



INDIAN INSTITUTE OF  
TECHNOLOGY  
BHILAI

# Social Network Generator using Jaccard Similarity

Satish Kumar Oraon Raunak  
Kumar  
Network Science

# Overview

---

- Introduction
- Statement of the Problem
- Review of Related Literature
- Framework
- Hypothesis
- Methodology
- Scope and Limitations
- Results
- Summary and Conclusion
- Implications and Recommendations
- References
- Whiteboard
- Brainstorm
- Q&A



---

# Introduction

- Social network analysis studies patterns of connections among individuals/organizations in a network.
- By analyzing these patterns of connections, we can gain insights into how information, resources, and influence flow through a network, which can be useful in a variety of fields

# Review of Related Literature

---

## Attribute Synthetic Generator

- The first phase of it involves network generation using preferential attachment
  - After the first phase, each node is assigned attributes
  - The attributes are then used to create links to the existing node by calculating the correlation
- 
- D. Chakrabarti, Y. Zhan, and C. Faloutsos. R-mat: A recursive model for graph mining. In SDM, volume 4, pages 442–446. SIAM, 2004
  - A.L.Barabási and R. Albert. Emergence of scaling in random networks. Science, 286(5439):509–512, 1999.

## Barabasi - Albert Model

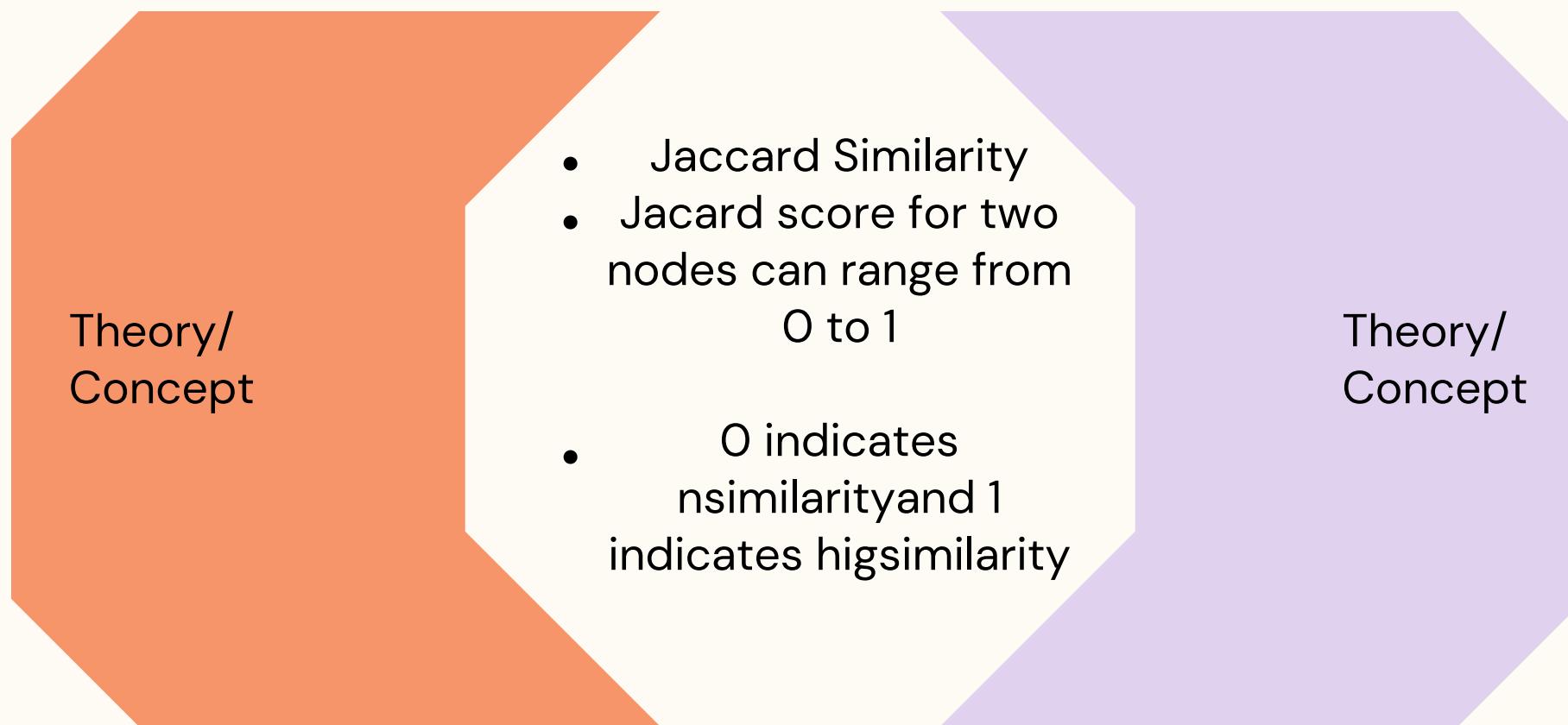
- Creates scale free networks
- Generates networks with power law degree distribution
- One disadvantage – does not consider clustering coefficient

## RMAT

- A recursive approach to create a network which follows a power law , small diameter
- The basic idea behind R-MAT is to recursively subdivide the adjacency matrix into four equal-sized partitions
- The edges are placed on this subgraph with unequal properties
- Displays community structure

# Framework

---



- Nodes not only tend to attach to popular nodes , but to nodes which are similar to them in some sense
- The idea is to attach preferential attachment with some similarity measure

Source: Add your references here.

# Hypothesis

---



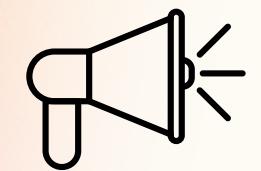
## More Diverse and realistic network

Considering attribute similarity creation of network that can better reflect the real world behaviour of individuals



**Nodes tend to create links with many similar nodes as they get old**

The new node has high entropy, but as they get many links, they tend to create links with more similar nodes .



## Friends of friend is my friend

If a node finds that a friend of his friend is similar to him, makes a link to him

# Methodology

Step 1

- Reading the interests of nodes from a file

Step 2

- Creating an initial social network with two nodes and adding edges between them.

Step 3

- For each additional node, computing the Jaccard similarity with all existing nodes to find candidate nodes that are potential neighbors.

Step 4

- Computing the probability of adding an edge to each candidate node, based on the degree of the candidate node and its neighbors.

Step 5

- Adding edges to the chosen candidate node until the desired number of nodes is reached.

Step 6

- Creating links with nodes at a distance 2

# Scope and Limitations

---

## Jaccard index limitation

- The Jaccard index is used as a similarity metric between nodes' interests. This may not always be the best choice for modeling social networks, especially when considering more complex relationships and dependencies among individuals.

## Lack of Node Attributes

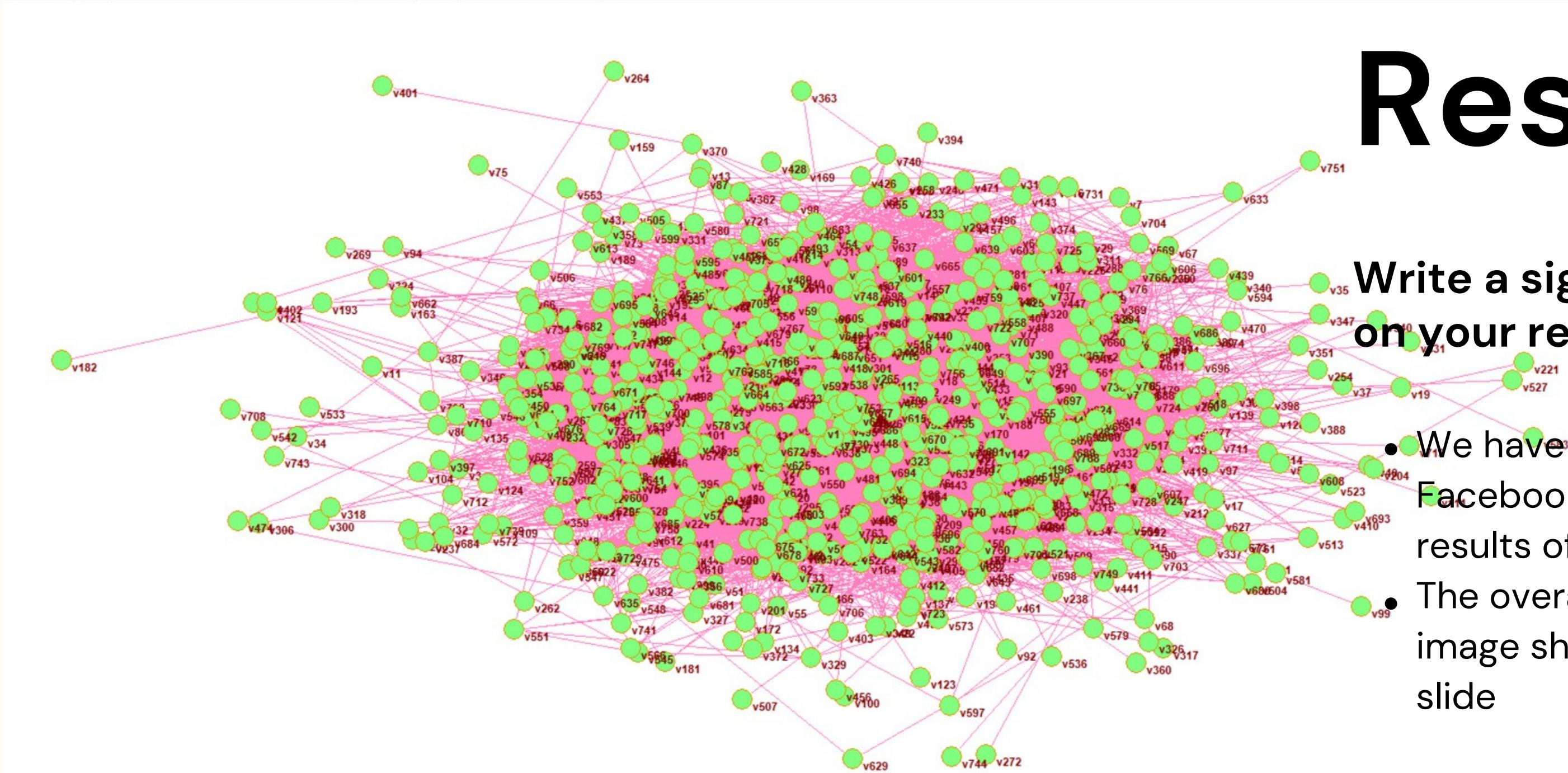
- In many cases, node attributes such as personal information or sensitive data are not available or cannot be easily obtained due to privacy concerns.
- can limit the types of analyses that can be performed on the network, and can also make it difficult to draw meaningful conclusions about the behavior of individual nodes or subgroups within the network

# Results

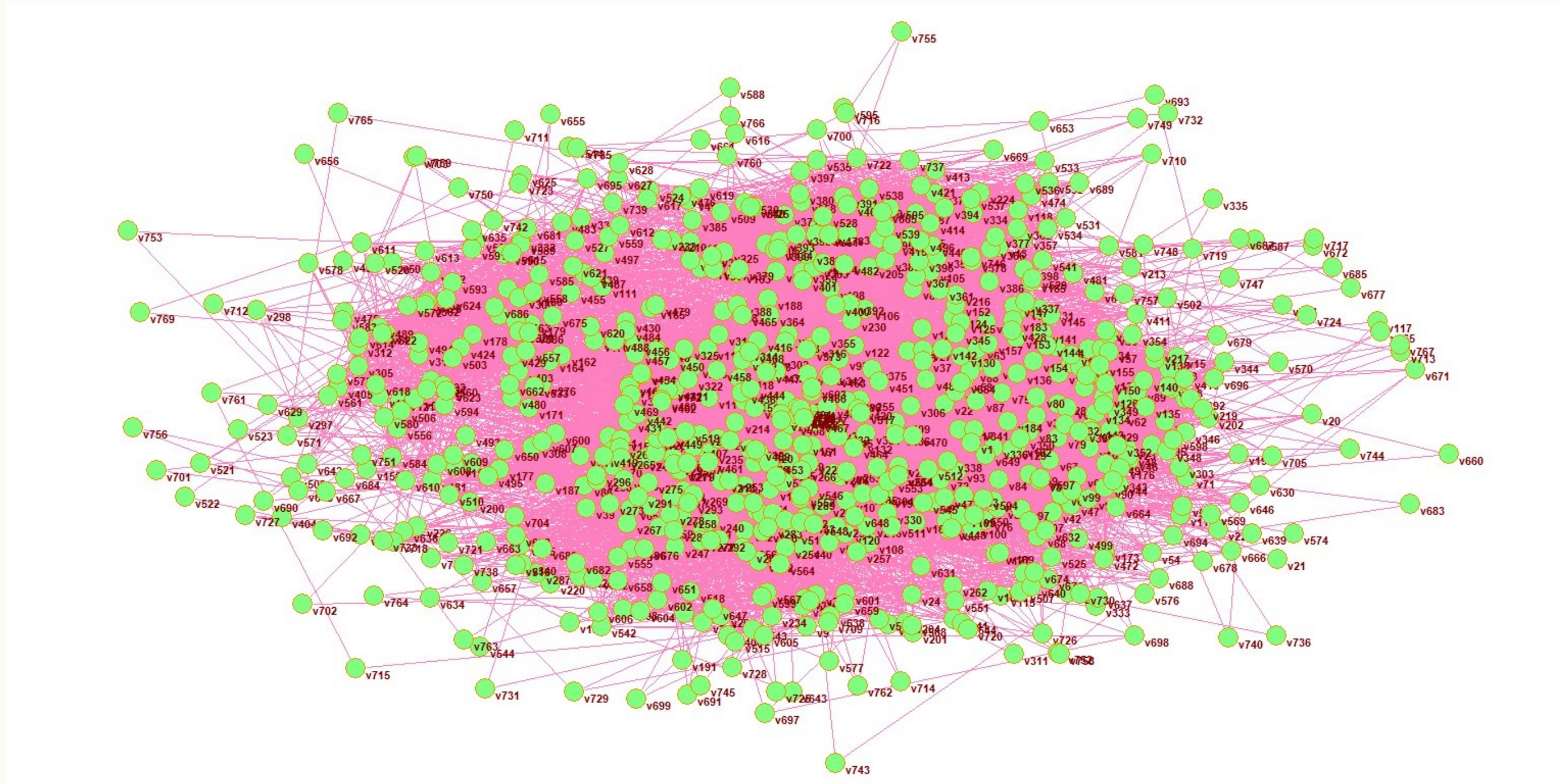
**Write a significant observation on your results here.**

- We have used intra college Facebook network to study the results of our network generator
- The overall network looks like the image shown in the left side of the slide

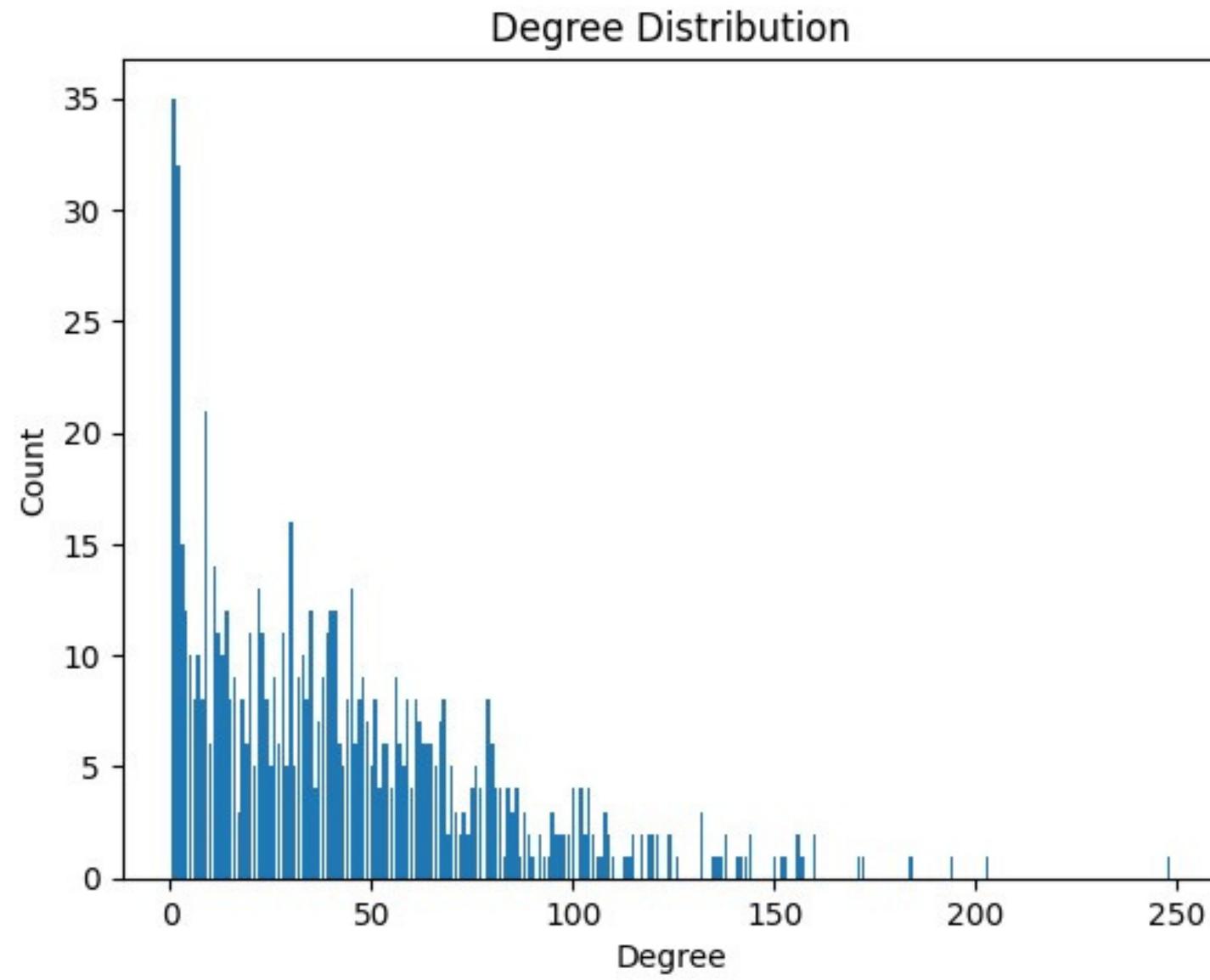
Layout GraphOnly Default Previous Redraw Next Options Export Spin Move Info FishEye Wait



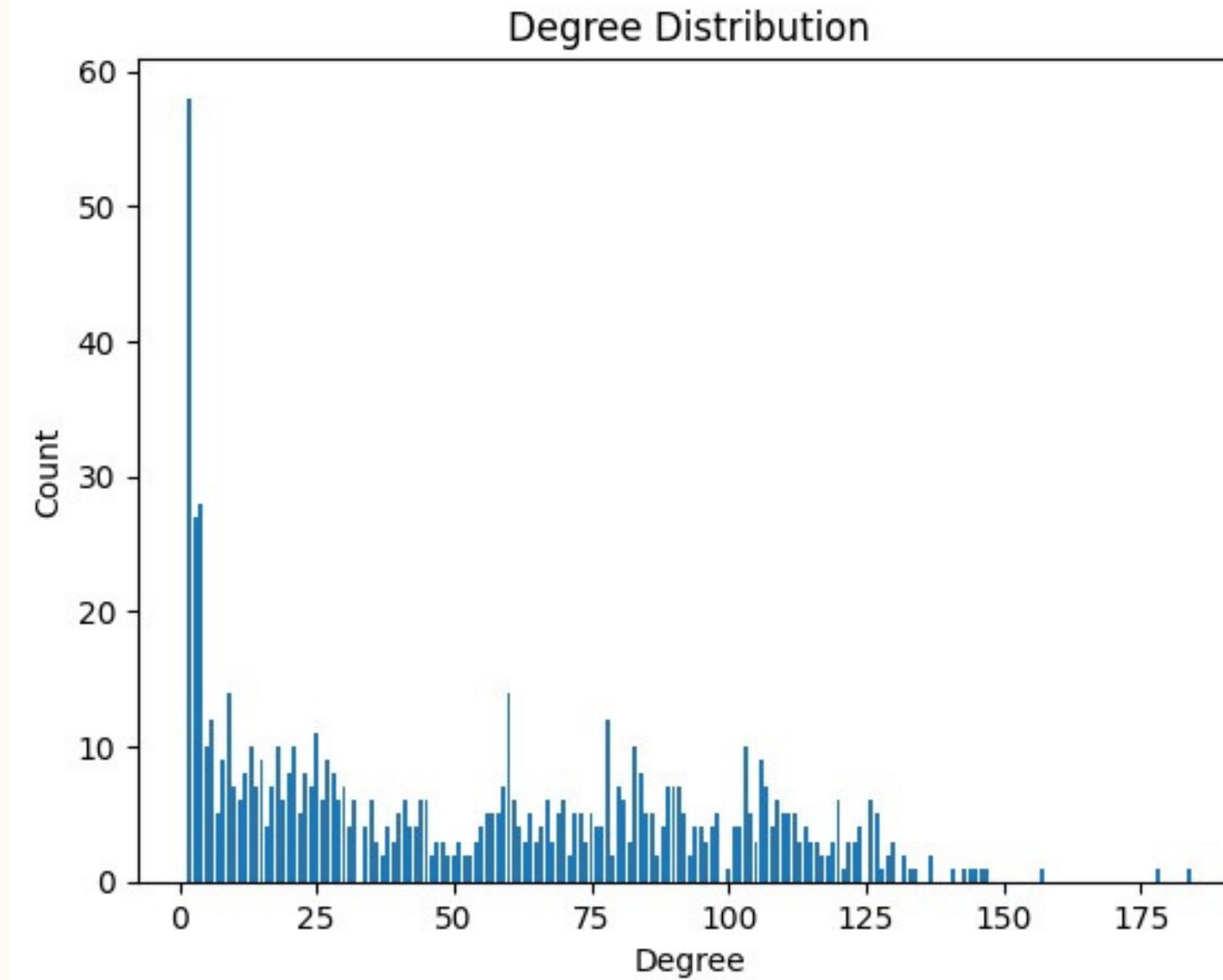
# The Network Generated from the attributes



# Caltech University



# Generated



# Implications and Recommendations

[Back to Overview](#)

## Theoretical and Methodological Issues

Heterophilicity in linking nodes refers to the tendency of nodes in a network to link with nodes that are dissimilar to themselves. In other words, nodes in a network with heterophilic linking patterns tend to form connections with nodes that have different attributes or characteristics than themselves.

## Practical Implications

Sociology Research: The generator can be used in sociology research to simulate social networks and study the impact of social connections and interests on various social phenomena, such as the spread of information or the diffusion of innovations.

Recommender Systems: Recommender systems can be trained on synthetic social networks generated by the generator to provide personalized recommendations to users based on their interests and social connections.

[Back to Overview](#)

# Q&A Session

Thank you for listening!