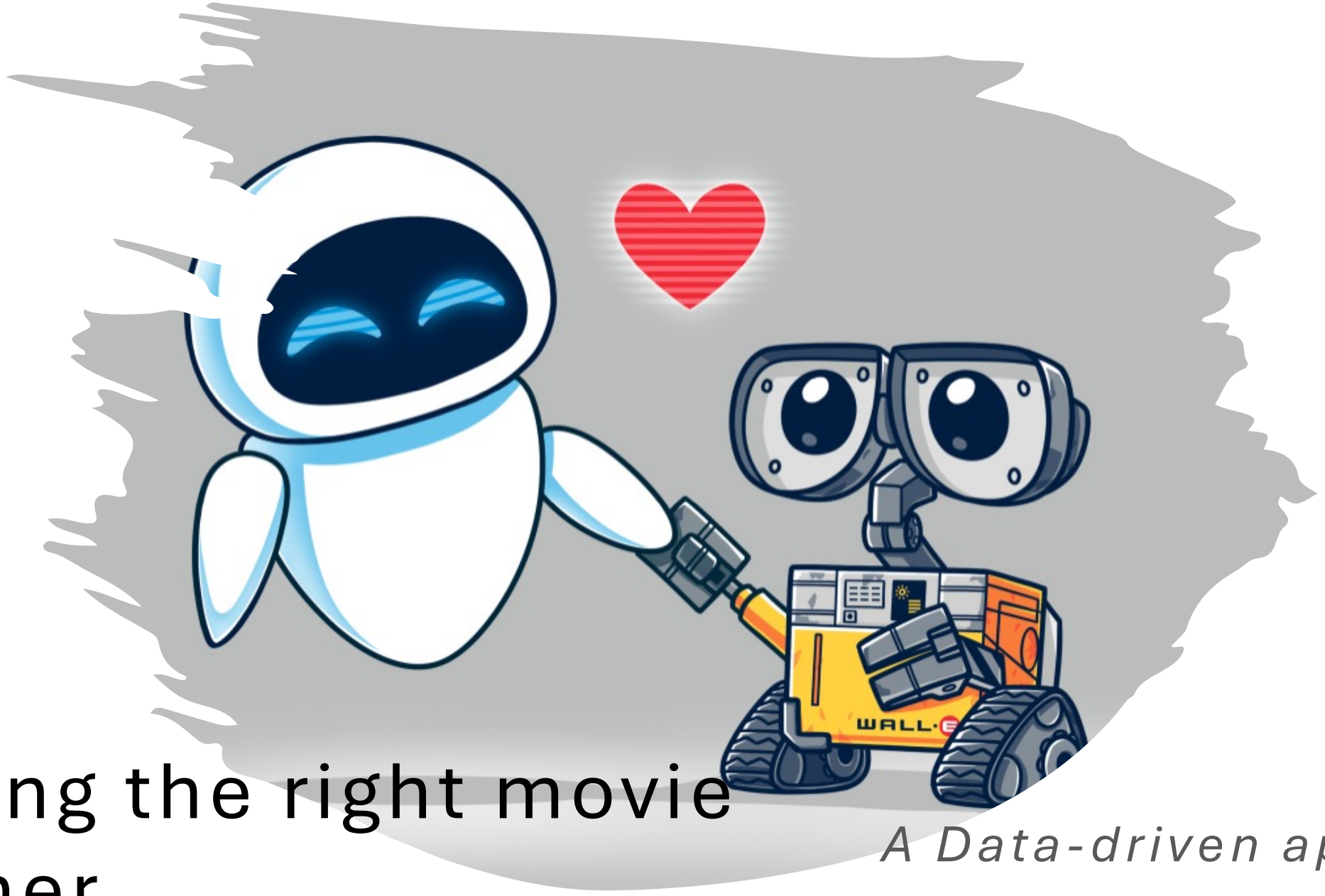
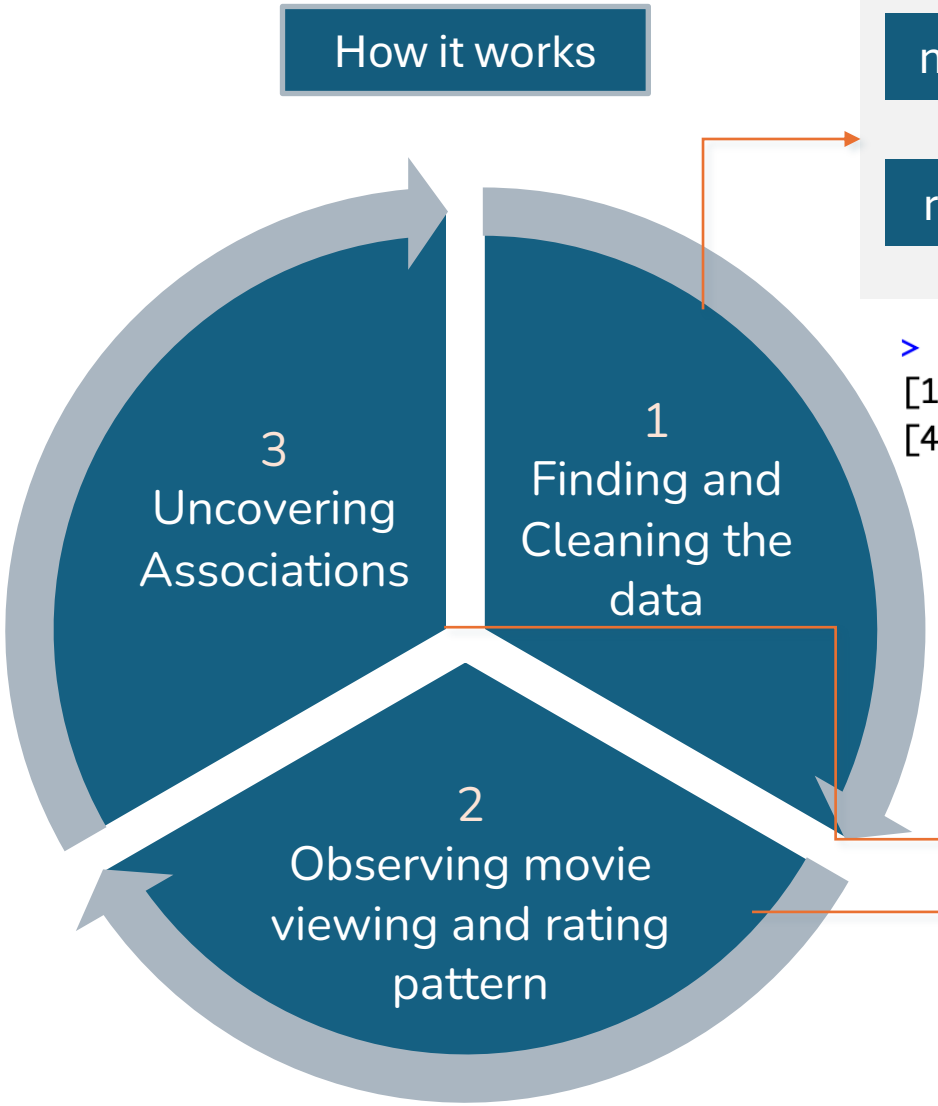


Finding the right movie partner

A Data-driven approach

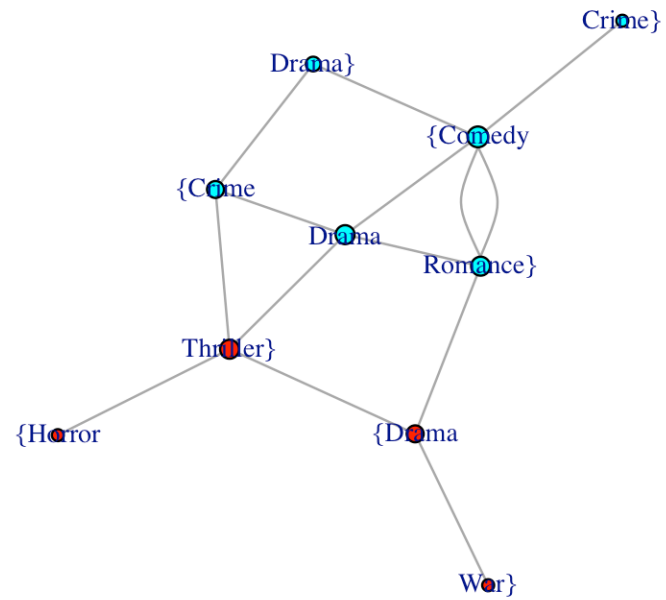
How it works



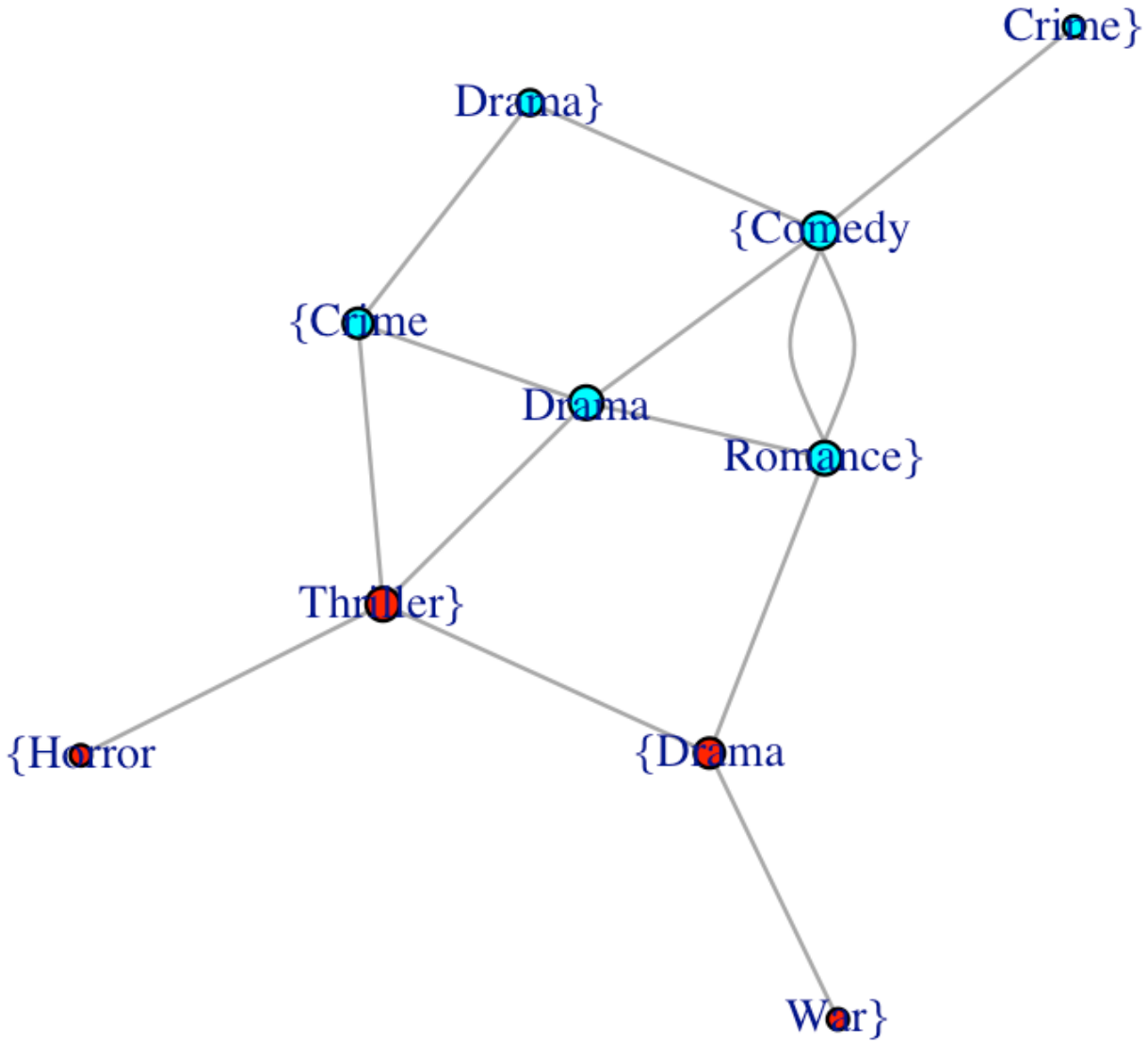
movies.csv	movieId	title	genres

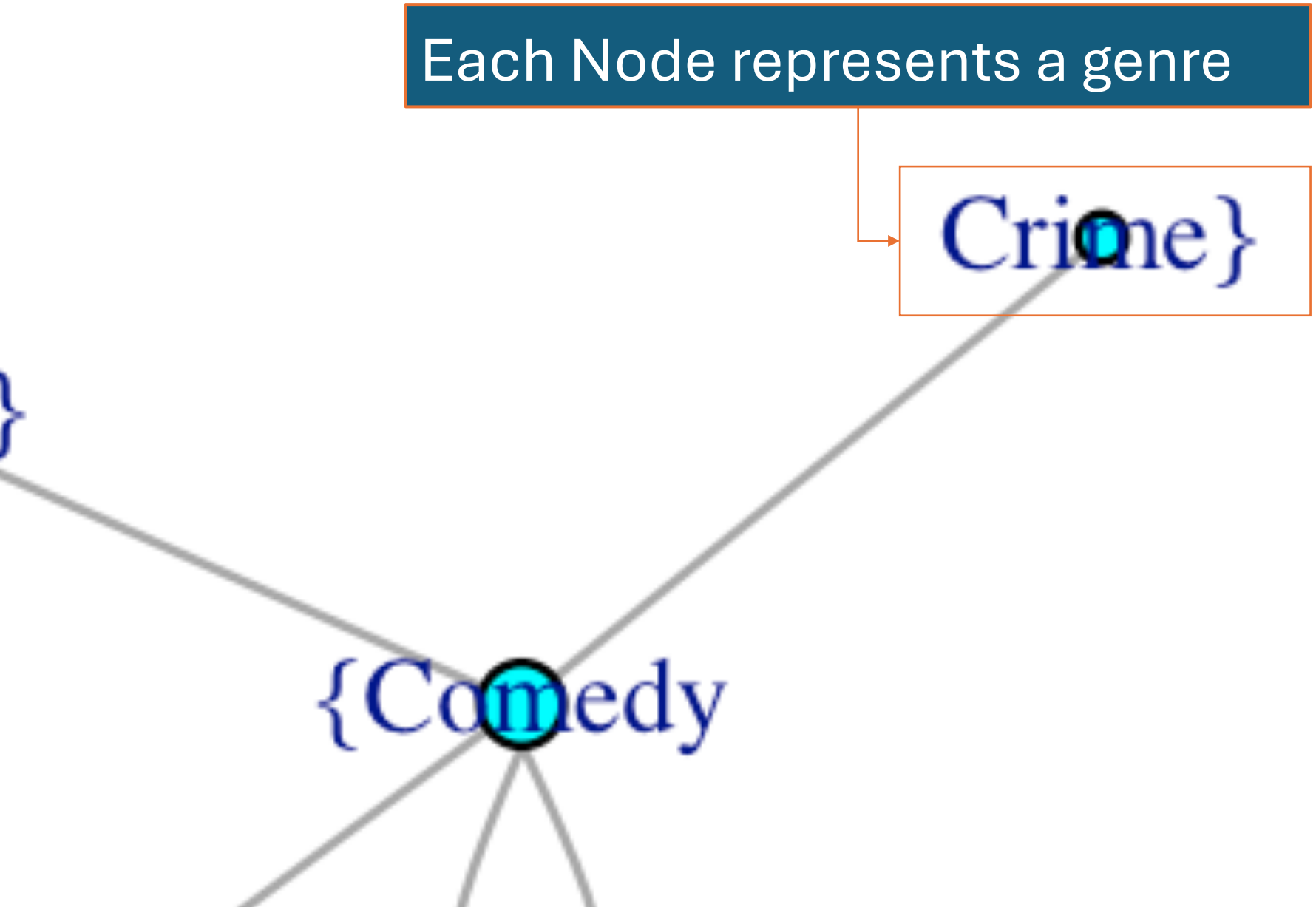
ratings.csv	userId	movieId	rating	timestamp

```
> head(genre_combinations)
[1] "{Comedy,Crime}"          "{Drama,War}"          "{Crime,Drama,Thriller}"
[4] "{Crime,Drama}"           "{Horror,Thriller}"    "{Horror}"
```

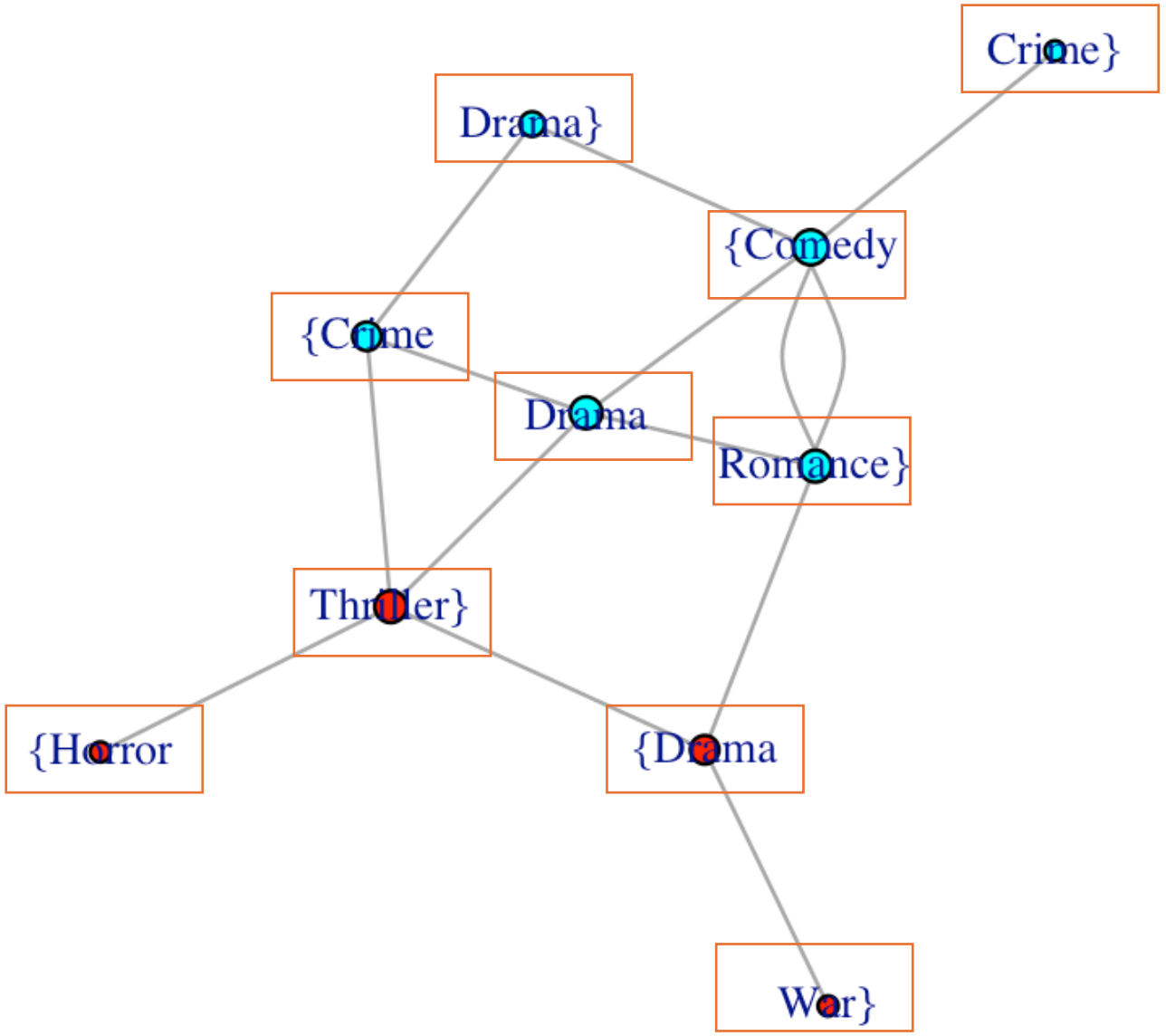


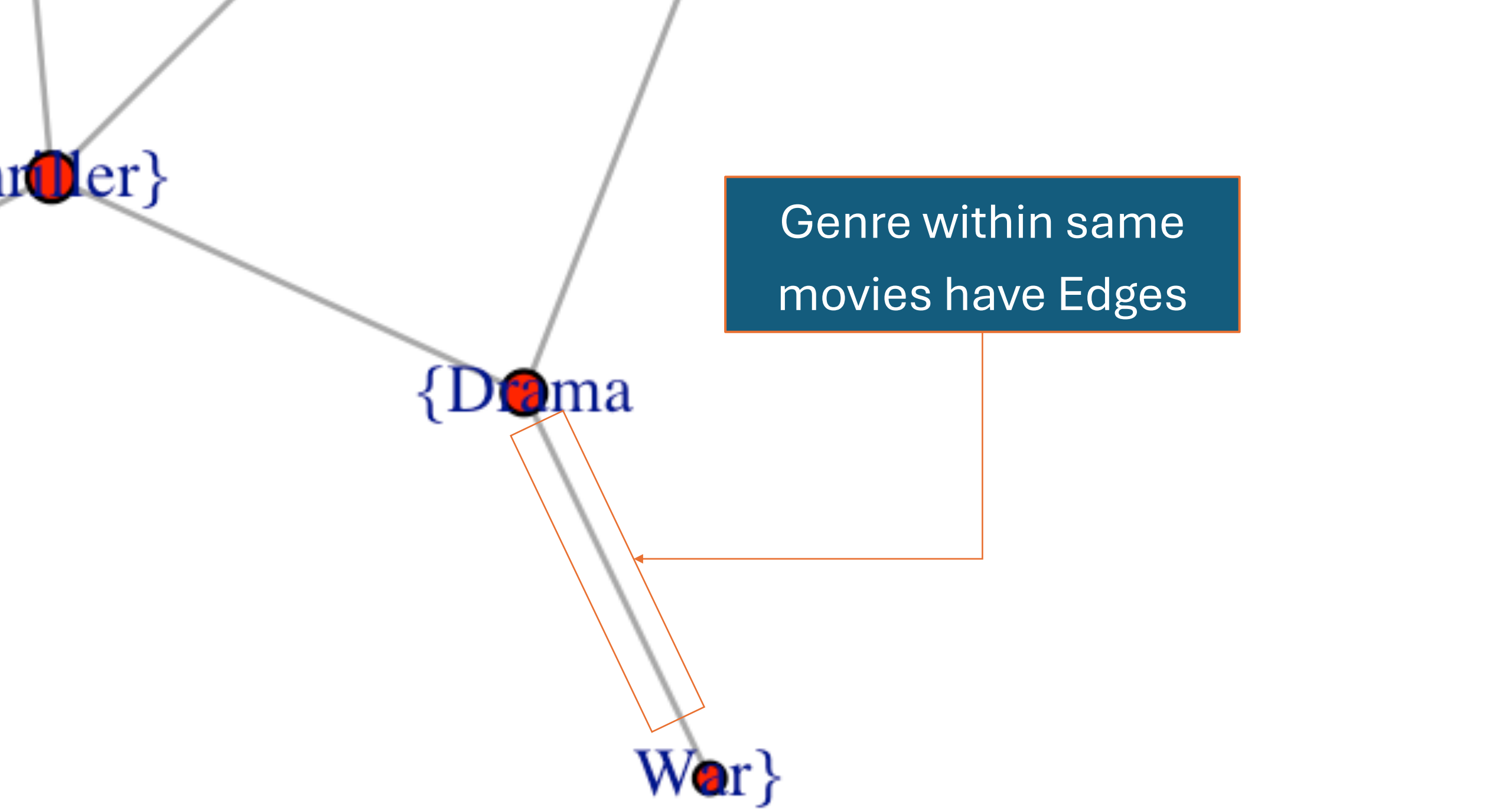
Interpreting the result of Enhanced Network of Genre Combinations



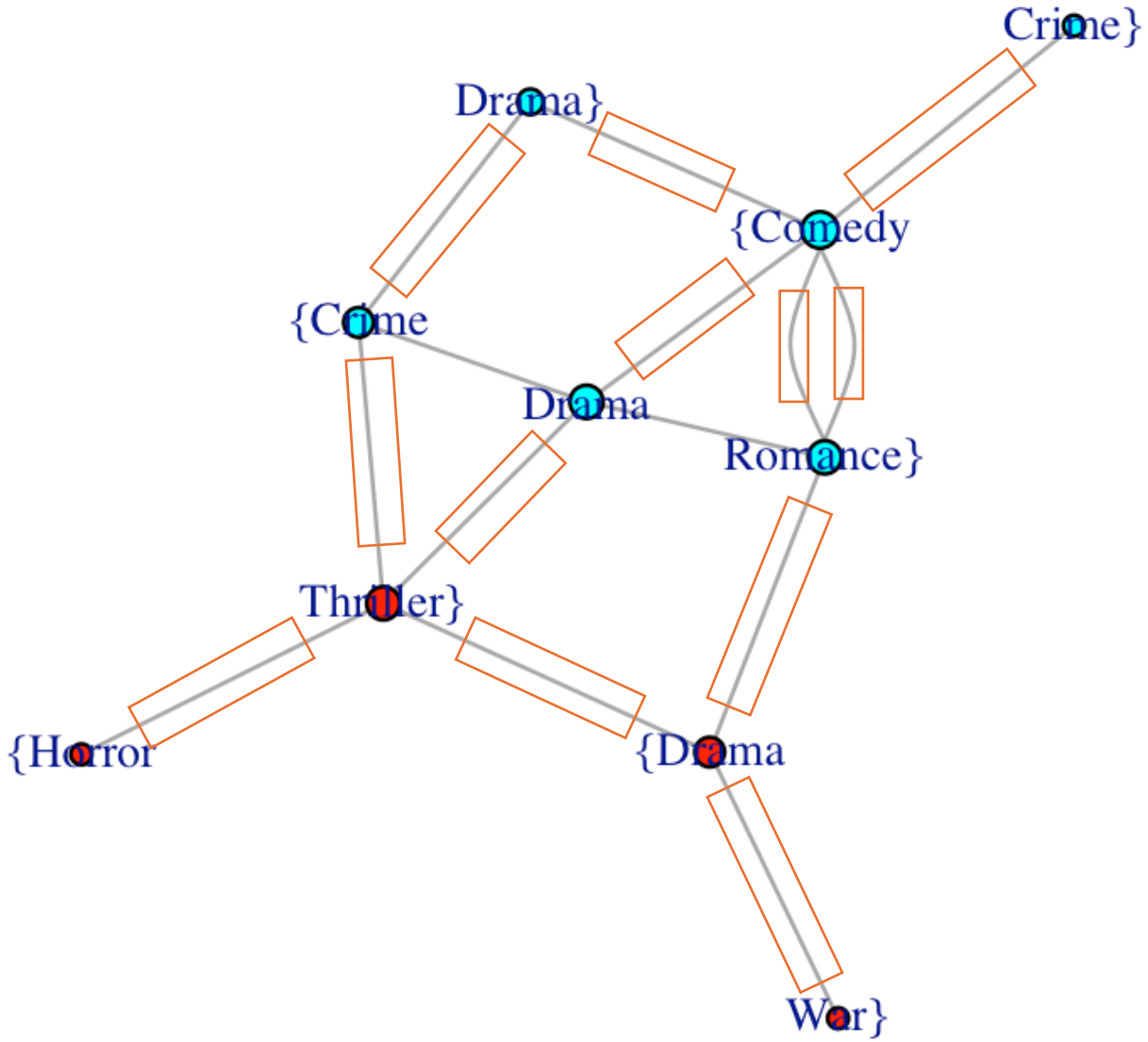

Each Node represents a genre
Crime}
{Comedy

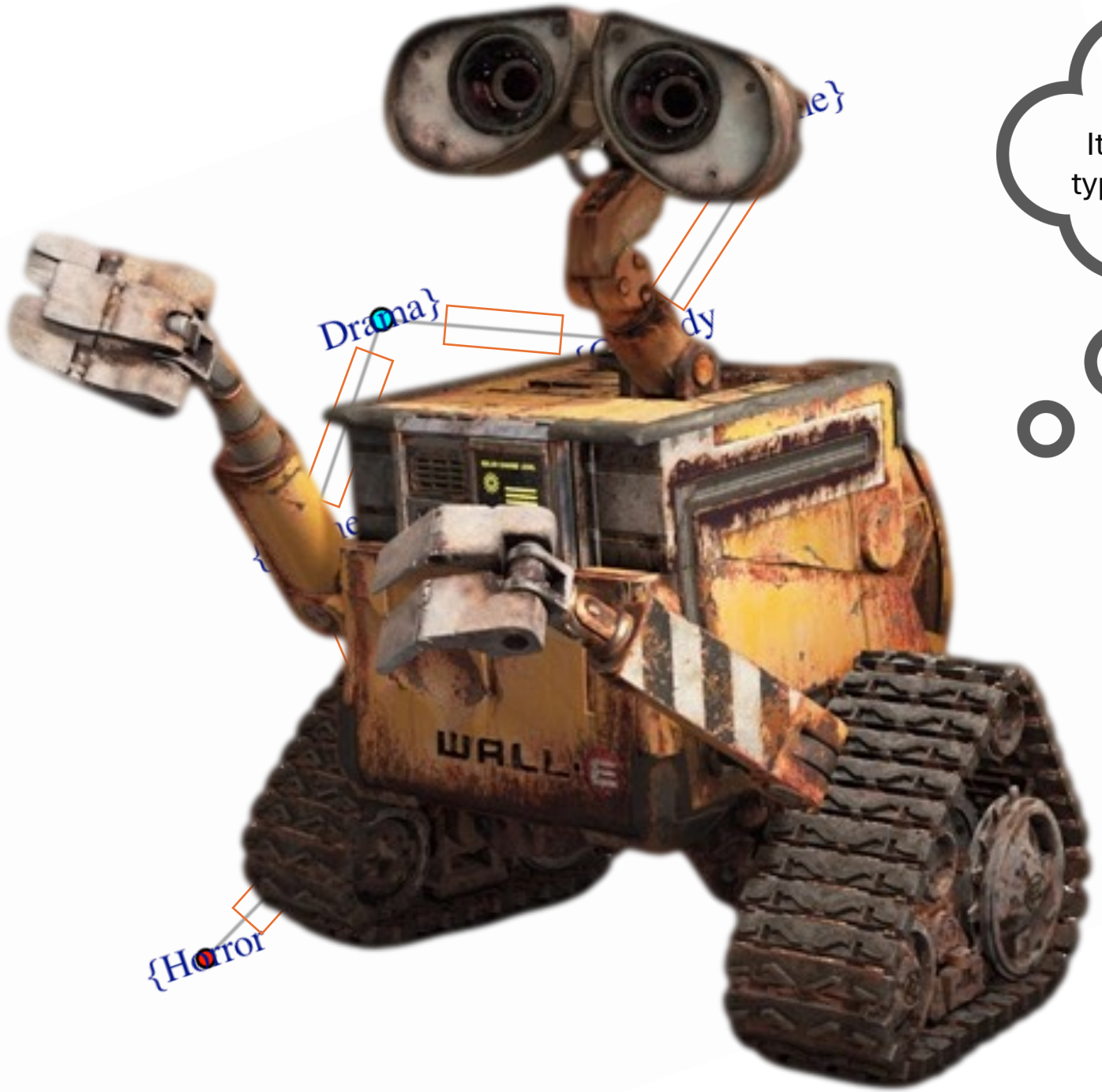
Nodes are Movie Genre




ller}
{Drama
War}
Genre within same movies have Edges

Edges are genre within same movies

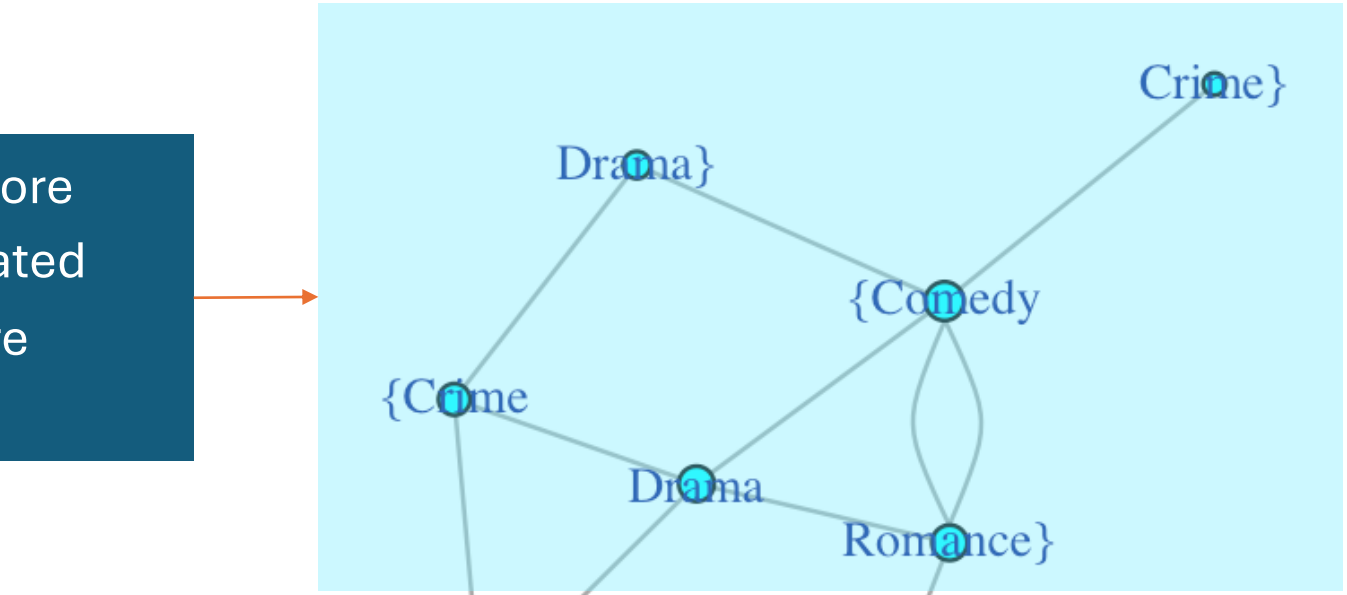


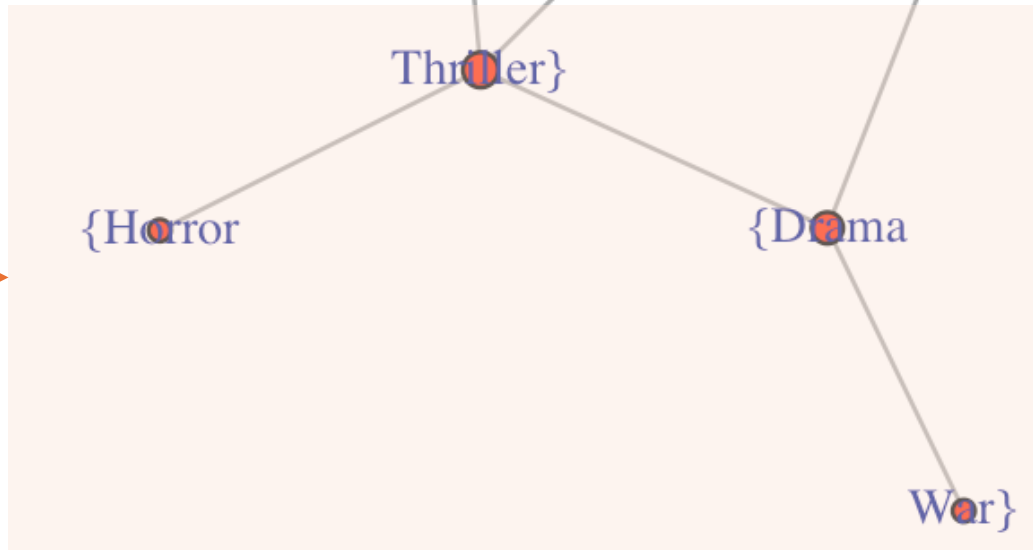


It's like all these movie types are holding hands. Eva...

Detecting communities within genre

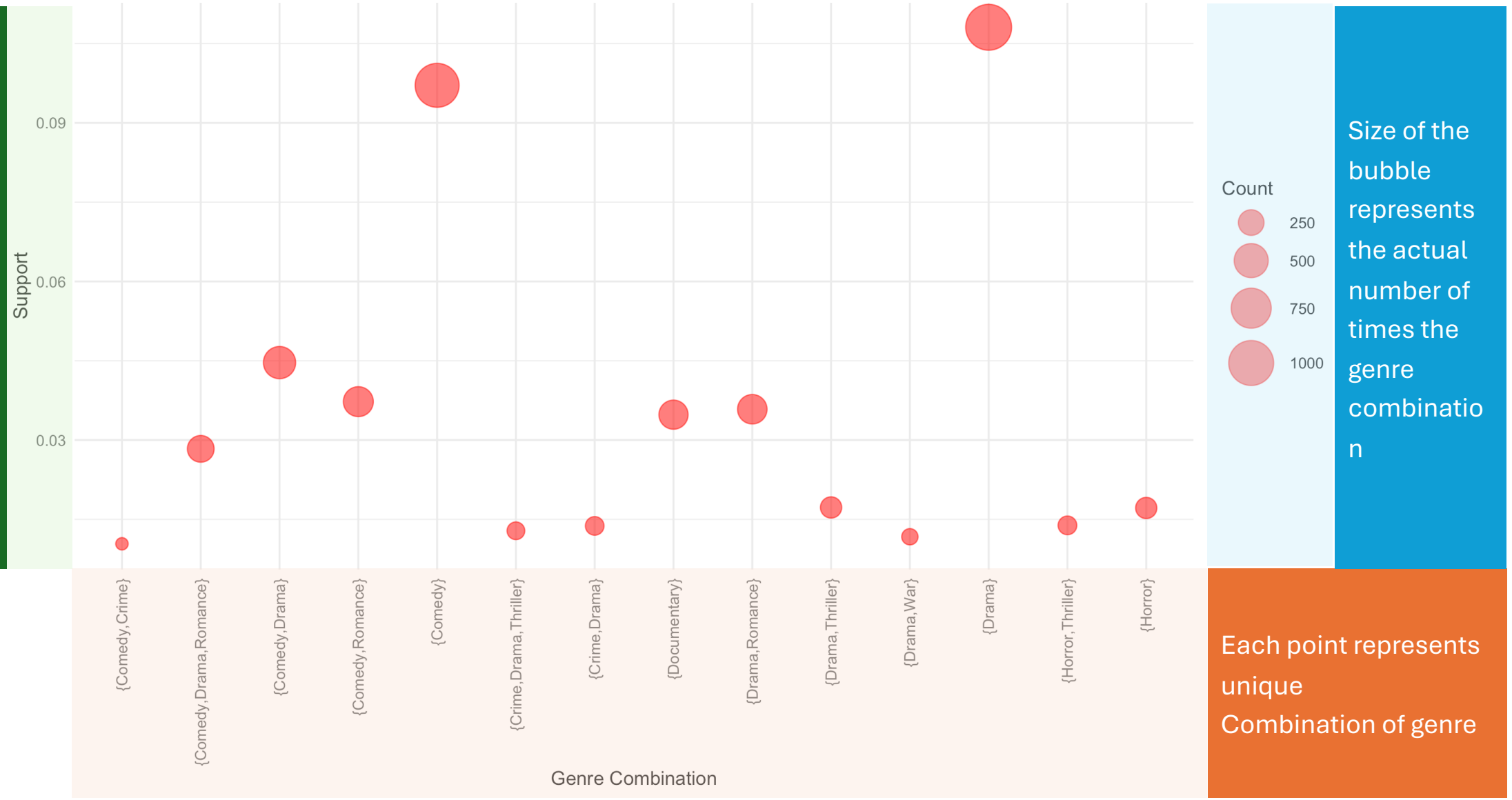
Genres that are more frequently associated with each other are colored similarly





Bubble Plot of Frequent Genre Combination

In simple terms, like the popularity or frequency of each genre combination within the dataset

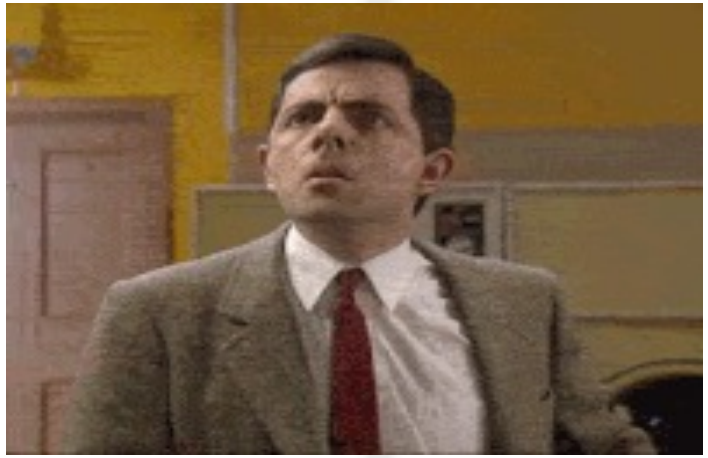


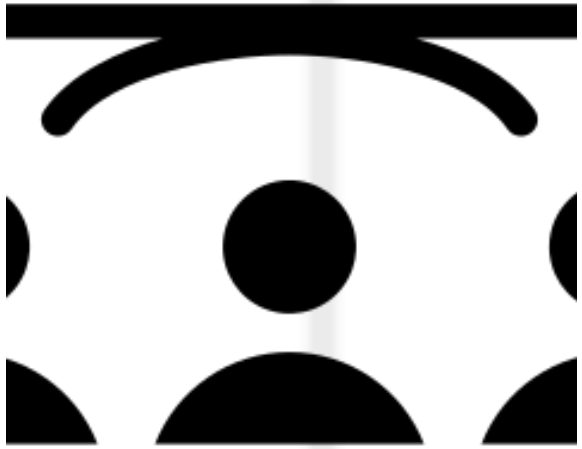
Size of the bubble represents the actual number of times the genre combinatio n

Each point represents unique Combination of genre

Genre: Comedy

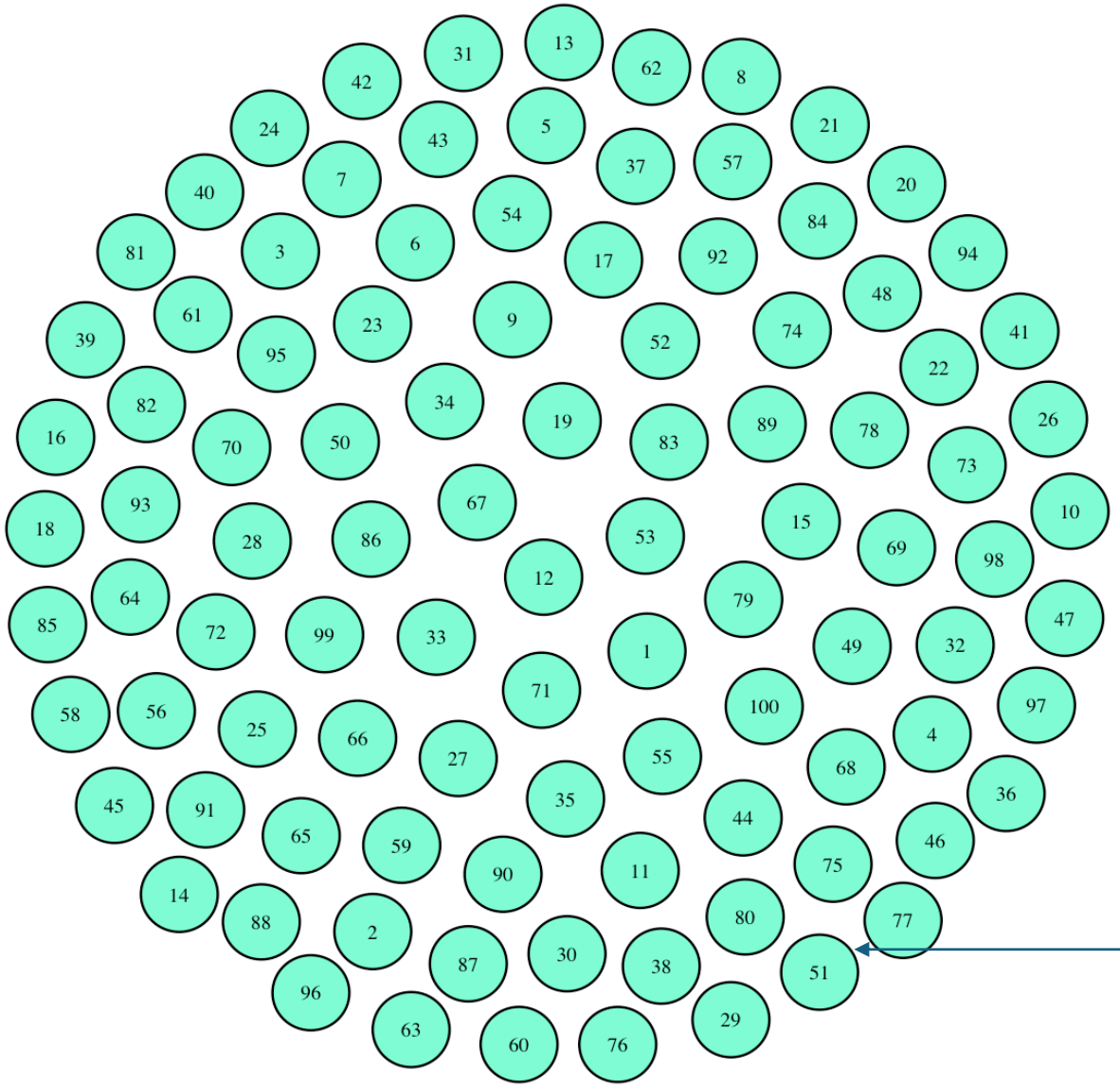
Movie: Bean (1995)





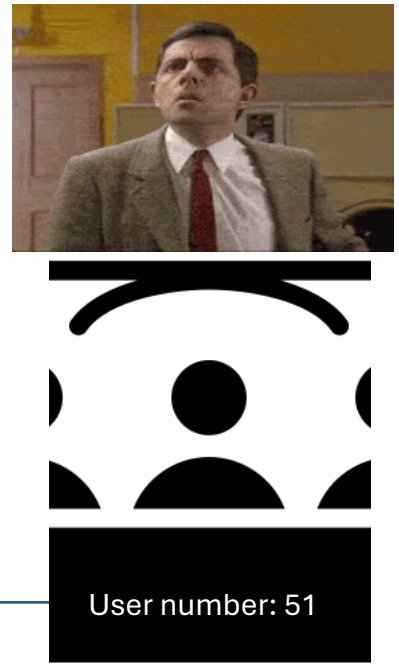
User number: 51

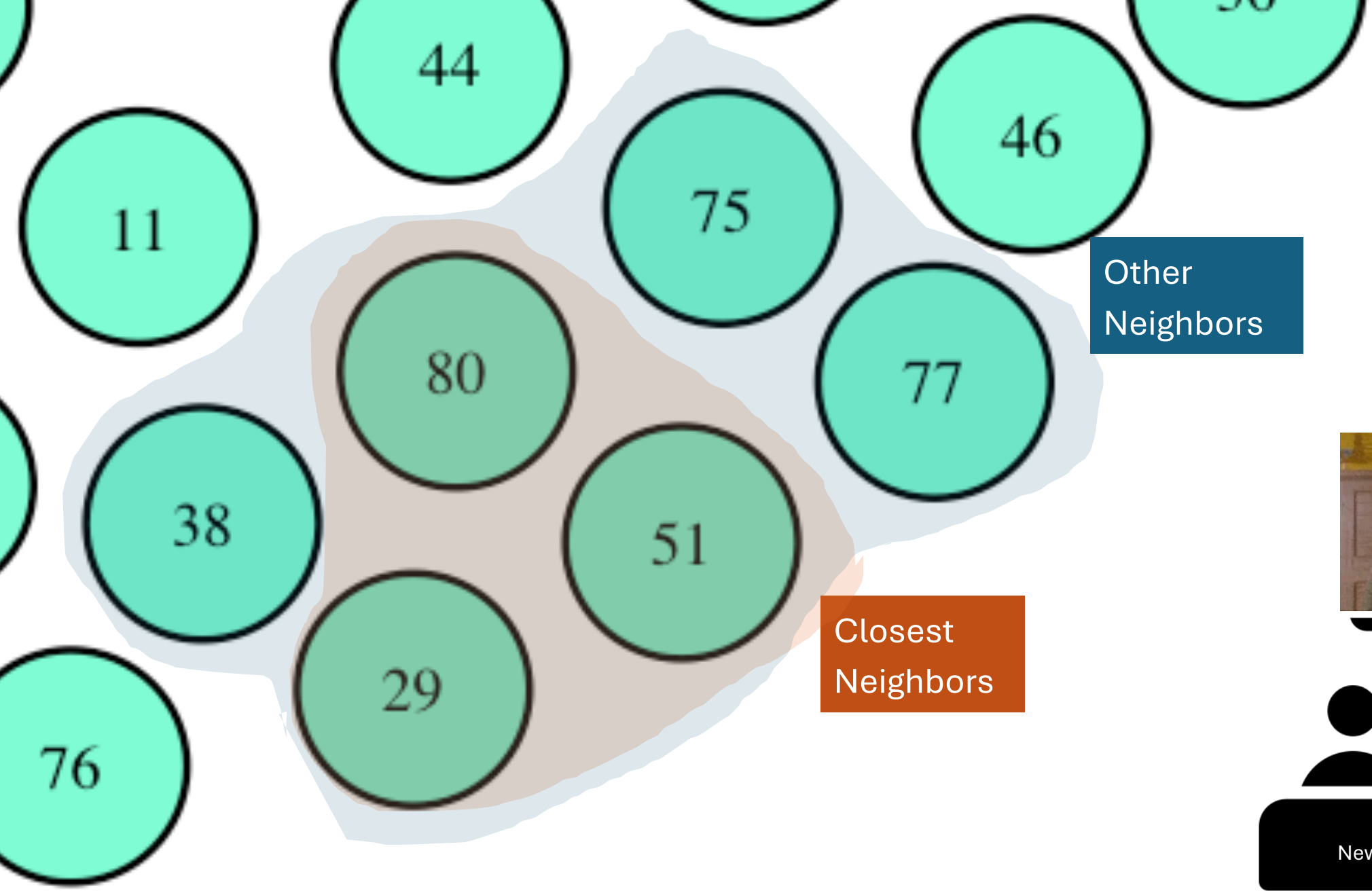
Nearest Neighbor Graph



Nearest neighbours is like finding the closest friends who have a similar movie preference as you.

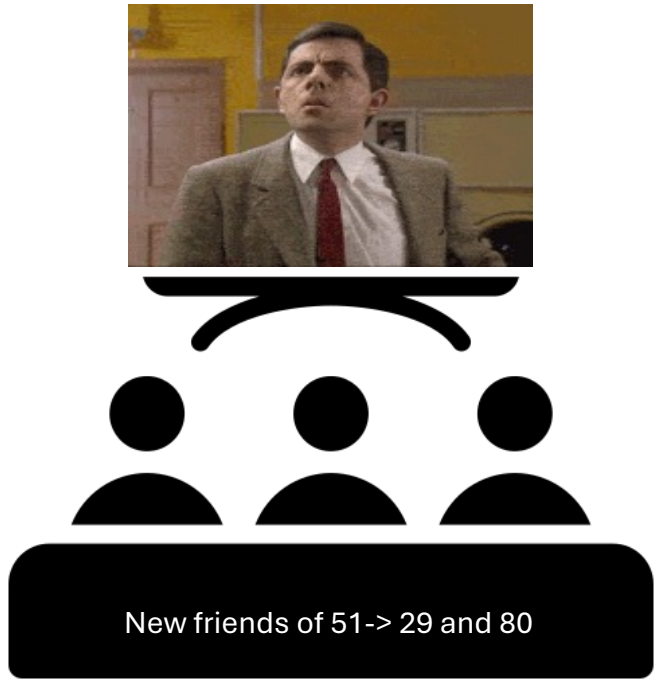
Here each number represents a person, and the neighbors tend to the have a similar genre preference.





Other Neighbors

Closest Neighbors

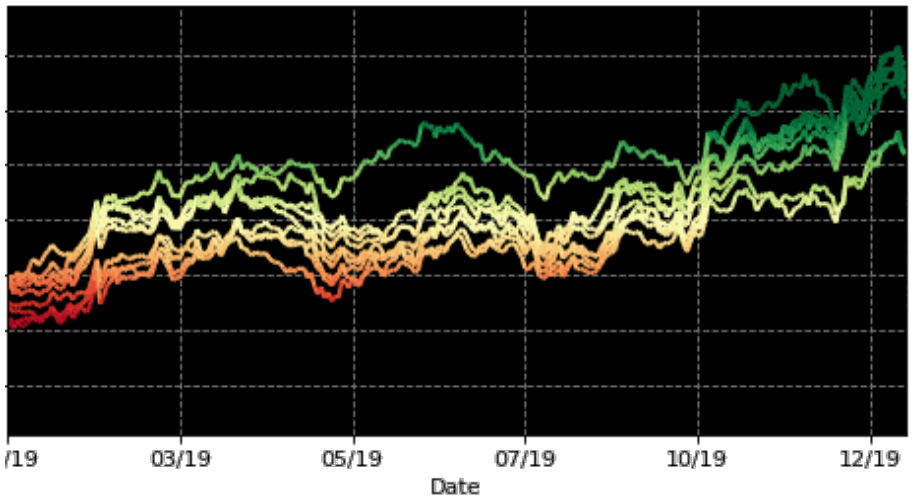


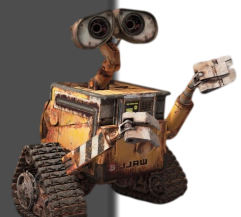
How our model can be further extended

- Time Series Analysis: Analyze ratings over time to see how preferences change, which could lead to dynamic recommendations that evolve with trends.

- Meta Data: Incorporate movie metadata (e.g., director, cast, description) and use text mining and natural language processing to recommend movies similar to those a user likes.

- FP-Growth Algorithm: We could use the FP-growth algorithm, which is generally faster than Apriori and does not generate candidate sets explicitly, thus improving the efficiency of the model.





Our current Model

After Extension

THANK YOU