Problem Presentation

CS552: Network Science

Sankalp Acharya
Dakshraj Sadashiv Kashid

Synthetic Social Network Generator

Introduction

Classic Models

- **Barabasi-Albert Model**: Creates scale-free networks with a power-law distribution.
- **Erdos-Renyi Model**: Creates small world random graphs.
- **Watts-Strogatz Model**: For generating random graphs that have both the clustering properties of regular graphs and the small-world properties of random graphs.

Recent Models

- **R-MAT Model**: A recursive approach to create a network which follows a power law and having a small diameter.
- **Kronecker-Graph Model**: Another recursive approach to create a network which follows a dense power law and having a fixed diameter.

Structural Properties of Real World Networks

- **Scale-free** networks, often exhibit the power-law distribution.
- **Small-world** phenomenon, having short path lengths and high clustering coefficients.
- Have a **community** structure.
- Exhibit **homophily, transitivity** & **reciprocity** (in case of directed networks).

Our approach will focus on improving the **clustering coefficient**, while maintaining the community structure as well as the scale-free nature of the network.

This will be achieved through applying **dynamic homophilicity** & **transitivity** in our algorithms.

Our Approach

Overview

1. Design a **Stochastic Block Model (SBM)** to create an initial model of the desired network, using a **dynamic homophily** approach.
2. Run a **Triadic Closure** algorithm on the obtained SBM.
3. Add the remaining number of nodes using *variants* of the **Preferential-Attachment-based** algorithm (we have designed 2 variants).

Stochastic Block Model

- It is a generative model for graphs that aims to capture the underlying community structure of a network.
- In an SBM, nodes are divided into different blocks or communities, and the probability of an edge between any two nodes depends on their **community memberships**.
- We assign the probabilities based on 2 uniform-random distributions: one for **intra-block** links and the other for **inter-block** links.
- With the addition of every link between blocks *i* & *j*, we change the value of p_{ij} (**Dynamic Homophily**).

Triadic Closures

- Refers to the phenomenon of **transitivity** in social networks, *i.e.* the tendency for two people who have a mutual friend to become friends themselves.
- It is one of the most fundamental properties of social networks and can be seen as a mechanism for the formation of clusters or communities within a network.
- We apply a **triadic closure algorithm** where the *tendency/probability* of *A* & *B* to have a link between them (if they didn't prior) is **proportional to the number of common neighbors** they have.

Preferential Attachment

- Preferential attachment is a mechanism that can generate scale-free networks, where a few highly connected nodes (also known as "hubs") coexist with many poorly connected nodes.
- In a network with preferential attachment, the probability of a new node connecting to an existing node is proportional to the degree (number of connections) of that node.
- We have proposed two variants: **Constant Preferential Attachment** & **Variable Preferential Attachment** algorithms.

Constant

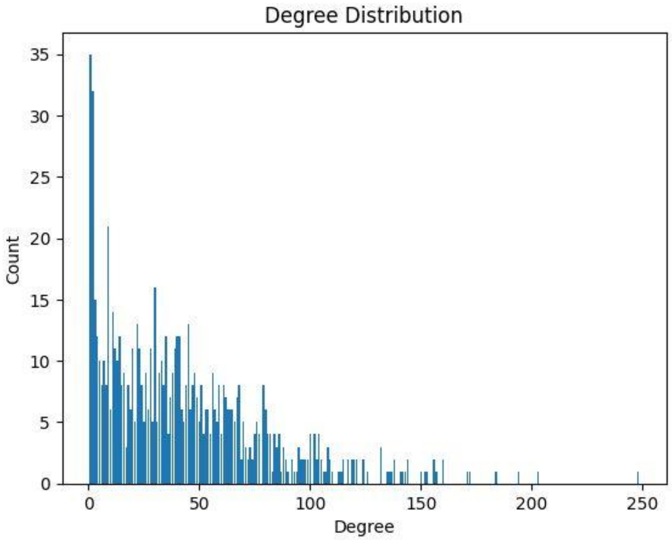
- Choose an existing node at random to link with (as per the $PA_{algorithm}$ in BA model), and link with it.
- Link to all neighboring nodes of the chosen node, until you either run out of links, or you run out of neighbors.
- In the latter case, repeat the above steps until the links are exhausted.

Variable

- Choose an existing node at random to link with (as per the $PA_{algorithm}$ in BA model), and link with it.
- Link to one of the neighboring nodes of the chosen node at random, with higher chance of being chosen being those nodes with higher degrees.
- Repeat the above steps until the links are exhausted.

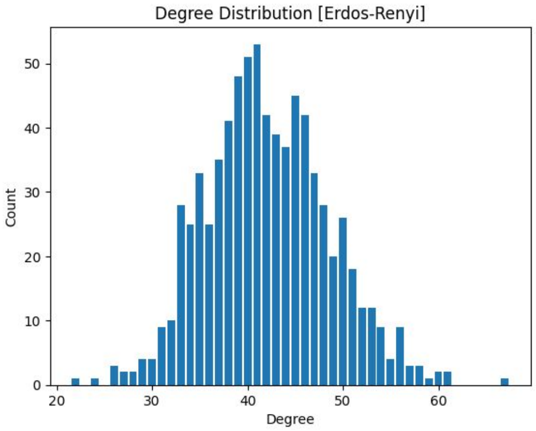
Results

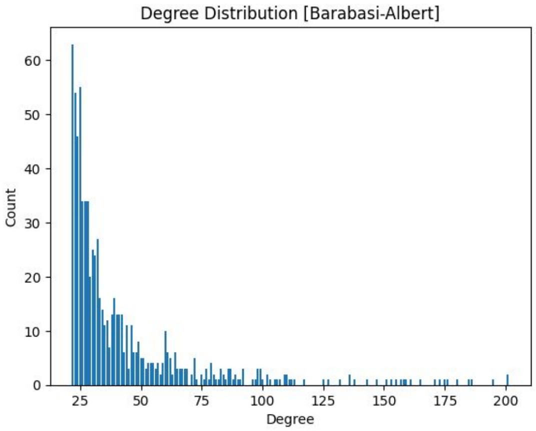
Caltech Dataset

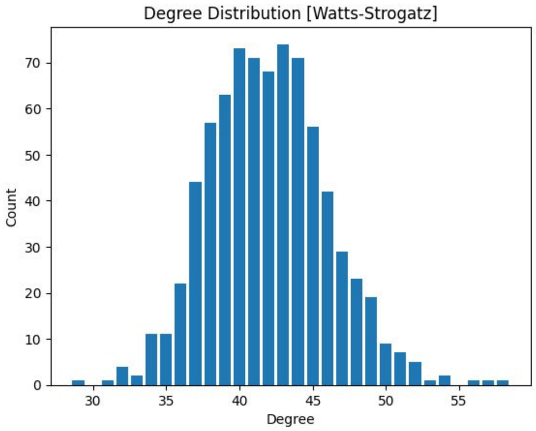


- **Nodes**: 769
- **Average Degree**: 43.31
- **Average Clustering Coefficient**: 0.41
- **Number of Communities**: 11

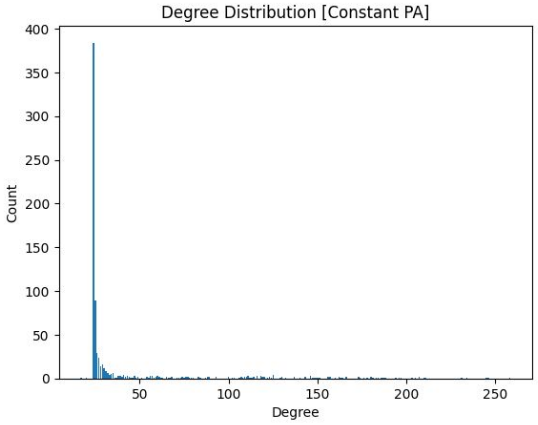
Degree Distributions generated by Classic Models

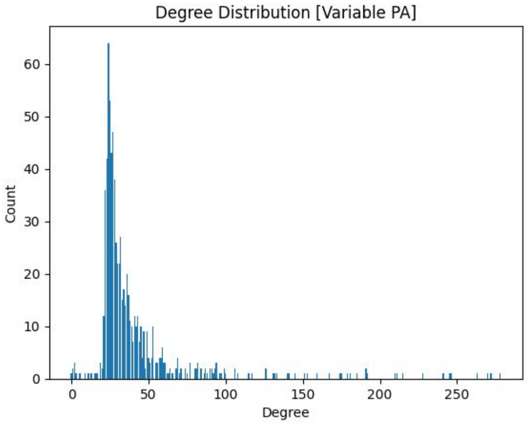






Degree Distributions generated by Our Models





Performance Analysis

S.No.	Network Models	Average Clustering Coefficient	Number of Communities
1.	CalTech Dataset	0.41	11
2.	Barabasi-Albert*	0.13	5
3.	Erdos-Renyi*	0.06	5
4.	Watts-Strogatz*	0.13	3
5.	SBM with Constant PA*	0.37	6
6.	SBM with Variable PA*	0.22	8

**** Indicates the values of these models displayed are averages over 10 runs.***

Problem 2:
Optimal Edge addition for maximum reduction in stress centrality

What is stress centrality?

$$SC(u) = \sum_{\{s \in V | s \neq u\}} \sum_{\{s \in V | t \neq s \ \& \ t \neq u\}} \sigma_{st}(u)$$

- Measure of importance and/or load called (*stress*) for a node.
- Edge addition: Distribution of traffic.
- Basic algorithm for computation:
  - BFS: *storing parent array*
  - Recursively tracing shortest paths.

Complexities:

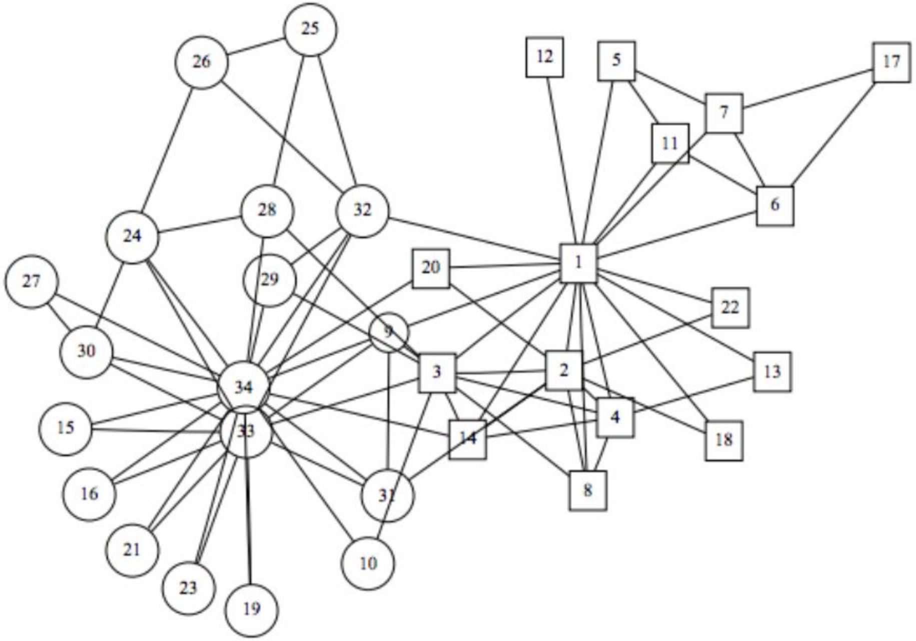
The **Brute Force** approach is to check for all the pair of edge.
: $O(n^2 \cdot C(stress\_centrality))$.

Our Solution:

- Greedy Algorithm.
- The Parallel Computing approach.
- The GNN approach using *GraphSAGE* to create node embeddings.

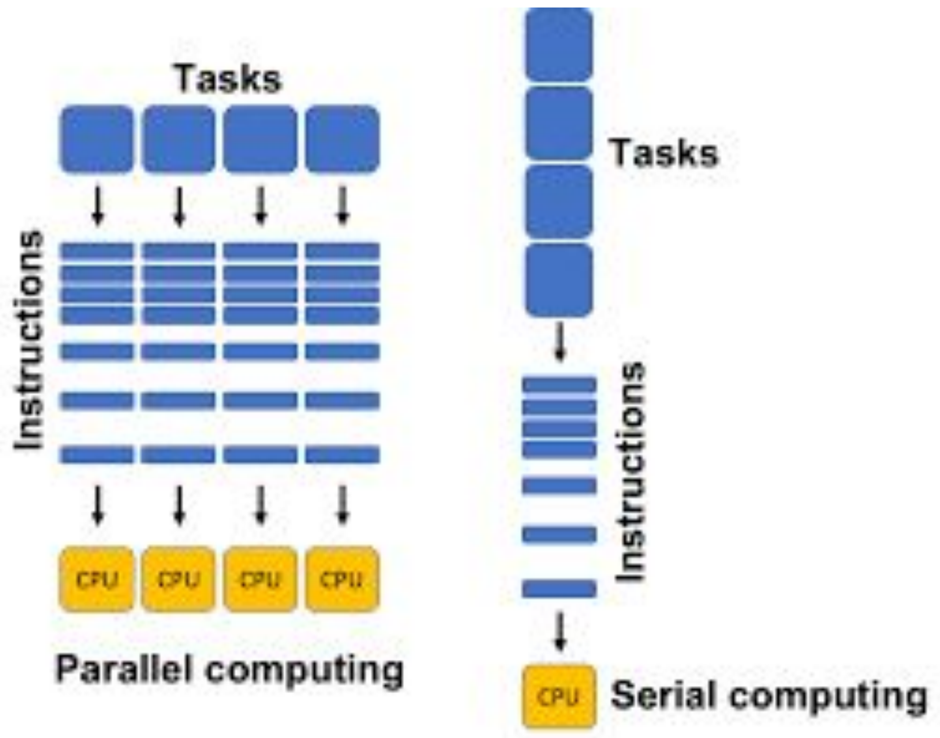
GREEDY ALGO

- Checking over all pairs of neighbours to cut-off maximum connection.
- Checking for maximum contribution from one branch.
- **$O((n\,log(n) + \Delta^2) \cdot C(stress\_centrality))$.**



Parallel Computing

- Brute Force approach using parallel architecture.
- Kernel for checking independent edge.
- **O(n· C(stress_centrality)).**



Further Development: An optimized kernel for stress centrality incorporating the dynamic addition of edges.

GNN Approach

- GraphSAGE to create **node embeddings.**
- Node prediction using reduction in node embeddings.
- **O(n· C(model_eval)).**

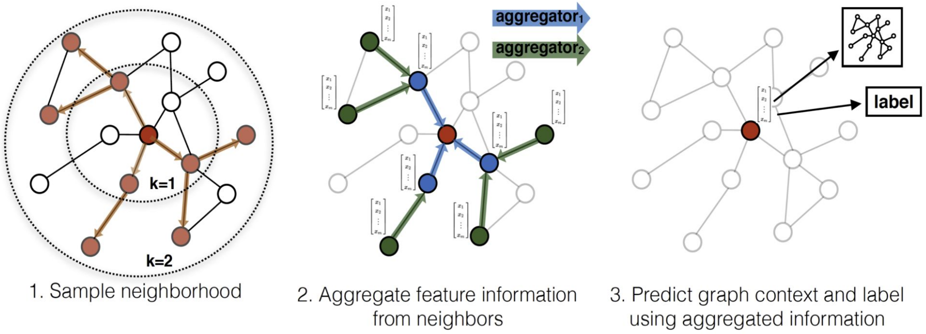
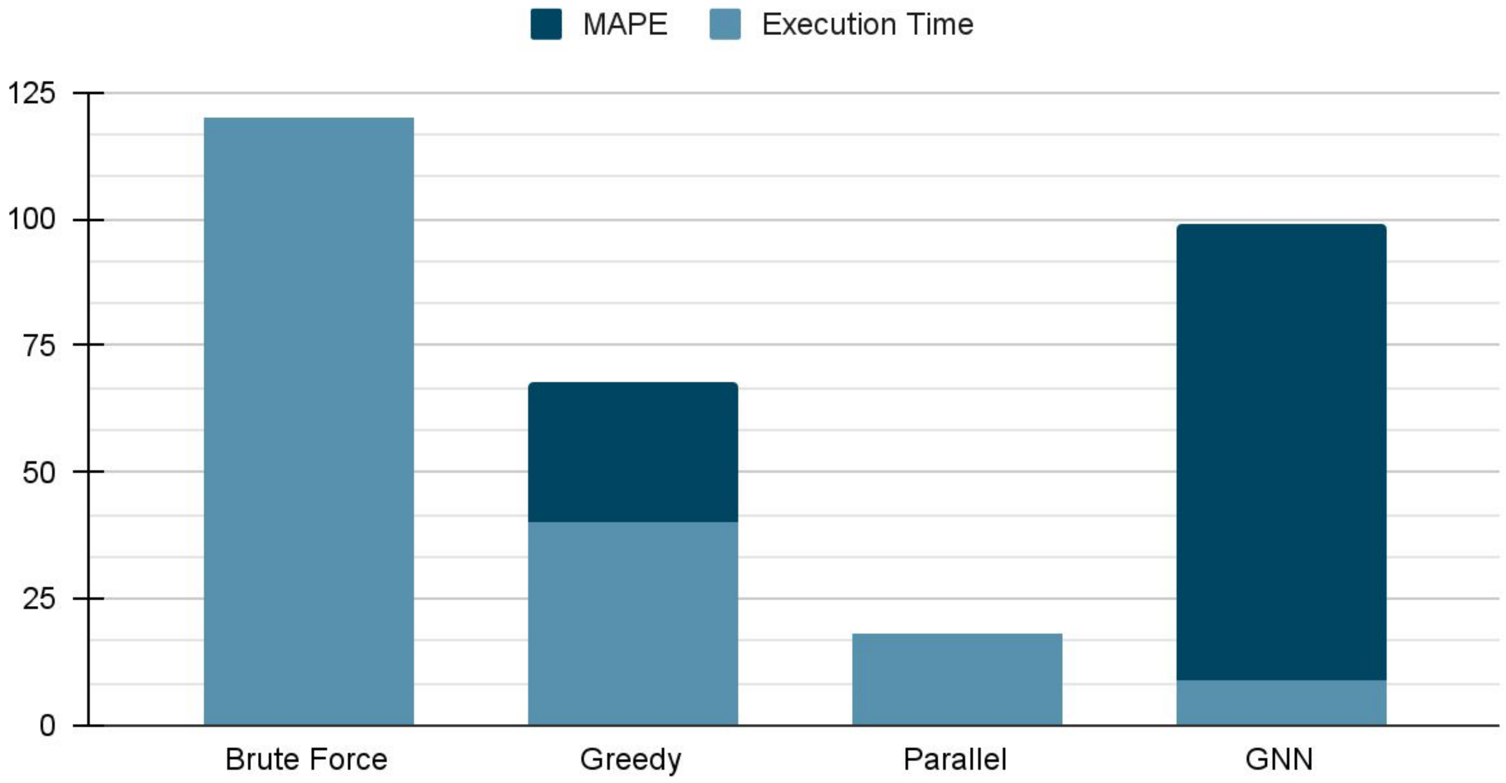


Figure 1: Visual illustration of the GraphSAGE sample and aggregate approach.

Further Development: **GNN + RL**

Performance



Thank You