**Project Title**

**Names - Student IDs**

# Short Project Description

Describe your project shortly here. Focus on what problem you will try to solve and its scope here. Tell why this problem is important and needs to be solved.

You can choose between:

* Data analysis/Modeling project: discover interesting relationships within a significant amount of data
* Algorithmic project that extends/builds on a known method or algorithm.
  + Extend/improve/speed-up some existing algorithm
  + Define a new problem and solve it

# Methodology

## What is the problem/question you are solving?

Give a brief but precise description or definition of the problem or question

## Dataset

What dataset you will use. Why is the data you plan to use appropriate? Does it have the right labels or information?

You can find some interesting datasets at:

* <https://ourworldindata.org/coronavirus-data>
* <http://snap.stanford.edu/class/cs224w-2018/data.html>
* <https://archive.ics.uci.edu/ml/datasets.php>
* <https://www.kaggle.com/datasets>

## How will you solve the problem? Plan of action.

Describe the methods, algorithms, techniques in detail. Also explain how will you scale it up.

Examples:

1. We will create edit histories of every article. We will then compare article edit histories and analyze how users are leaving Wikipedia since all the easy/obvious articles have already been written.
2. Our hypothesis is that friends have similar tastes. We will include a regularization term to a Latent Factor Rec Sys which will encourage neighboring users to have similar parameters.
3. We will implement a scalable frequent itemset based approach to identify cluster seeds (complete bipartite subgraphs). In the second pass we will then use a random walk based approach to expand around the seed and extract the clusters.

## How will you evaluate your method?

How will you measure the performance or success of your method? What baselines you will use.