NAME: G.SAI SUSHANTH REDDY

REG.NO: 22MCB0005

SUBJECT: SOCIAL NETWORK ANALYTICS (MCSE618P)

ASSESSMENT-3

(REPORT)

VISUALIZING MULTI-DIMENSIONAL HETEROGENEOUS SOCIAL NETWORKS, CLUSTERING & CLASSIFICATION

INTRODUCTION

Visualizing multi-dimensional heterogenous social networks can be challenging due to the complex nature of the data. In this context, we explore the visualization of a dataset based on Marvel fictional characters, encompassing diverse attributes such as character traits, powers, affiliations, and relationships. By representing this information visually, we can uncover hidden patterns, clusters, and connections, enabling a deeper understanding of the social fabric within the fictional hero universe.

This visualization endeavour involves addressing the challenges posed by the heterogeneity of the dataset, where heroes possess various data types, including textual descriptions, categorical labels, and numerical attributes. By employing innovative visualization techniques, such as node-link diagrams, force-directed layouts, and color-coded encoding, we can map the multidimensional hero attributes onto visually intuitive representations. Through interactive exploration and analysis, we aim to unveil the intricate web of relationships, communities, and characteristics that define these fictional hero networks.

METHODOLOGY & RESULT

The following marvel fictional heroes dataset contains various columns like Page ID, Name, Identity, Character, Eye & Hair Colour, Alive, Number of appearances, First Appearance & Year.

	page_id	name	ID	ALIGN	EYE	HAIR
0	1678	Spider-Man (Peter Parker)	Secret Identity	Good Characters	Hazel Eyes	Brown Hair
1	7139	Captain America (Steven Rogers)	Public Identity	Good Characters	Blue Eyes	White Hair
2	64786	Wolverine (James \"Logan\" Howlett)	Public Identity	Neutral Characters	Blue Eyes	Black Hair
3	1868	Iron Man (Anthony \"Tony\" Stark)	Public Identity	Good Characters	Blue Eyes	Black Hair
4	2460	Thor (Thor Odinson)	No Dual Identity	Good Characters	Blue Eyes	Blond Hair
5	2458	Benjamin Grimm (Earth-616)	Public Identity	Good Characters	Blue Eyes	No Hair
6	2166	Reed Richards (Earth-616)	Public Identity	Good Characters	Brown Eyes	Brown Hair
7	1833	Hulk (Robert Bruce Banner)	Public Identity	Good Characters	Brown Eyes	Brown Hair
8	29481	Scott Summers (Earth-616)	Public Identity	Neutral Characters	Brown Eyes	Brown Hair
9	1837	Jonathan Storm (Earth-616)	Public Identity	Good Characters	Blue Eyes	Blond Hair

| SEX	ALIVE	APPEARANCES	FIRST APPEARANCE	Year	First Name
Male Characters	Living Characters	4043	Aug-62	1962.0	Spider
Male Characters	Living Characters	3360	Mar-41	1941.0	Captain
Male Characters	Living Characters	3061	Oct-74	1974.0	Wolverine
Male Characters	Living Characters	2961	Mar-63	1963.0	Iron
Male Characters	Living Characters	2258	Nov-50	1950.0	Thor
Male Characters	Living Characters	2255	Nov-61	1961.0	Benjamin
Male Characters	Living Characters	2072	Nov-61	1961.0	Reed
Male Characters	Living Characters	2017	May-62	1962.0	Hulk
Male Characters	Living Characters	1955	Sep-63	1963.0	Scott
Male Characters	Living Characters	1934	Nov-61	1961.0	Jonathan
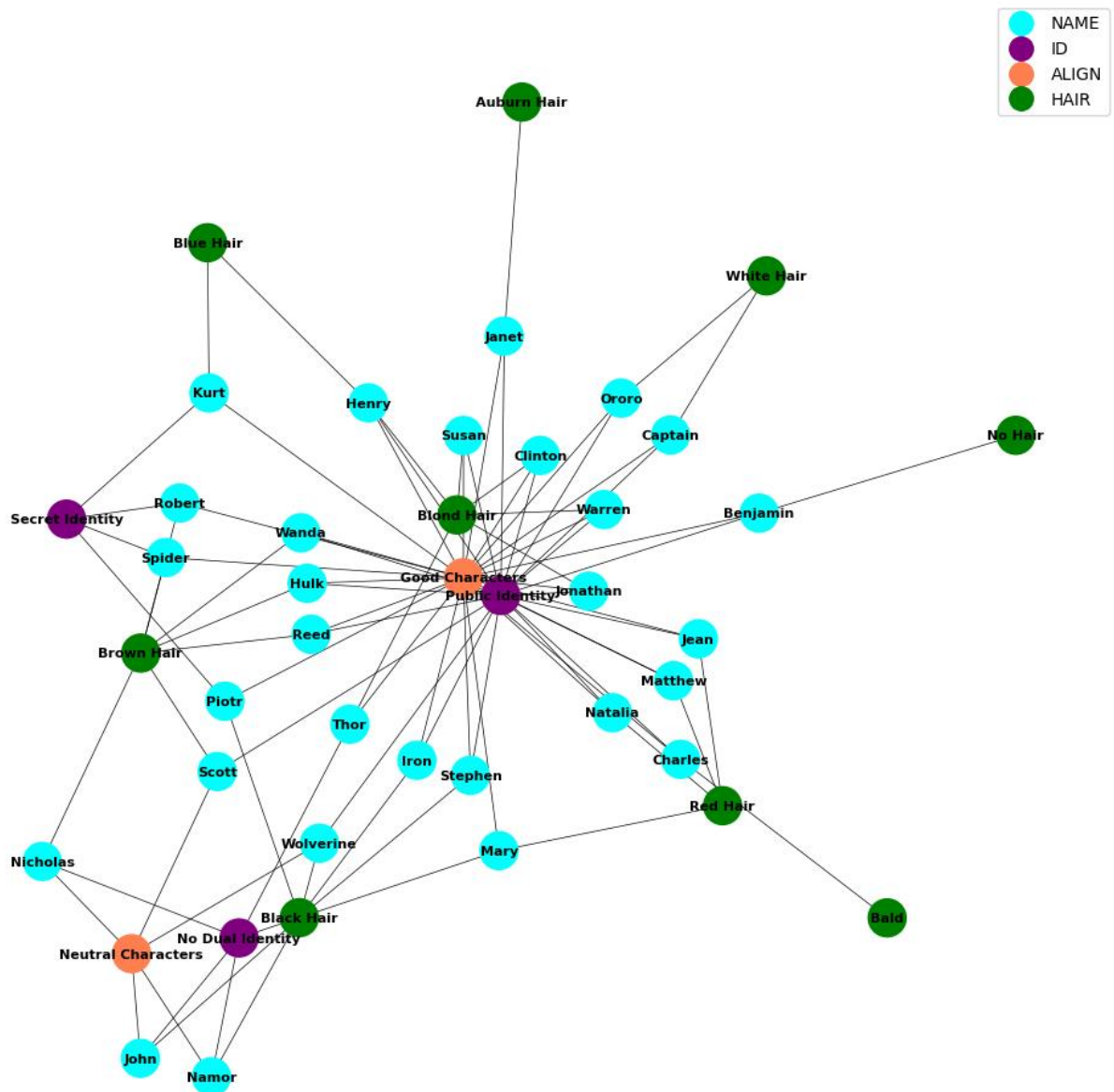
We have taken following features like NAME, ID, ALIGN (character), HAIR and have tried to visualize them graphically represent them and identify and unveil the intricate web of relationships among the columns.

And through the graph we are able to identify certain patterns which cannot be identified in a tabular manner.

Clustering and Classification:

We have used **K-means** clustering to identify various groups or clusters based on Name, Identity and Hair Colour.
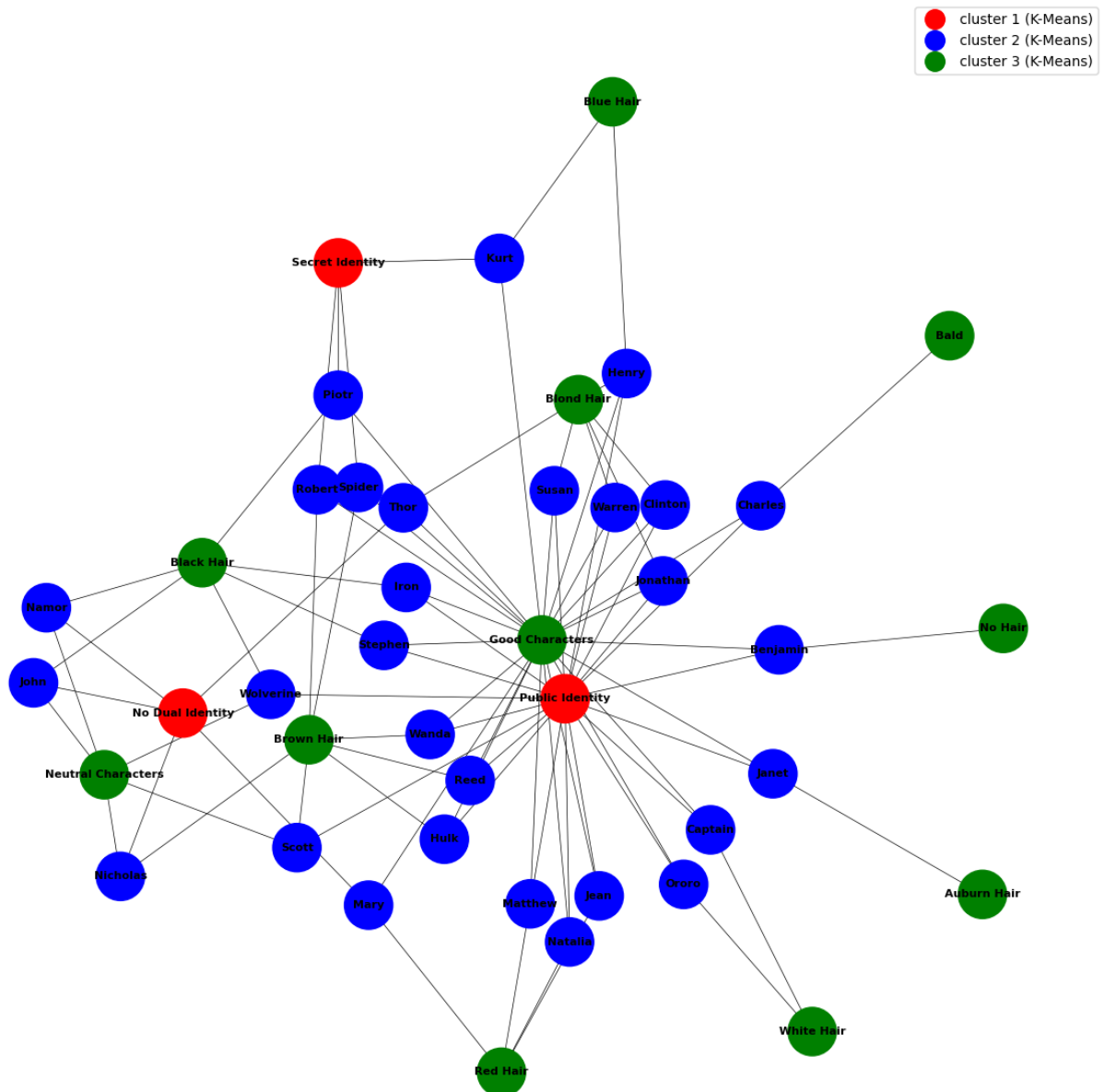
Once clusters are identified, you can perform classification within each cluster to predict specific outcomes or labels. This involves training classification models on labelled data within each cluster. So I have used Decision Tree Classifier

I have used appropriate metrics to evaluate the performance of your clustering and classification models such as Accuracy, Precision, Recall, F1-score and constructed a confusion matrix.

```
Accuracy: 0.6666666666666666
Precision: 0.4444444444444444
Recall: 0.6666666666666666
F1-score: 0.5333333333333334
Confusion Matrix:
 [[6 0]
 [3 0]]
```

CONCLUSION

In conclusion, visualizing multi-dimensional heterogeneous social networks, exemplified by our exploration of a Marvel fictional characters dataset, providing a powerful means of unravelling the complexities inherent in such intricate systems.

Through the utilization of diverse visualization techniques and approaches tailored to handle the heterogeneity of the dataset, we have gained valuable insights into the relationships, clusters, and patterns that shape the fictional hero universe.

These visualizations have not only deepened our understanding of the dataset but have also demonstrated the potential for visual exploration to enhance comprehension and foster new discoveries in multi-dimensional social networks.

Source Code:
https://github.com/SushRed10/22MCB0005/blob/main/22MCB0005_SNA_Assessment3.ipynb