Math 168 Essay 1: The Filmtipset Network

Pratyusha Majumder

October 10, 2020

The Filmtipset movie network is a social network that depicts comments made by the user on Filmtipset.se, a Swedish movie rating website where users can befriend others in an asymmetric manner. There are $|V| = 104890$ nodes, representing users and movies, and $|E| = 1266753$ edges denoting when a user rates and comments on a movie. User-user edges denote friendship between users. This network is an unweighted multigraph and the edges are time stamped to depict when the comments and ratings were made. Figure 1 shows the full relationship network that the Filmtipset dataset represents.

A study by Said, Luca and Albayrak of Technische Universitat Berlin, was conducted using this network to explore how social relationships affect user similarities in movie tastes [**?**]. A subgraph of the Filmtipset network was used, which is emboldened in Figure 1. The subgraph contains nodes representing users and movies, with user-movie edges depicting user ratings for movies, and user-user edges depicting friendship or a fan-relationship between two users. User-user edges are directed, and their relationships are either symmetric or asymmetric. $u_a \rightarrow u_b$ shows an asymmetric friendship that illustrates that user u_a was a fan of user u_b. In this case, u_b is known as a friend of u_a, and unlike u_a, u_b had an unchanged social graph. In a symmetric relationship, $u_a \leftrightarrow u_b$, both users have a changed social graph and are known as friends of each other. Both symmetric and asymmetric user-user relationships were used to investigate users' similarities in movie tastes [**?**].

The researchers created a binary representation of the user-movie relationship from the Filmtipset network, where 1 indicated that users had seen the movie and 0 indicated that users had not seen the movie. This was denoted as C_{all}. Another version of this matrix only included users who had greater or equal to 5 friends (i.e. at least 5 symmetric relationships with other users), denoted by C_{r5}, and another version with movies that had been seen atleast 5 times and users who had seen atleast 5 movies, denoted by C_{um5}. Then, a Jaccard similarity coefficient was calculated for each user pair with symmetric and asymmetric relationships, using Equation 1.

$$J(A|B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

Their results are shown in Figure 2. The average Jaccard relationship for user pairs with symmetric relationships in C_{all}, C_{um5}, C_{r5}, were the highest in their respective groups (0.18, 0.18 and 0.20 respectively). These coefficients were approximately double the average Jaccard coefficient of all user pairs (0.10). This shows that users are more likely to have watched the same movies as those who they have symmetric friendships with on the platform. An interesting observation is that user pairs with asymmetric friendships had a higher average similarity score than the average score for all users, showing users have some similarity in taste with those they form asymmetric friendships with. Said, Luca and Albayrak concluded that reducing the social graph to just symmetric friendships could result in a higher density user-movie matrix [**?**].

This study shows that user-user friendships, whether symmetric or asymmetric, can be leveraged to create better recommendation systems for users on streaming platforms. To extend the findings, one could keep the original user-movie matrix with the ratings on a 1,2,3,4,5 scale, where 1 depicts a low score and 5 depicts a high score, and explore the user-user social graphs to investigate similarities between users' ratings of movies. If they are found to be significant, then those findings can also give one an idea about how users' enjoyment of movies compare to those in their social graphs.

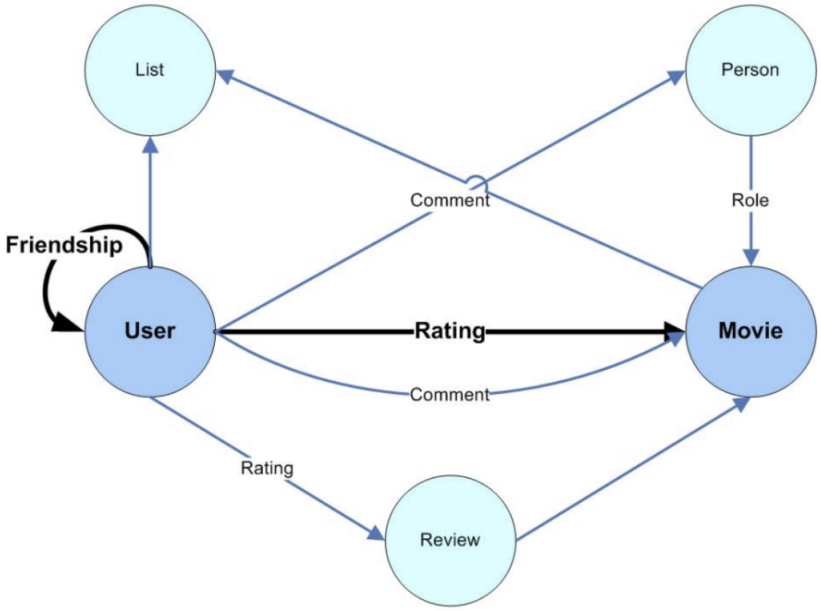

Figure 1: Graph representing all relationships in the Filmtipset dataset

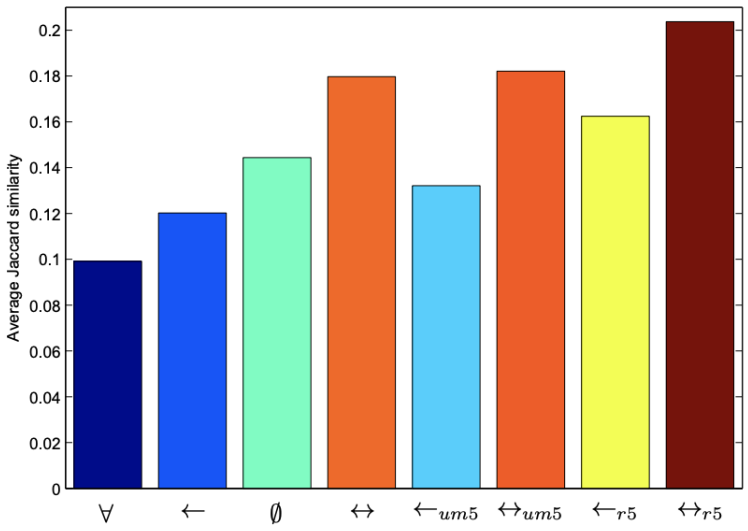

Figure 2: Average Jaccard similarity scores for user pairs, ($\forall$) denotes all user pairs in C_{all}, ($\leftarrow$) denotes asymmetric friendship pairs in C_{all}, ($\emptyset$) denotes user pairs with no friendships, ($\leftrightarrow$) denotes symmetric friendship pairs in C_{all}, ($\leftarrow_{um5}$) denotes asymmetric friendship pairs in C_{um5}, ($\leftrightarrow_{um5}$) denotes symmetric friendship pairs in C_{um5}, ($\leftarrow_{r5}$) denotes asymmetric friendship pairs in C_{r5}, ($\leftrightarrow_{r5}$) denotes symmetric friendship pairs in C_{r5}.

1 References

[1] Said, A., De Luca, E., and Albayrak, S., How social relationships affect user similiarities, in *Proceedings of the IUI Workshop on Social Recommender Systems*, Association of Computing Machinery, New York (2010).