**CPSC 572**

Project Proposal

**Netflix Movie Network Analysis**

Group # 11

Ishrat Naba, Jose Perales, Simrat Singh

**1. Nodes and Links**

One side of the bipartite network will consist of nodes representing “**movies”**, and the other side will have nodes representing “**people”** involved in the movies. The node “**people”** will be subdivided into “Actors/Actresses” and “Directors”, differentiated based on the color of the nodes.

The links/edges will represent the involvement of a person in a movie.

Example of the links include: Actor <--> Movies and Director <--> Movies

**2. Dataset**

Our team will analyze certain properties of movies in Netflix, and the relation between movies and their cast. Our group decided on using the dataset available on: <https://www.kaggle.com/datasets/shivamb/netflix-shows/data>

We can use the above dataset without any restrictions, and no permission is required to make use of it. The dataset is in CSV format and includes information about TV Shows and various other fields that are irrelevant to our analysis. These irrelevant data fields and insufficient data will be cleaned before analyzing the data.

**3. Expected size of the network**

The expected number of movies in the dataset is around 6,100, while the expected number of people, including directors, actors and actresses is roughly 42,700. The number of links is expected to be around 45,600+. Other than irrelevant information, our team plans on using the entire dataset available.

**4. Questions you plan to ask and why we care**

This analysis aims to unravel the behind-the-scenes dynamics of Netflix Movies, shedding light on the personalities driving their success and the patterns that dictate their popularity. We intend on answering questions related to performance metrics by category, prevalence of ratings, the actor decision-making process and stardom/career trajectories. This analysis will provide both valuable insights for Netflix and offer a comparative lens for other streaming platforms seeking to optimize their content strategy. By understanding the symbiotic relationship between directors, actors, and the success of specific categories, our findings aim to empower the industry with actionable data for informed decision-making.