Agenda (3pm. Sept. 29):

- Discuss questions and directions of interest based on early EDA.
- Identify interests to pursue in pairs.
- Create pairs based on interests and strengths and weaknesses!

Everyone:

- Skim the paper: https://arxiv.org/abs/2007.04448

Marlin:

- I created a repo for this project: https://github.com/marlinfiggins/West-Coast-Datathon-F2020
- Deliverable: Working on model framework for generating endorsement hierarchies
    - Assume we're working with bipartite network

Xingzi:

- Look at the statistical relationship between covariates.

Aditya and Shashank:

- Create a (well) parameterized function to generate the social network from scratch

Agena (7pm. Sept 30):

- Discuss progress
- Adjust $ terms for inflation if necessary
- Look at deliverables
- FIgure out the next directions!
- Current concerns:
    - Figuring out which covariates go into utility function
        - How we we factor for ensemble casts and significance of role
    - Figuring out which visualizations best describe our project for datafolio / report
    -
    - Going from rankings to success, how does Oscar winning affect hierarchy?
        - Model 1: Oscar win/nomination at time t as a covariate in utility function at time t+1
        - rankings affecting Success and success affecting rankings (success is removed from the mechanical model

Things to cover in the report:

- Issue of directed vs undirected
- Talk about how we got covariates of interest
    - What influences hierarchy
- Is hierarchy indicative of success
- Is success relative of hierarchy
- How we computed probabilities of actors working
- Decay parameter lambda

- Individuals work as an actor is different than as a director

What has to be done?

- ~~Figure out covariates~~
- Graph development
- Inference scheme for Hierarchical model
- Inference for Oscars
- Report writing
    - Figure out what you want visualize
        - Hierarchical Ranking
            - Highlight interesting cases
        - Decay parameter lambda
        - Graph (possible based on variables)
        - Justification of different covariates
        - DAG for inference pipeline.

Marlin:

- Working on inference pipeline, working with Aditya on graph data production

Aditya

- **Graph dev, configure output properly and optimize inputs/processing speed**

Xingzi & Shashank:

- Identify Covariates

Agenda (7:00pm, Oct 1.)

Marlin: Visualization of the ranks (assuming I get covariates and graph). Writing script for running the model. Running the model.

Aditya: Graph finalized, look at viz options for graph

Xingzi: Visualization and work with Shashank on the reformatting of covariates dataframe.

Shashank: Give Xingzi starter code for the reformatting of covariates. Brainstorm ideas for Oscar <-> inference with hierarchy.

To be done:

- Coerce covariates to proper format
    - Output dataframe as .csv in data folder
    - Put helper function for formatting in /src/
- Coerce graph to proper format
    - Get it in and out of a file w/helper functions

- Graphs and beginning report writing
    - Motivation for final covariates
        - For visualization, things like gross on the log scale.
        - Otherwise just make the viz cute. (axis, titles, etc.)
    - Writing plots for hierarchy model.
    - DAG of workflow

Covariates are:
All variables are running as in only matter up to time t.

- Have you been nominated for an Oscar?
    - Director
    - Actor
- Have you won an Oscar?
    - Director
    - Actor
- Total gross (inflation, logs)
- Proportion of movies in particular genre(
- Proportion of movies in particular rating (G, PG, ….)
- Mean score
- Mean votes

Data format for covariates:

Agent | is_director | ID | Covariate 1| … | Time

| | | | |

Ideally, write a helper function so that we return numpy array which we can index as:

S[t][n_agents][k_covs]

Note: Individuals who are both actors and directors will have their covariates computed separately and will have two ID numbers associated with them.

Data format for graph:

Write helper function so that we return numpy array which can be indexed as:

Delta[t][n_agents][n_agents]

Agenda (7pm Oct. 2.)

Things to discuss:

- Deliverables
- Timeline
- Telling everyone to go back and take notes on what they've done
    - Identify important aspects of your work that should be included in report
- Identifying core contributions and observations
- Figure critique
- Discuss final features and how implement

Next Steps/Deliverables:

Marlin:

Aditya: Technical Expositions -> transitioning from data to graph / why it was a reasonable structure to use to model hierarchies, runtime considerations, combine visualizations with this portion

Xingzi: Update and t-test plot & Data exploration & Covariates significance

Shashank: Covariates transformation

Limitations of Analysis

- Top 220 movies in a year (in general are higher rated actors/directors) versus population
    - Our population is lead actors and directors of popular movies since 1986 (each movie only features one star, either the leading actor or the leading actress)

- 82/488 is oscar nominated people who have not featured as a director/lead actor for a popular movie
- How oscar win affects the hierarchy of lead actor and directors of popular movies

Paper outline:

Non-Technical Executive Summary – What is the question that your team set out to answer? What were your key findings, and what is their significance? You must communicate your insights clearly – summary statistics and visualizations are encouraged if they help explain your thoughts.

- We construct a network of actor-director relationships.

b. Technical Exposition – What was your methodology/approach towards answering

the questions? Describe your data manipulation and exploration process, as well as your analytical and modeling steps. Again, the use of visualizations is highly encouraged when appropriate.

- **Data exploration & Covariates significance**
- **Covariates transformation**
- **Graph**
- **Math Modeling**
- **Inference**

Agenda (Oct. 3., 3pm)

Goal: Review of rough overleaf doc.

% TODO: Example task.

Takeaway from this:

- Final decision of figures and edits from one another.
- Combining ideas for Non-technical summary.