

Individual Contribution and Learnings

Milton Ossamu Tanizaka Filho - r0822517

- Read raw dataset and processed to igraph data format
- Created weights feature for analyzing performance in clusters
- Informing global and local metrics of the network
- Created exploratory data analysis visualizations
- Implemented community detection algorithms
- Plotting network visualization discriminating by community

Learnings:

- Data wrangling with graph dataset
- Using igraph to process networks and apply community detection algorithms
- Analyzing different methods to choose a community that might fit for recommender system means
- Visualizing networks

Loris Gallo - r0906521

- Literature Review on methods of analyzing and grading methods based on various metrics
- Using igraph to calculate the hub-dominance and transitivity
- Data visualization
- Creating heatmap on correlation of NMI and Adjusted Rand index between different methods
- Reporting results created by the team

Learnings:

- Performing a thorough analysis on a big dataset by performing community detection algorithms.
- There is no right answer when it comes to selecting a specific algorithm, it depends on the situation and the preferences of the end-user.
- Learning how to visualize the message you want to transfer.

Kin Ho Chan - r0772534

- Literature Review on the difference of clustering methods and evaluation metrics

- Reporting results created by the team

Learnings:

- While the algorithms were good on paper, we discovered that the actual performance of algorithms may rely on the dataset and there is no “golden” answer of what algorithms to be used. Take the Louvain and Leiden as example, even tho the Leiden was reported to be superior than Louvain, our results seem to show differently. Proper evaluation of methods should be done before adopting such method for the recommendation system(although wrong prediction/clustering may do no harm to the customers).
- Learn about the implementation of community detection methods, the analysis pipeline for graph data analysis with the use of igraph and the respective evaluation metrics.

Hrishikesh Nambiar - r0909789

- Did Literature Review of the various clustering methods.
- Collaborated with team to create the clustering visualisation.
- Reporting on clustering and datasets.
- Tried Node classification and link prediction on the dataset.
- Implemented link prediction on the sample of the large amazon dataset.

Learnings:

- The Co-purchasing dataset due to its size requires scalable algorithms for its analysis.
- We were able to understand the nature and structure of the dataset majorly its cluster nature mainly through the lens of average rating.
- We noticed that real-world datasets are not easy and clean and sometimes require assumptions made to derive meaningful insights. Example in our case we had to assume only small sample to implement link prediction or majorly implementing random walks for node embedding is heavy computation process. Thus to know which nodes are more probable for connection requires high computation power or simplifying assumptions.
- Understand advantages of graph datasets when compared to tabular and relational database. Create data pipelines. Learning to use, implement and visualise clustering on python iGraph, use stellargraph api for node embedding and link prediction.