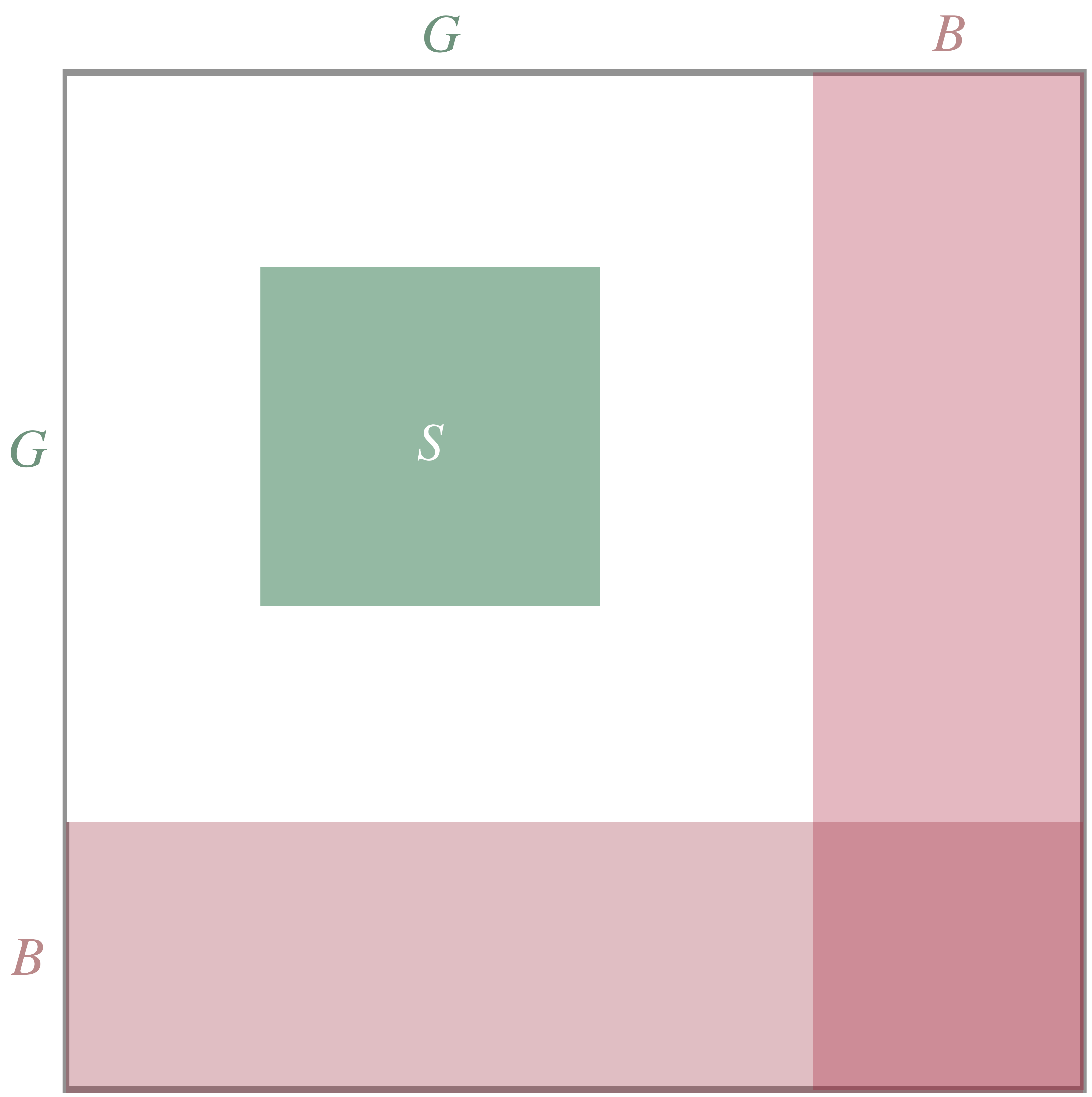
Errors and Time Complexity

- Error: $O(\varepsilon^2)$, $O(\varepsilon)$
- Time complexity: $O(n \log(n\varepsilon))$

Spectral Method

Main idea

- We know that an uncorrupted subgraph should have an estimated p close to the frequency of 1s in the adjacency matrix
- $||(A - P_S)_{S \times S}||_2$ is small for uncorrupted subgraph S


G
B
G
S
B

Spectral Method

- We start with the entire graph
- Compute our metric $\|(A - P_S)_{S\times S}\|_2$
- Use the largest eigenvector to find the troublemaker node
- Remove the node
- Repeat

Our Algorithms

Notation

- $G = (V, E)$ is our original graph
- $G' = (V, E')$ is our perturbed graph
- D, D' are degrees of V, V'
- $\bar{d}, \bar{d}'$ is the mean of D and D'
- $\tilde{d}, \tilde{d}'$ is the median of D and D'

Median - Mean

Main idea

- Median and mean have errors $O(\varepsilon)$
- Observed median is ~2x closer to the true p than mean
- We can exploit this by taking the median minus the mean to find how far away the median is from p

How well can we estimate p?

Proof Sketch

- $\bar{d}' \approx d + \varepsilon(q-p)(2n-1) - \varepsilon^2 n(q-p)$
- $\tilde{d}' \approx d + \varepsilon n(q-p) - \varepsilon\sqrt{np(1-p)}$
- Claim: $\hat{d} := \tilde{d}' - \dfrac{\bar{d}' - \tilde{d}'}{1-\varepsilon}$ is good
- $\tilde{d}' - \dfrac{\bar{d}' - \tilde{d}'}{1-\varepsilon} = \tilde{d}' - \left(\sum_{k=0}^{\infty} \varepsilon^k\right)(\bar{d}' - \tilde{d}') = d + \dfrac{1}{1-\varepsilon}(q - p + \varepsilon\sqrt{np(1-p)})$
- Then $\hat{p} = \dfrac{\hat{d}}{n} = p + \dfrac{\frac{\varepsilon}{1-\varepsilon}(q-p+\varepsilon\sqrt{np(1-p)}) - p}{n} = p + O\left(\dfrac{\varepsilon p}{(1-\varepsilon)n} + \dfrac{\varepsilon^2\sqrt{p(1-p)}}{(1-\varepsilon)\sqrt{n}}\right)$

Algorithm 1 Median - Mean Algorithm

Require: adjacency matrix A, epsilon ϵ

$D \leftarrow$Degrees of nodes in A

$\bar{p} \leftarrow$ normalized mean

$\tilde{p} \leftarrow$ normalized median

$y \leftarrow \frac{1}{1-\epsilon}|\bar{p} - \tilde{p}|$

if $\bar{p} \leq \tilde{p}$ **then**

  return $\tilde{p} - y$

else

  return $\bar{p} - y$

end if

Errors and Time Complexity

- Error: $O\left(\dfrac{\varepsilon p}{(1-\varepsilon)n} + \dfrac{\varepsilon^2\sqrt{p(1-p)}}{(1-\varepsilon)\sqrt{n}}\right)$
- Time complexity: $O(n)$

Variance Method

Variance Method

- The variance of degrees on an unperturbed graph should be: $(n-1)p(1-p)$
- Variance significantly increases on a perturbed graph with an adversary

Variance Algorithm

- Remove the εn bad nodes
- Find the node that decreases the variance the most
- Remove that node
- Repeat

Algorithm 2 Variance Algorithm

Require: adjacency matrix A, epsilon ϵ

$n \leftarrow$number of nodes

while $t = 0, t < \epsilon n$ **do**

  for node x in A **do**

    Modify degrees of nodes after removing x

    Compute variance on subgraph degrees

  end for

  $v \leftarrow$ node with largest variance deviation

  $A \leftarrow A \setminus v$

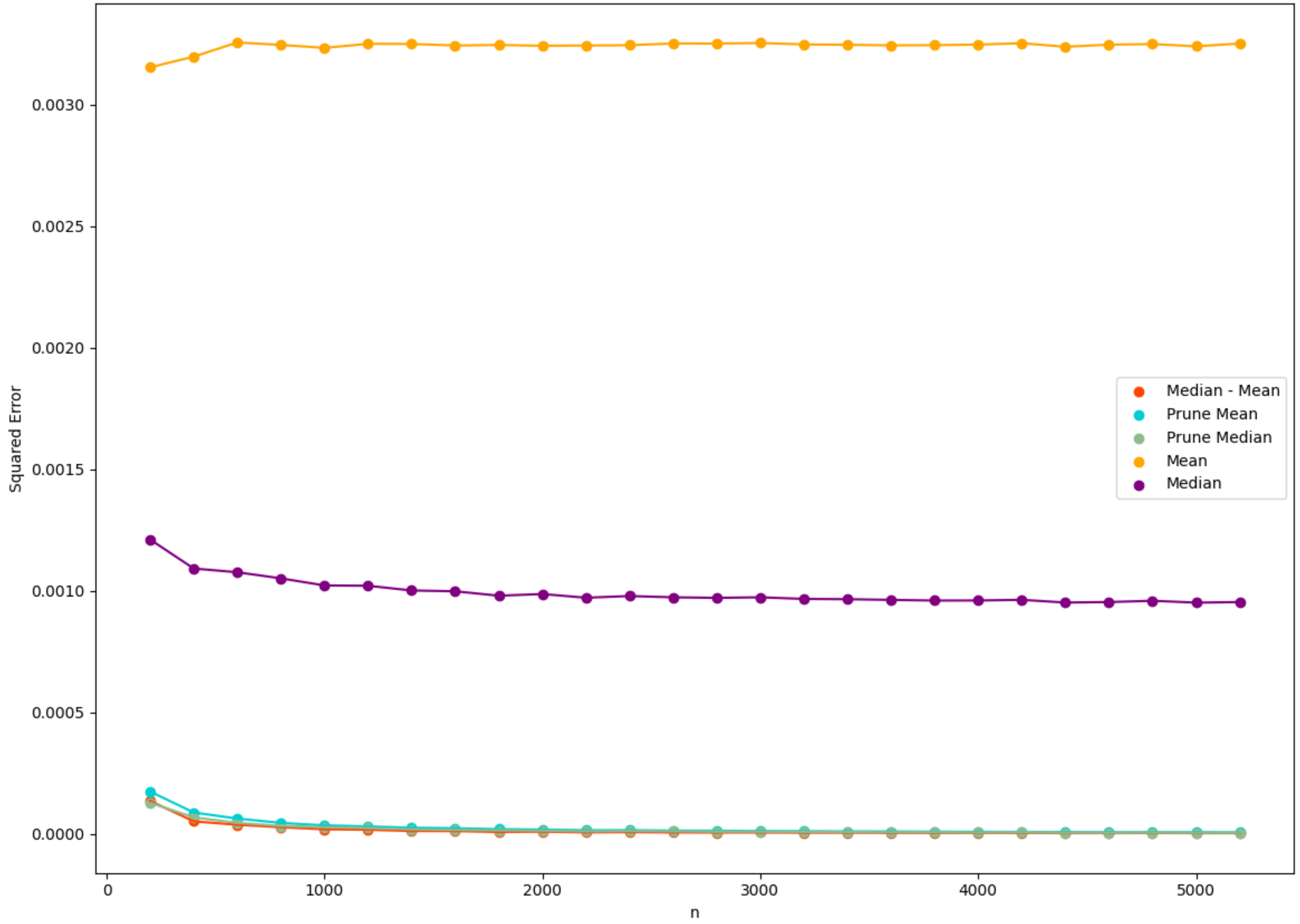
end while

Errors and Time Complexity

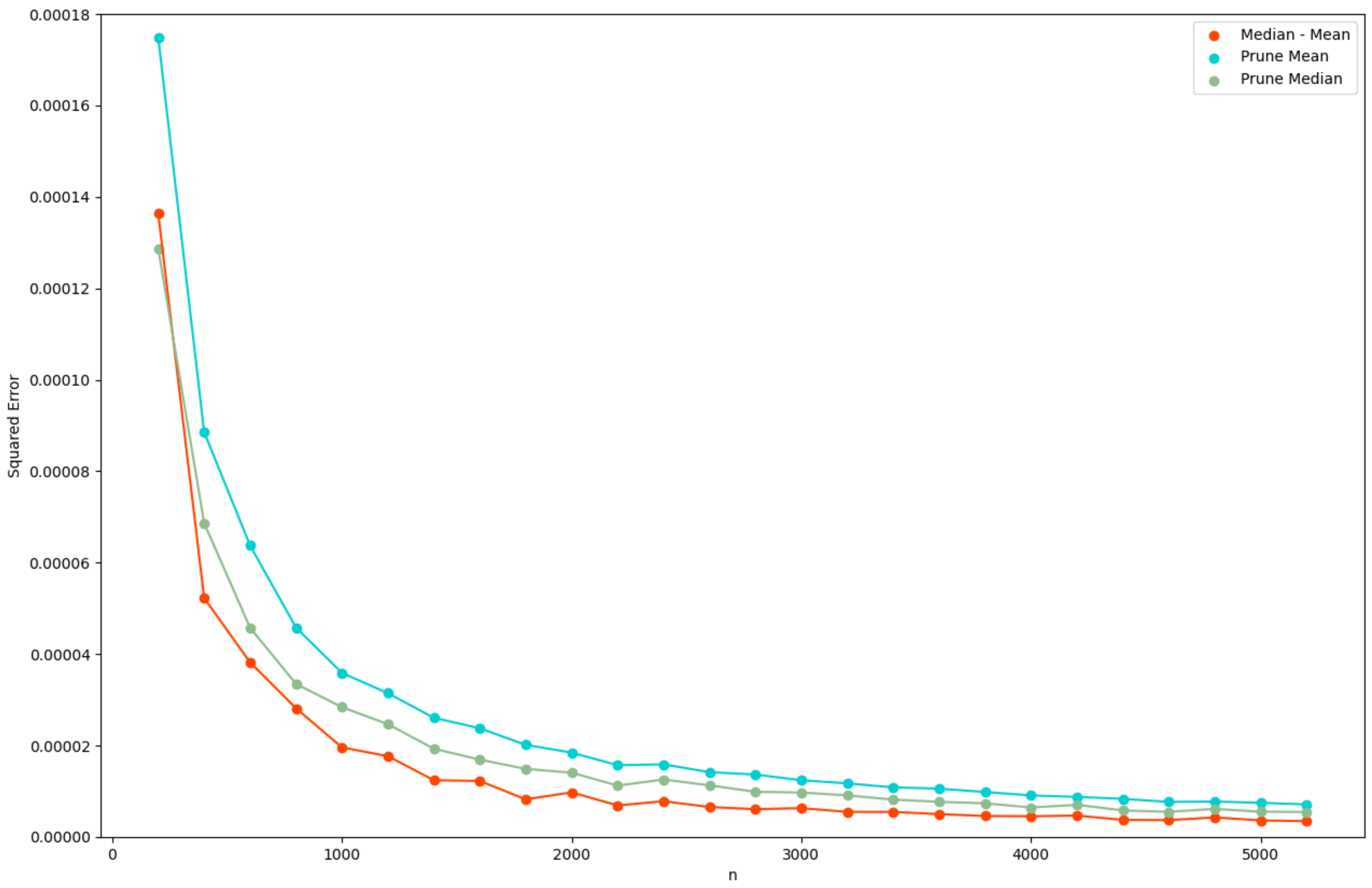
- Error: ?
- Time complexity: $O(\varepsilon n^3)$

Results

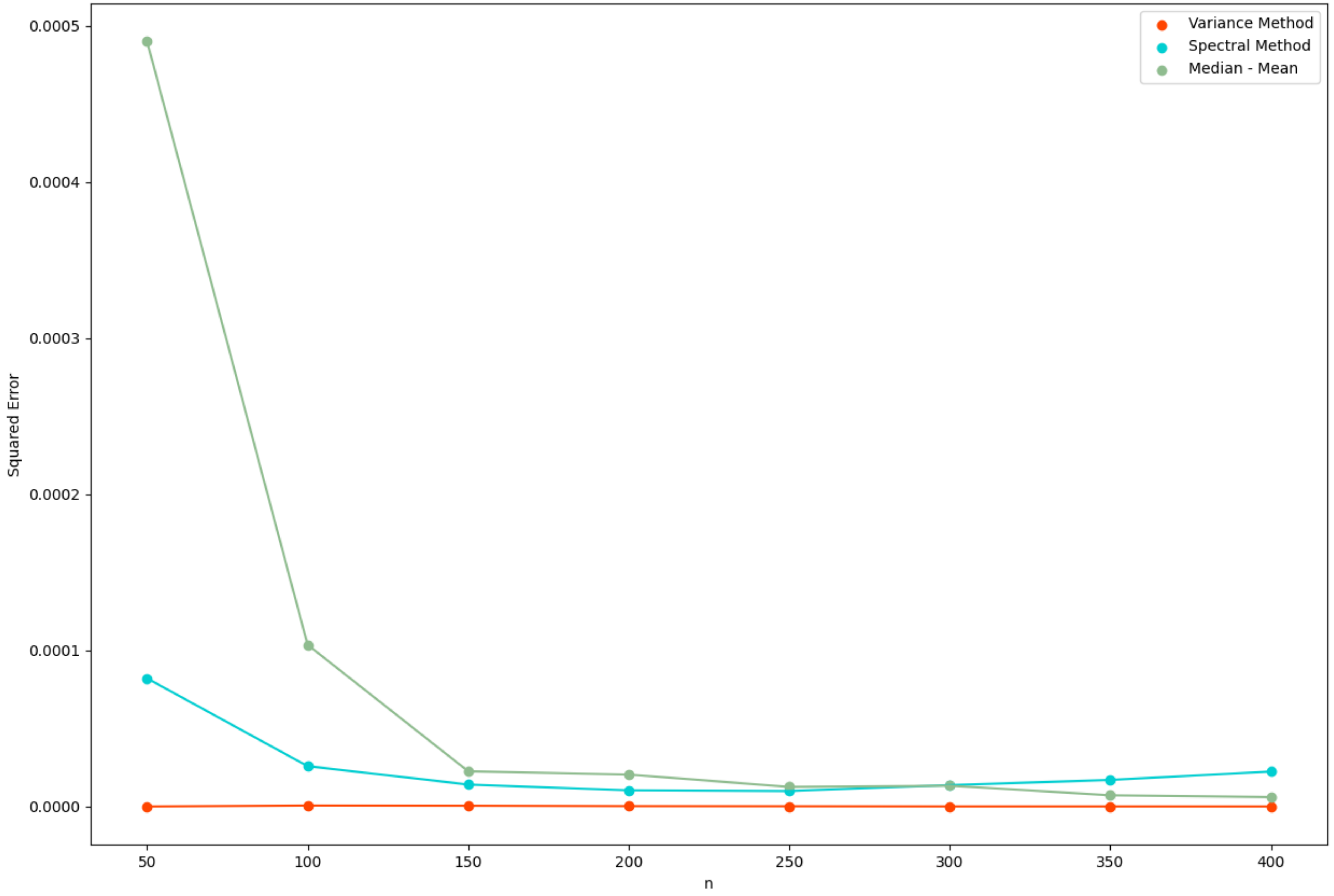
Results (Median - Mean)



Results (Median - Mean)



Results (Variance)



Results (Variance)

Results

Methods	Error	Runtime	Authors
Mean/Median	$O(\varepsilon)$	$O(n)$	Acharya et al.
Prune then Mean/Median	$O(\varepsilon^2)$	$O(n\log(\epsilon n))$	Acharya et al.
Spectral Method	$O\left(\frac{\sqrt{p(1-p)\log n}}{n}+\frac{\varepsilon\sqrt{p(1-p)\log(1/\varepsilon)}}{\sqrt{n}}+\frac{\varepsilon}{n}\log n\right)$	$O\left(\varepsilon n^3\frac{\log n}{\gamma}\right)$	Acharya et al.
Median – Mean	$O\left(\frac{\varepsilon p}{(1-\varepsilon)n}+\frac{\varepsilon^2\sqrt{p(1-p)}}{(1-\varepsilon)\sqrt{n}}\right)^*$	$O(n)$	Us
Variance Method	?	$O(\varepsilon n^3)$	Us

Next Steps

- Update error bounds on Median-Mean algorithm for general adversary
- Attempt to prove error bounds for variance method

Thank you!