Quilting Stochastic Kronecker Product Graphs to Generate Multiplicative Attribute Graphs

Hyokun Yun[1], S.V. N. Vishwanathan[1,2]

Departments of Statistics[1] and Computer Science[2], Purdue University

PURDUE UNIVERSITY

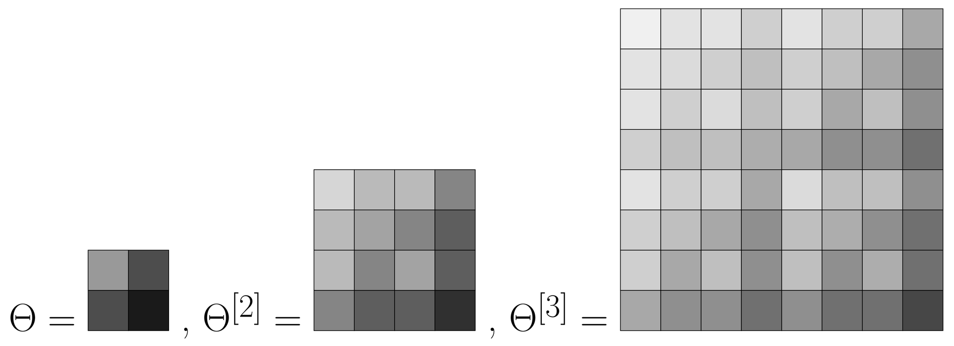
Abstract

Question: How to efficiently sample graphs from Multiplicative Attribute Graphs Model (MAGM)?

- **Our Answer: Quilting Algorithm**
  - The first **sub-quadratic** sampling algorithm for MAGM
  - **Time complexity**: $O\left((\log_2(n))^3 |E|\right)$ on mild conditions
    - n: number of *nodes* in the graph
    - $|E|$: number of *edges*
  - Exploit the close connection between Kronecker Product Graphs Model (KPGM) and MAGM
    - Sampling a graph from KPGM can be done very efficiently
    - We theoretically prove that it suffices to sample *small* number of KPGMs and *quilt* them
  - Can sample a graph with 8 million nodes and 20 billion edges in under 6 hours

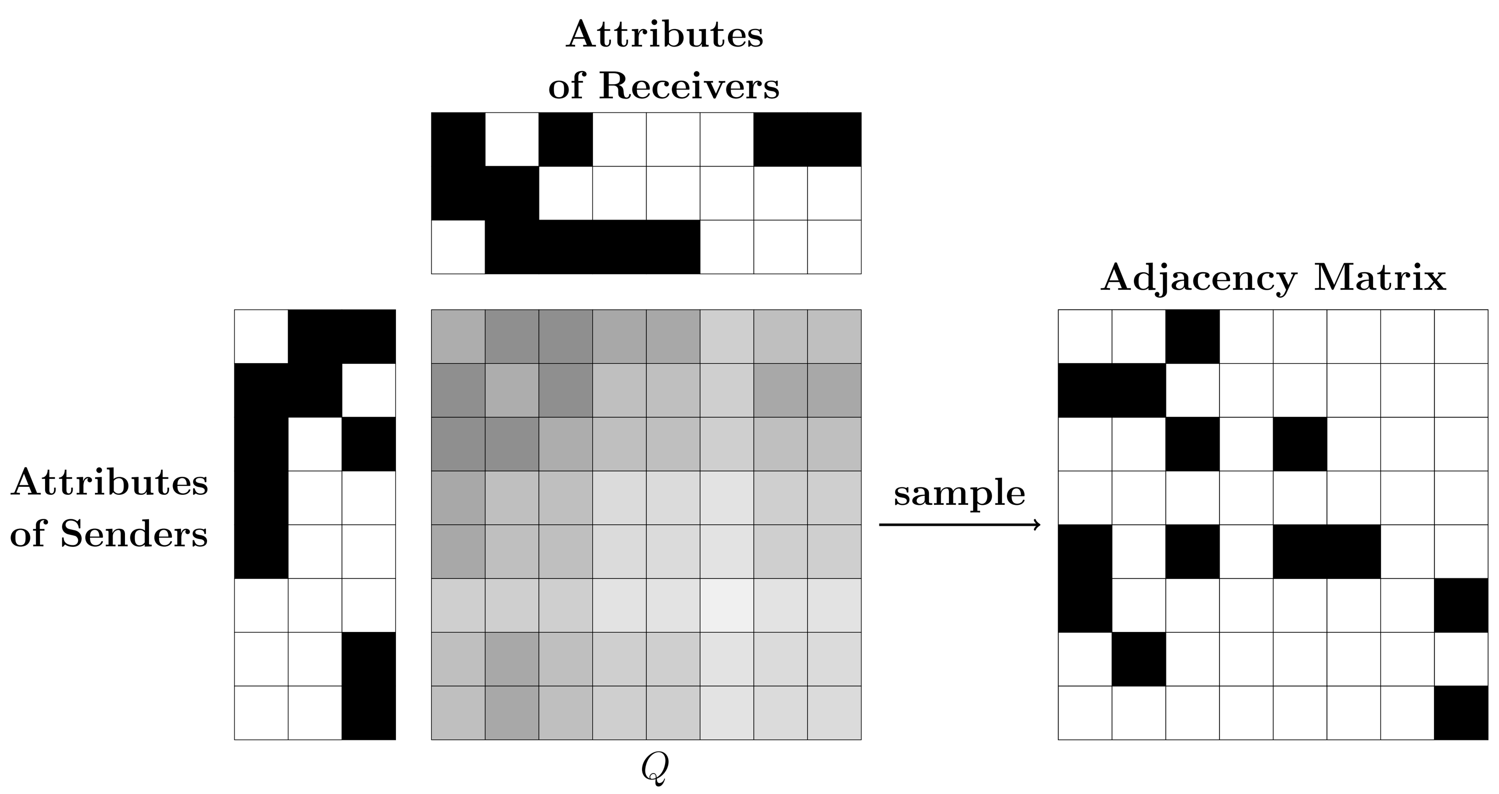
Kronecker Power of a Matrix

- Let Θ be a 2×2 matrix
- Kronecker power of the matrix shows the fractal structure



- KPGM[1] uses this Kronecker Power of a matrix to define a graph model

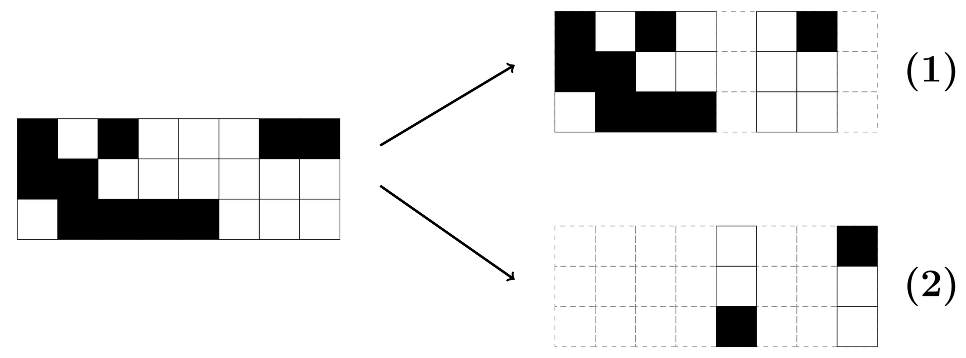
Multiplicative Attribute Graphs Model (MAGM)[2]



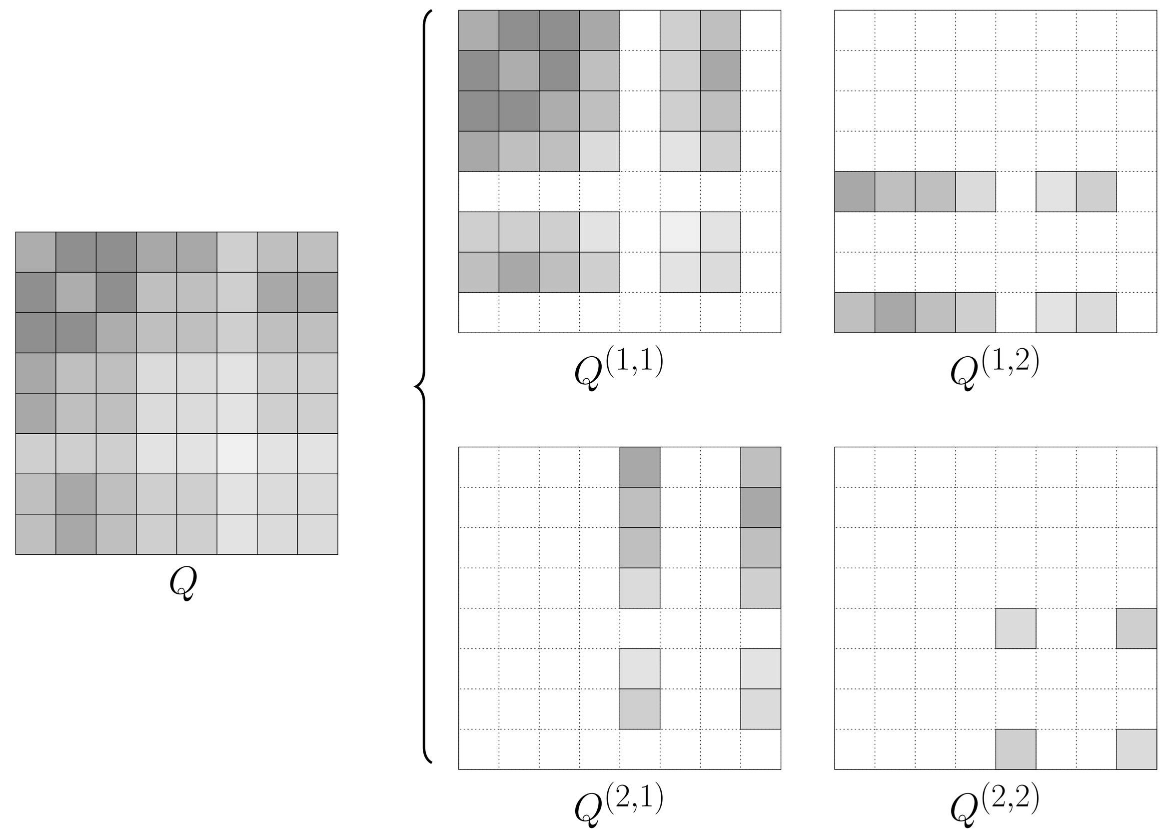
- Generalization of KGPM[1], increased modeling power (ex: power-law degree distribution)
- d attributes characterize nodes in the graph: each node either possesses or lacks (0 or 1) each attribute
- Probability of an edge between two nodes is determined by attributes
  - The effect of each attribute is multiplicative → *Multiplicative* Attribute Graphs
  - We call Q *the edge probability matrix*
- Naïvely sampling each entry of matrix will take $O(n^2)$ time!

The Quilting Algorithm[3]

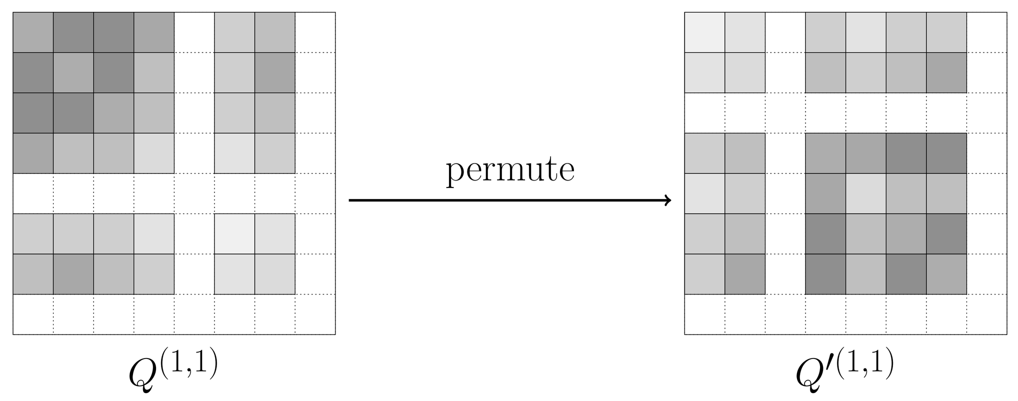
Step 1. Partition nodes such that each node has unique attribute configuration in its partition



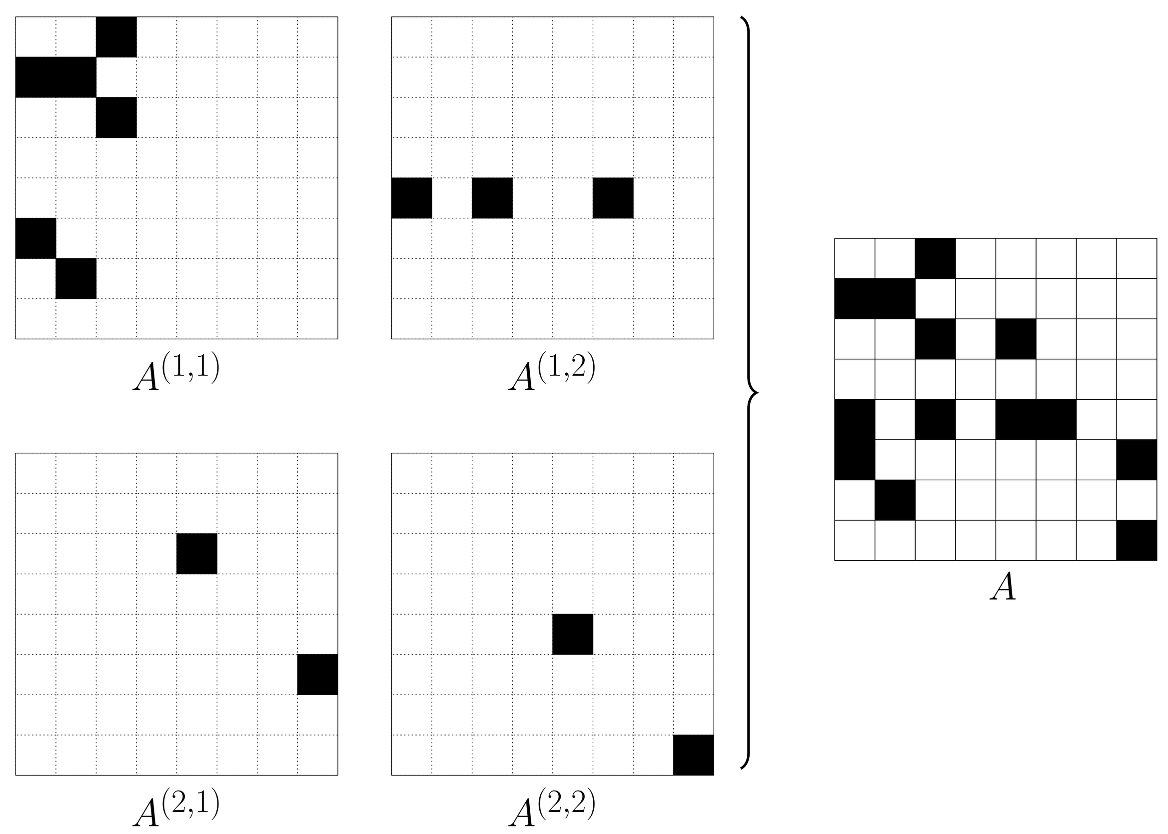
Step 2. Use this partition to *divide* the edge probability matrix Q



Step 3. When permuted, each $Q^{(k,l)}$ matrix becomes the submatrix of Kronecker Power Matrix. That is, it is a **subgraph of KPGM**, for which efficient sampling method $O(\log_2(n)\,|E|)$ exists[1].

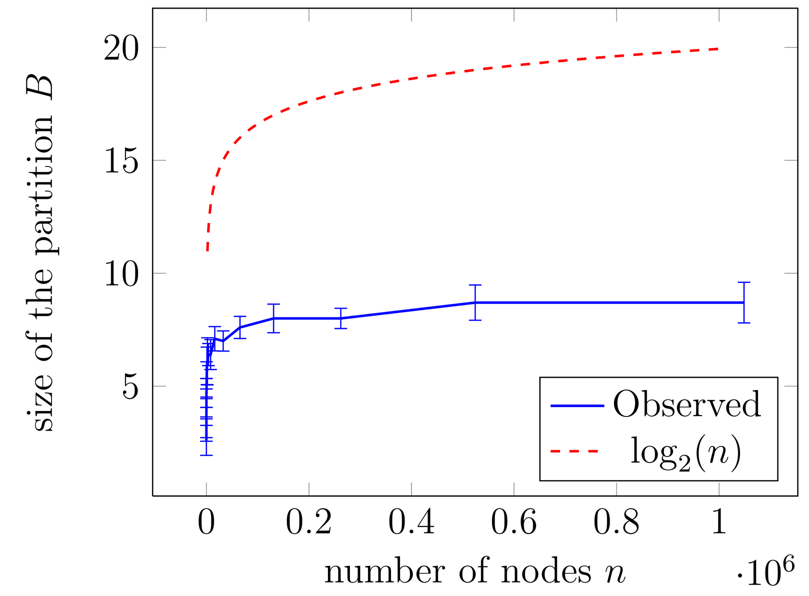


Step 4. We sample a graph for each $Q^{(k,l)}$ and *quilt* these pieces together to form the final graph



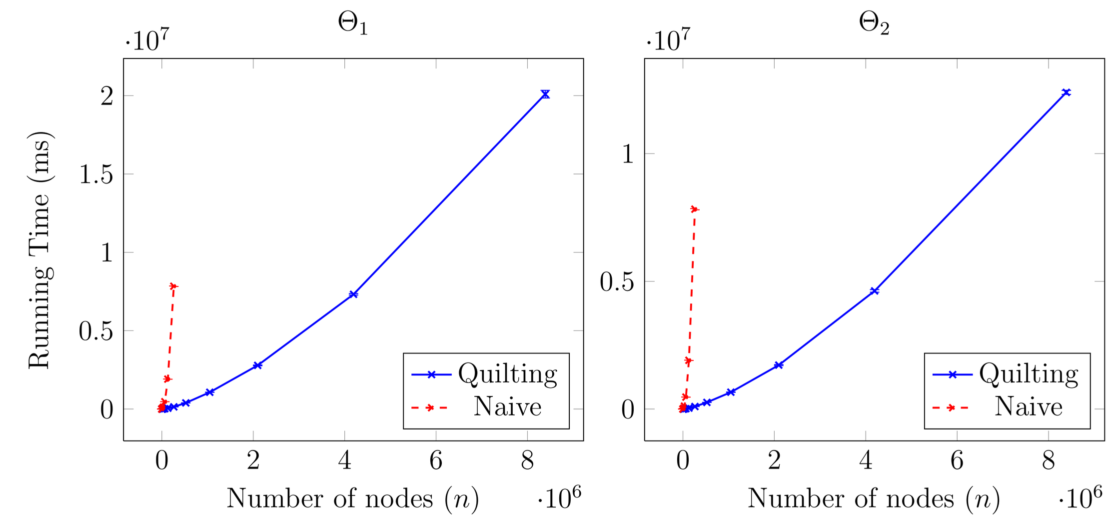
Performance

- **Question: How many partitions?**
  - With high probability $O(\log_2(n))$ under mild conditions
  - Therefore quilting $O\left((\log_2(n))^2\right)$ KPGM graphs suffices
  - Theoretically bounded by $O(\log_2(n))$ on mild conditions
  - Empirical simulations confirms the theory:

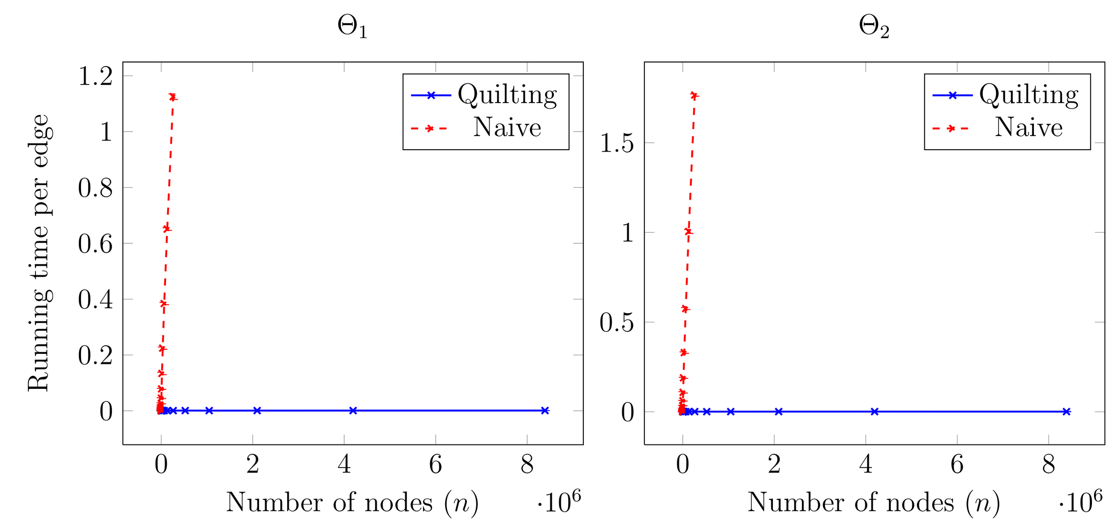


Experiments

- Choose two parameter values from the literature (Θ_1 and Θ_2)
- Increase number of nodes n to evaluate scalability (repeated 10 times each)
- Size of the graph vs. Total running time



- Size of the graph vs. Running time per edge



References

[1] J. Leskovec, D. Chakrabati, J. Kleinberg, C. Faloutsos and Z. Ghahramani, Kronecker Graphs: An Approach to Modeling Networks. *Journal of Machine Learning Research*, 11(3), 2010.
[2] M. Kim and J. Leskovec, Multiplicative Attribute Graph Model of Real-World Networks. *Algorithms and Models for the Web-Graph*, 62-73, 2010.
[3] H. Yun and S.V. N. Vishwanathan, Quilting Stochastic Kronecker Product Graphs to Generate Multiplicative Attribute Graphs. *Under Review*