Hydergraph – Mapping Culture Through Food Heritage

Sohan Thummala
2023101019
Rudra Choudhary
2024121012
Saiyam Jain
2023101135

October 25, 2025

Abstract

Unlike existing latent feature based recommenders that tend to be opaque, network-based recommenders store explicit relationships between nodes which in our case are elements of Hyderabadi culture. This project aims to build a network-based recommender for Hyderabad's culture which is transparent, interpretable and can provide deeper insights into the connections between cultural elements in Hyderabad.

1 Introduction

Recommender Systems (RS) are a type of information filtering system designed to predict and suggest items or content—such as products, movies, music, or articles—that a user might be interested in. These predictions are based on the user's past behavior, preferences, or the behavior of similar users [1]. This project aims to create a recommender system for cultural elements in Hyderabad.

An opaque Neural Recommender System(RS) is limited in the reasoning it can provide for suggestions. We propose Hydergraph, a network based recommender for cultural elements in Hyderabad which proposes suggestions based on network data analysis and thus should be able to generate rich explanations for suggestions.

We aim to create a Network based RS for the city of Hyderabad. To this end we detail methods to generate network data and heuristic based recommendations. Fig.1 shows an example pipeline for a simple recommender based on adjacency. The user is directed towards elements in nodes that are direct neighbors of cultural elements that they like based on a simple keyword matching.

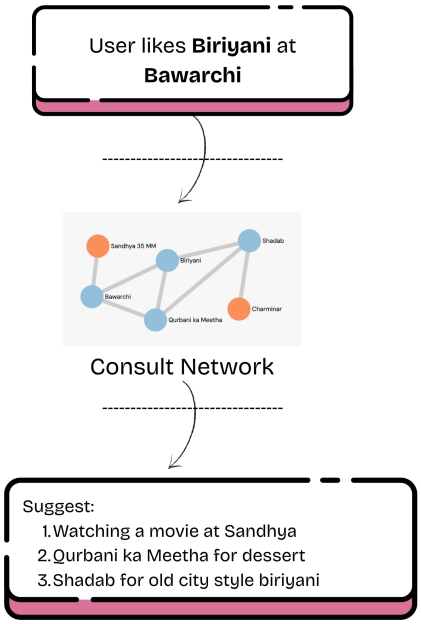

Figure 1: Basic Network based recommender workflow

1.1 Related Work

Our approach includes a network based recommender that does not involve neural network approaches but a comprehensive study of current approaches to building RSs is mentioned.

In the industry, RS enhance customer satisfaction and drive revenue growth by providing tailored suggestions. Major corporations such as Amazon, Netflix, and Spotify integrate RS into their operations, significantly contributing to their business models. For instance, Amazon reports that 35% of its revenue comes from its RS, while Netflix attributes revenues of approximately $33.7 billion and its success in customer retention significantly to its RS [2]. The methods used in these recommender systems are: Content-Based Filtering(CBF), Collaborative Filtering(CF), graph based, reinforcement learning(RL) and mixed approaches.

CBF and CF methods started off with hand-crafted features with small interpretable pipelines to generate recommendations but the current approaches are based on feature extraction using neural-networks resulting in methods based on feature vectors from these networks.

RL based RS models the recommender problem as a sequential decision making process. User history is taken as context data and the RL agent provides suggestions which are actions, rewards are associated with user retention and feedback, the agent is trained to maximise long term total reward. Some graph based methods involve GNNs and multi-modality which is not the focus of this project.

For a network based approach to building an RS, the choice of network is important, we borrow methodology from Satish et al. [3] where co-occurrence based method was used to generate network data for citation network analysis which is also our approach for network data generation.

2 Objectives and Motivation

2.1 Motivation

Hyderabad is rich in cultural elements that interact with each other in interesting ways. These elements can be the subject of common discussion due to their geographical proximity, historical similarity, preferences of the people discussing them, etc. Analyzing these discussions via co-occurrence networks captures many of these interesting characteristics.

2.2 Objectives

1. Collect text data by scraping blogs, social media, news articles, etc.
2. Create network data using co-occurrence methods from text data.
3. Explore centrality based, and similarity based methods to analyze network data.
4. Create a simple recommender augmented by network data trying methods like node similarity measures as well as path based and nearest neighbour methods.

3 Materials and Methods

3.1 Text Data

Text data was collected from the following sources:

1. Wikipedia articles
2. Travel blogs like TripAdvisor, Travel Triangle, etc
3. Food blogs like Taste Atlas, Eating Asia, etc
4. Zomato

3.2 Network construction

Entities are elements of hyderabad culture that form nodes. These were manually identified from the scraped text. Later additions to this can include using entity recognition tools from NLP toolkits like Apache OpenNLP and Stanford CoreNLP. Entity co-occurrence was the method used to obtain connectivity data, entities are said to have an edge between them(undirected) if the occur together. 2 different networks were constructed as follows:

1. Sentence co-occurrence
2. Paragraph co-occurrence

4 Initial Results

Note that most of the attached images and discussion is on the sentence based co-occurrence network hereon referred to as sentence network, paragraph network has been created and analyzed, the sentence based network was chosen for discussion as its results are intuitively more sound. A visualization for the top 30 nodes by degree and their connections is shown in Fig. 4.

The networks were constructed as explained in section 3.2. One such network visualized using Gephi is shown in Fig. 2. Node betweenness centrality(normalised) was calculated for all nodes in both networks, in the sentence based network, Charminar has the highest weighted betweenness of 0.178(note that this is normalized such that the sum of normalized betweenness of all nodes is 1) and unweighted betweenness of 0.246 here edge weights are calculated using number of co-occurrences. Other observations to note in the sentence network are that Charminar has a weighted degree of 34 which is the highest and unweighted degree of 14 which is second highest after Haleem which has an unweighted degree of 15 and a weighted degree of 24, second after Charminar.

Various centrality distribution plots for the sentence network are given in Fig. 3. Network level metrics are given in the table 1

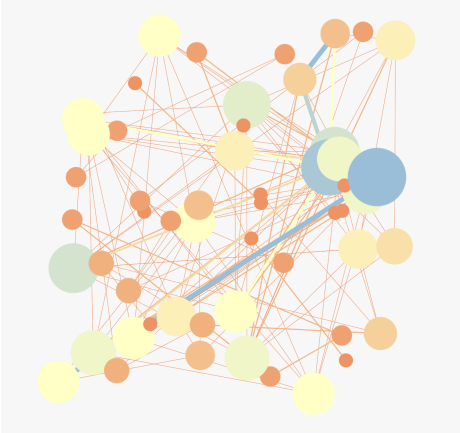

Figure 2: Sentence co-occurrence based network visualized using Gephi

5 Conclusion

The sentence co-occurrence network provides a network that has desirable properties like linking places that are close together, food that is available at the same restaurants,

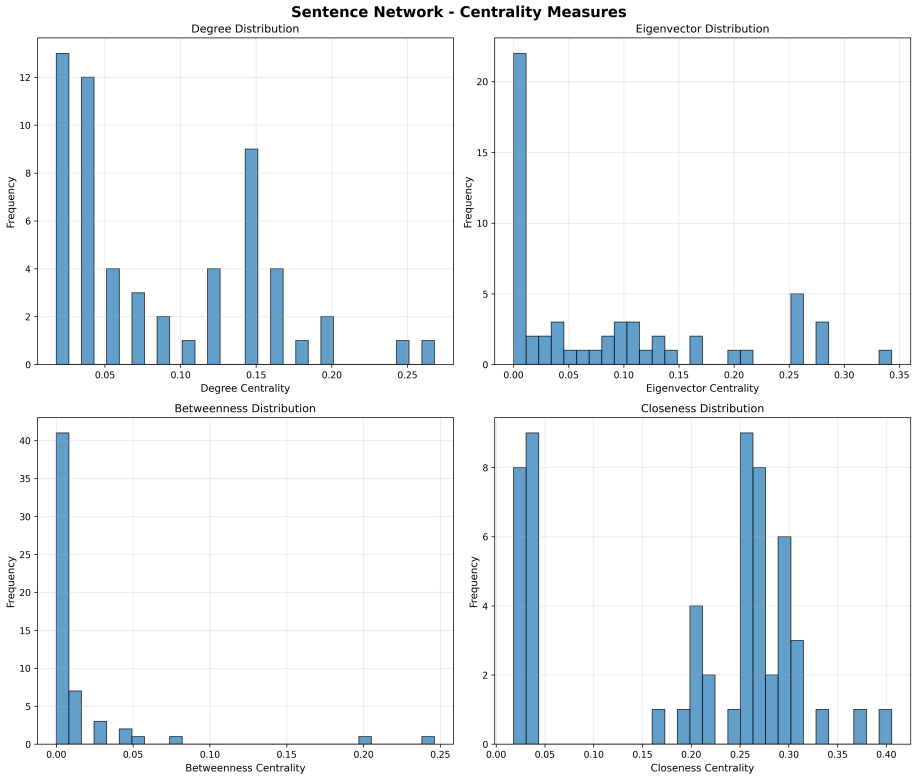

Figure 3: Centrality distributions on the Sentence Network

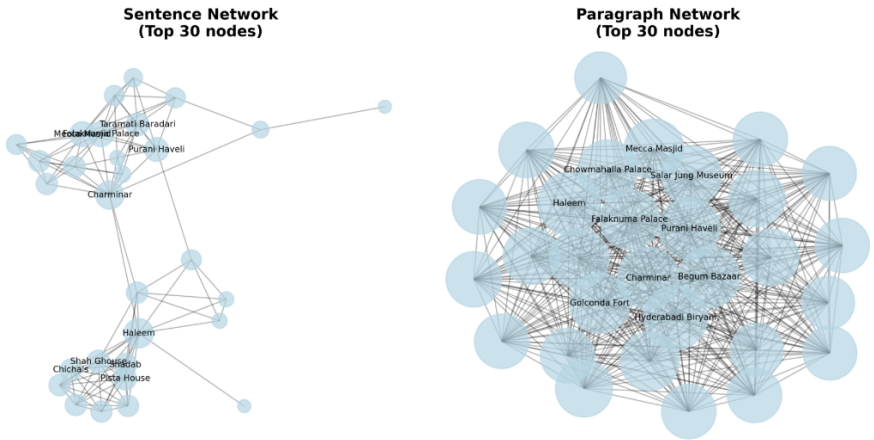

Figure 4: Top 30 nodes by degree from both networks

Table 1: Graph statistics at different text granularities

Level	Edges	Components	Avg. Clustering	Diameter	Avg. Path Length
Sentence	137	8	0.583	6	2.68
Paragraph	1496	1	0.862	3	1.45

restaurants that have similar cuisines, etc. Some data is also based on bloggers' personal recommendations/preferences. These properties make the network useful for a recommender system.

References

[1] Charu C. Aggarwal. Recommender Systems The Textbook. Number 4. Springer Cham, 2016.

[2] Shaina R. and Mizanur R. and Safiullah K. and Armin T. and Ananya R. and Farshad N. and Amirmohammad K. (2024) A Comprehensive Review of Recommender Systems: Transitioning from Theory to Practice. arXiv:2407.13699

[3] Satish Muppidi and K. Thammi Reddy. 2020. Co-occurrence analysis of scientific documents in citation networks. Int. J. Know.-Based Intell. Eng. Syst. 24, 1 (2020), 19–25. https://doi.org/10.3233/KES-200025