19ZO02 Social and Economic Network Analysis

Project Report(Batch – 6)

GOWTHAM T	**19Z214**
INDRA SHEKAR G	**19Z219**
SANJAI M	**19Z241**
YOGESH KUMAAR R	**19Z262**
MANOJ S	**19Z263**
MATHANA SEKARAN T	**19Z264**
HARISH N R	**20Z431**

Dissertation submitted in partial fulfilment of the requirements for the degree of

BACHELOR OF ENGINEERING

Branch: COMPUTER SCIENCE AND ENGINEERING

NOVEMBER 2022

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

PSG COLLEGE OF TECHNOLOGY

(Autonomous Institution)

COIMBATORE - 641 004

1. Problem Statement:

To visualize the structure of Marvel Comic Universe as a social network and perform,

- Link Prediction
- Medium of interaction between the Heroes
- Detect various communities in the Marvel Comic Universe

2. Dataset description:

The dataset consists of three types of csv(Comma Separated values) files

- Edges – marvel characters and the comic in which they have appeared

	hero	comic
0	24-HOUR MAN/EMMANUEL	AA2 35
1	3-D MAN/CHARLES CHAN	AVF 4
2	3-D MAN/CHARLES CHAN	AVF 5
3	3-D MAN/CHARLES CHAN	COC 1
4	3-D MAN/CHARLES CHAN	H2 251

- Nodes – tells whether a entity is a comic or character

	node	type
0	2001 10	comic
1	2001 8	comic
2	2001 9	comic
3	24-HOUR MAN/EMMANUEL	hero
4	3-D MAN/CHARLES CHAN	hero

- Hero network – two characters are connected if they appear in the same comic

	hero1	hero2
0	LITTLE, ABNER	PRINCESS ZANDA
1	LITTLE, ABNER	BLACK PANTHER/T'CHAL
2	BLACK PANTHER/T'CHAL	PRINCESS ZANDA
3	LITTLE, ABNER	PRINCESS ZANDA
4	LITTLE, ABNER	BLACK PANTHER/T'CHAL

The overall dataset contains 6423 nodes i.e., heroes,12651 comics and the Hero Network dataset has 574467 edges, Edges dataset has 96104 edges and Nodes dataset has 19090 edges.

3. Tools Used

Gephi: it is an open-source software for visualization of networks and graphs

Matplotlib: Python visualization library used to plot graphs

Pandas: Python library for loading, viewing, exploring, and manipulating datasets.

Networkx: Python library to create, analyse and manipulate network and graphs.

Sklearn: Machine learning package in python

4. Challenges:

The main challenge that we faced is the size of dataset. In link prediction first we will be removing some edges between the nodes, as out dataset has many entities it became time consuming and computationally intensive task it took nearly 160 hours to run that part. We cannot simply remove some entities from our dataset as each row has a specific role to play in the graph. Also, Networkx is a new package for us so initially it took some time to learn about it. We also faced some minor issues in downloading community package, but we resolved it.

5. Contribution:

Roll No	Name	Contribution
19Z214	Gowtham T	Medium of interaction, Gephi visualizations
19Z219	Indra Shekar G	Gephi visualizations, Community Detection
19Z241	Sanjai M	Gephi visualizations, Medium of interaction
19Z262	Yogesh Kumaar R	Link prediction
19Z263	Manoj S	Link Prediction
19Z264	Mathana Sekaran T	Community Detection, Gephi visualizations
20z431	Harish N R	Final Report, community Detection

6. Annexure I: code

Git repo Link to the project

7. Annexure II: Snapshots

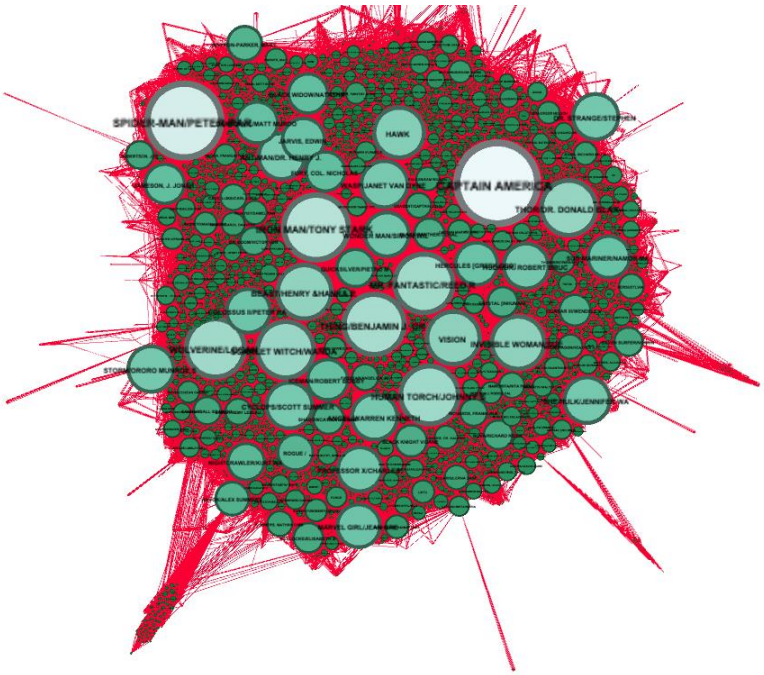

Fig 1. Network visualization using gephi

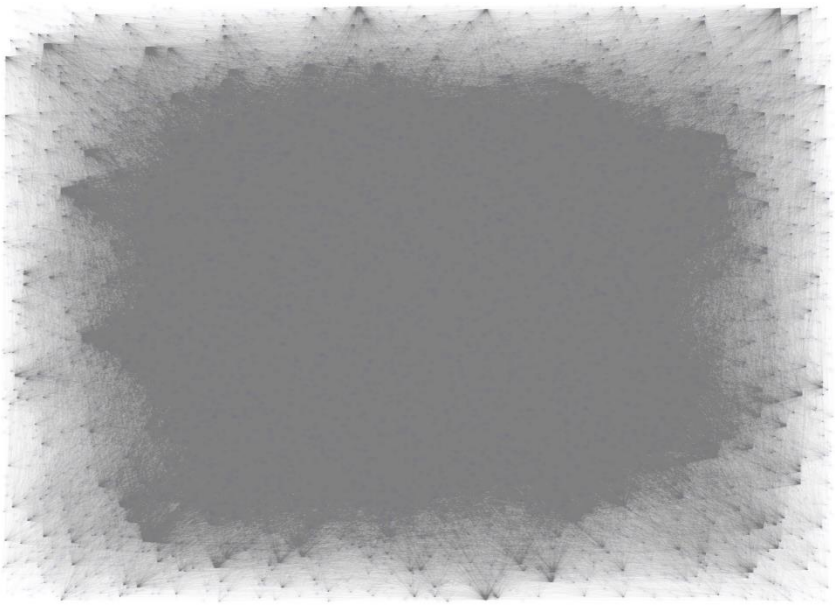

Fig 2. Whole network

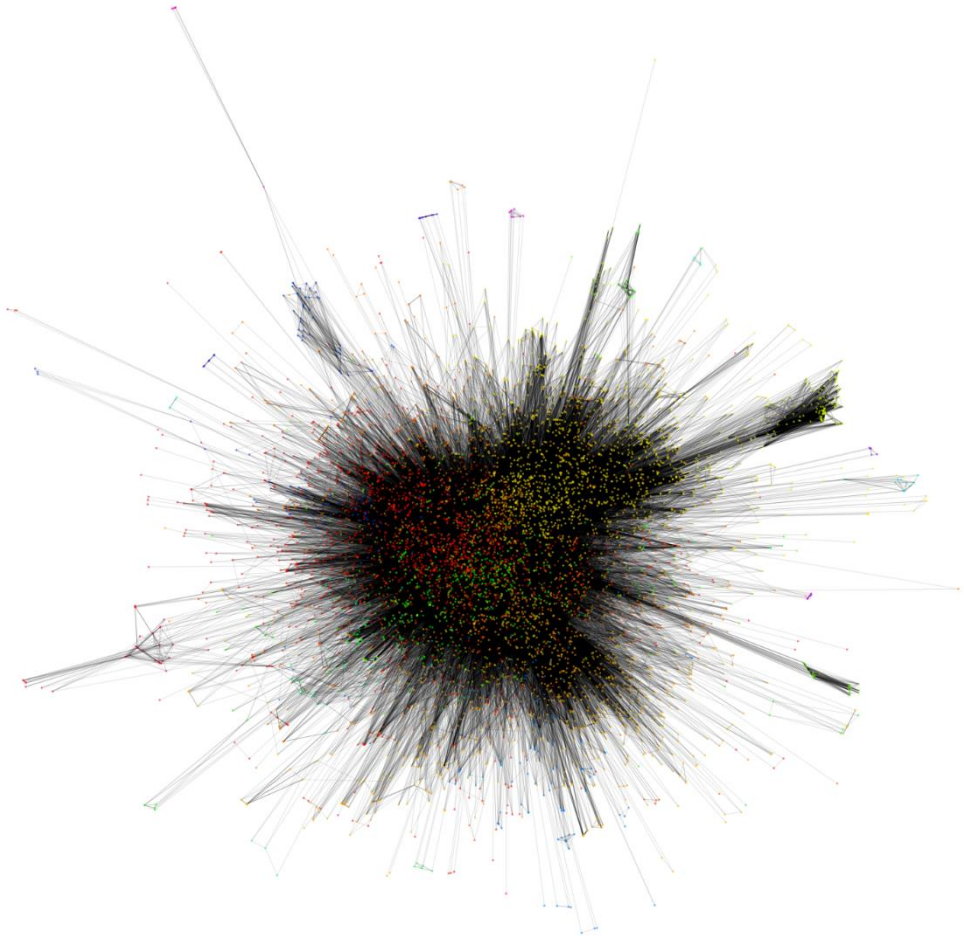

Fig 3. Communities in the network

Plagiarism report:

VeriGuide - Originality Report
Individual Report

Background Information

File Name:	SENA_report.docx
Report Generated On:	11/11/2022, 11:22:01 AM

Similarity Statistics Overview

Similar Sentence(s) Found By VeriGuide:	1 out of 38 sentences = 2.63%
Similar Sentence(s) Filtered by User:	1 out of 38 sentences = 2.63%
Sentence(s) Selected By User To Export:	0

Similarity Statistics for Each Source

Entry	Source	From	Similarity
1	https://gephi.org/	Internet	1 / 38 = 2.63%