Q1: How do you build the network from this given dataset?

Ans: Well, I used the given information in problem description which said, we need an actor for a node and similar movies done by two different actors being connections. That's it, I went ahead chopped down those un-necrssory features from the data set, leaving only then actors_1_name, actor_2_name, actor_3_name and movie_title feature. There on, I went ahead and Label Encoded using scikit learn built-in features to convert those categorical values of actor columns in to numerical values to create a metrics from them as demanded in problem set.

Q2: How do you find two subnetworks?

Ans: First I Preprocessed the data set finding columns that were needed, LabelEncoded the categorical values, convert them to float int type, flatten them all and extracted the movie matrix. Once I had the movie matrix based on actor columns and movie title, all I needed to loop through the data set for actors having done similar movies, and grab at-least 40 samples for 20 nodes / subnetwork.

Q3: What is/are your metric(s) to compute similarity between the two subnetworks?

Ans: I used Pearson-correlations to find out the correlation between each subnetwork. These subnetworks based on nodes (actors) and their connections (similar moives done). Check out LabelEncoded version of these subnetworks as subnetwork1 and subnetwork2 dataframes.

Q4: What is your design to show the similarity?

Ans: There could be a lot of possible design architectures to built this. I used Pair-wise correlations using pandas built-in methods. There could be whole lots of other techniques used. I discussed about them in detail in Jupyter Notebook file.

Q5: Simply hop over to the Jupyter Notebook file, run every functions associated with the Task # X specified and see your results. This could also be run in browser using Python frameworks like Django and Flask, or using Selenium built-in python library, and could also be modified making a complete front-back end interface.

Thank You!