Idaho Cities' Thresholds for Renewable energy Adaptation (City Names Network)

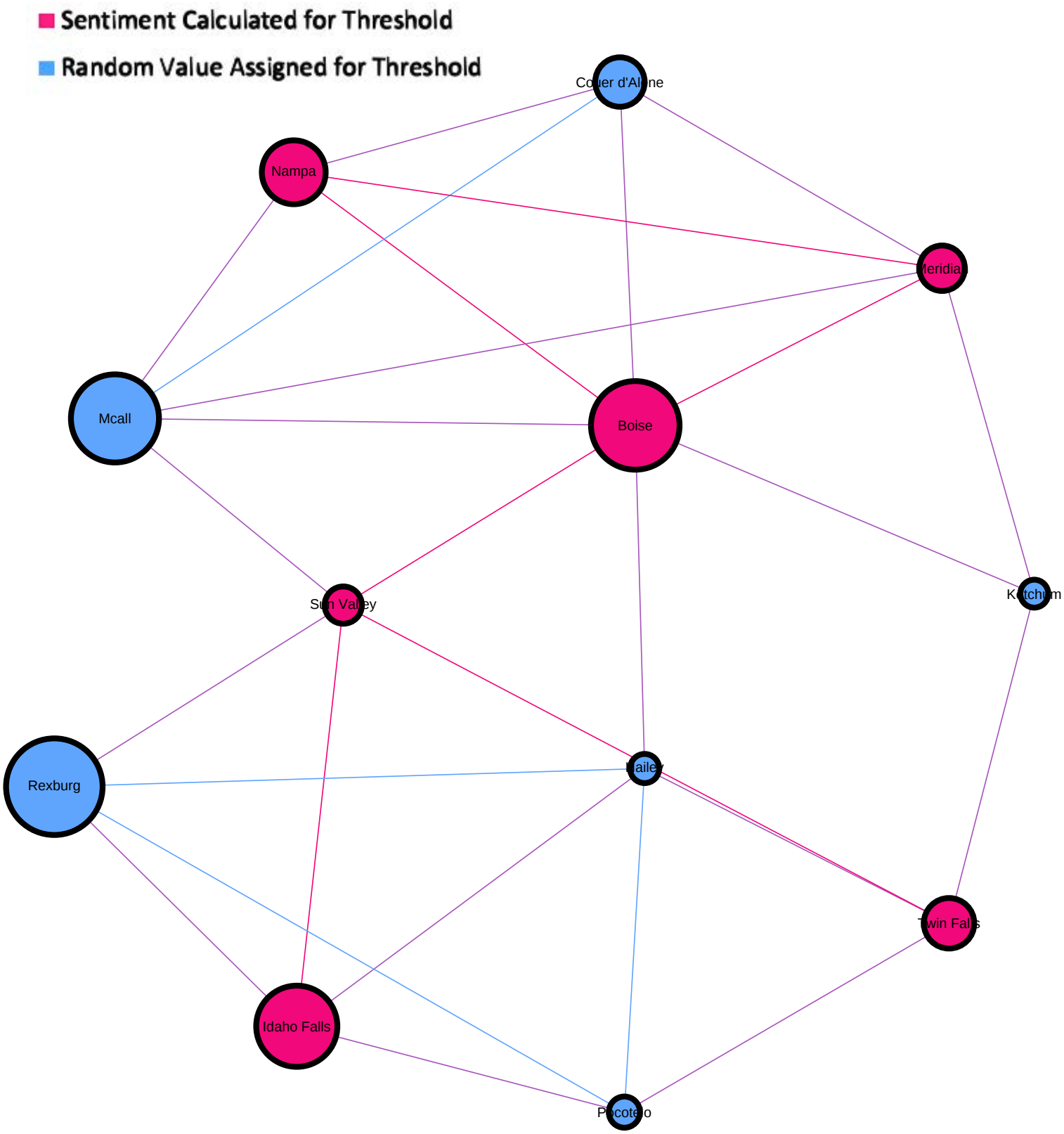

Figure 1: Gephi generated Visualization of 12 cities in Idaho and there relative thresholds for adopting Renewable Energy Sources. Nodes are connected based on their three nearest neighbors. Nodes are sized proportionally to their relative threshold values. For six cities Sentiment Analysis was run on Newspaper articles written about Renewable Energy as it pertains to each city. The other six were assigned random values for threshold values. Node Labels are the corresponding city.

Idaho Cities' Thresholds for Renewable energy Adaptation n (Threshold Values Network)

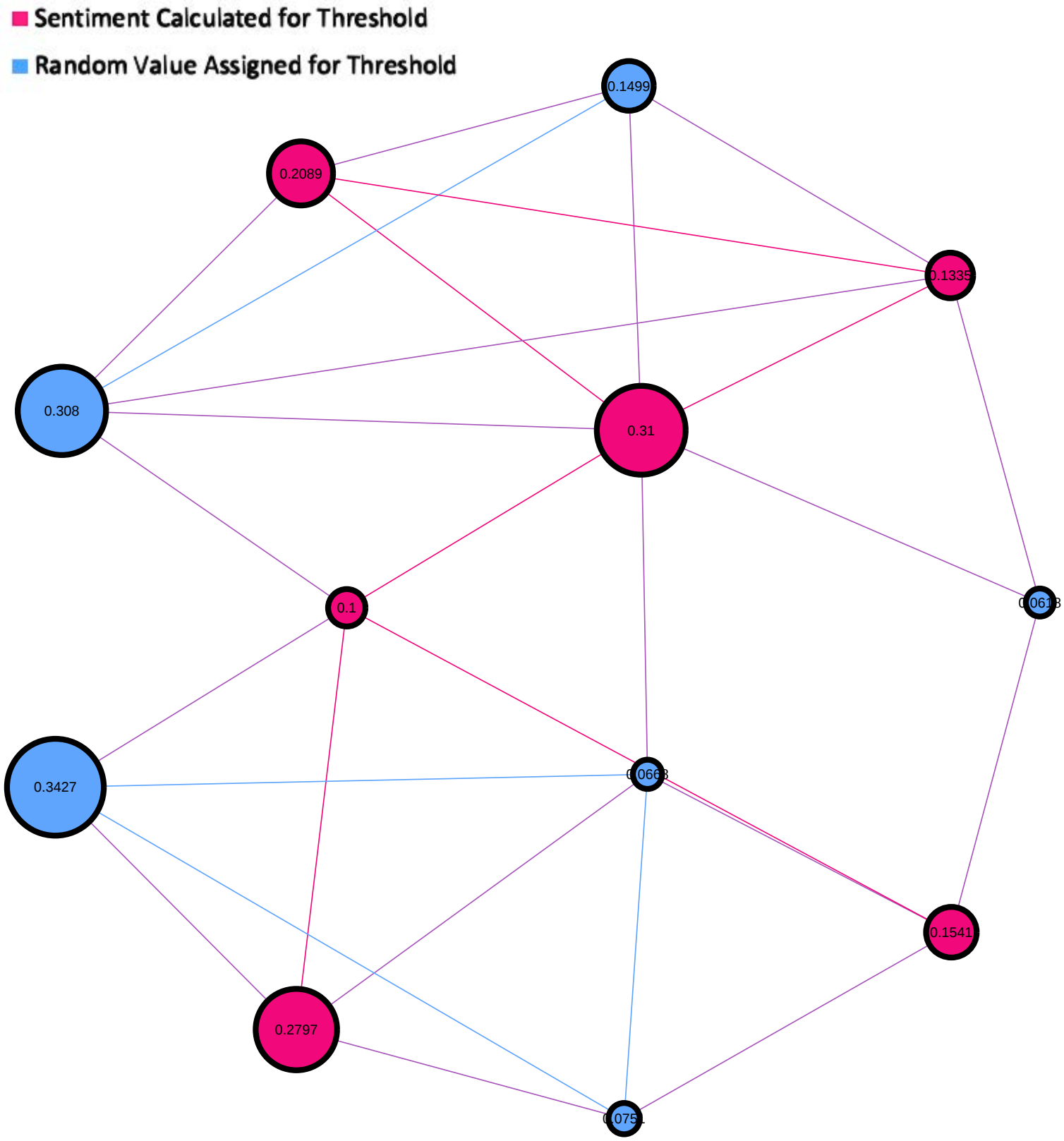

Figure 2: This is the same Gephi Visualization as above such that nodes are connected by connecting to there nearest neighbors, and nodes size are proportional to their threshold value. however, the nodes labels now display their normalized threshold score from [0 to 1]. . However, where a city name use to lie, now contains the normalized threshold value. I.e. the large central with a score of 0.31is Boise, Idaho.

Idaho Cities' Thresholds for Renewable energy Adaptation (Explanation and Reflection)

Introduction

The basic research question of my question is: How do we effectively get more cities in Idaho to adopt forms of Renewable energy? I decided to use my home state as there is so much available land, yet there seems to be issues in getting sustainable forms of energy implemented. There are several ways to approach this problem, but based on what I have learned in this class as well as other DCS classes was to turn this into an Influence Maximization problem. I will not go into depth here, but the essence of it is as follows. Given a network of nodes and edges, select the smallest, specific number of nodes, who if they adopt the behavior, in my case implementing forms of renewable energy. In this type of problem, a node will adopt the new behavior if the fraction of that nodes neighbors exceed their threshold. This problem becomes very important in modern social networks -- think Viral Marketing, who us the influencer you target who will cause their the max number of their followers to buy the product. You can read more about it here: https://www.cs.cornell.edu/home/kleinber/kdd03-inf.pdf

Methods

I started out by selecting 12 of the biggest cities in Idaho, as well as my home town, purely out of interest. My network model was structured such that each city is a node. A node is then connected to its nearest 3 cities (as the crow fly's). The bases of my whole model is a Python program. Once I had a connected graph of cities, the next step was to assign threshold values to each cities. I did so using a very similar technique to what we did in class. For each city, I used an EBSCO database to pull articles. For the sake of fairness, I used the exact same search term for each city: <city name> "Renewable Energy". I tried to select the four most recent examples, but often ran in to hurdles and had to thoroughly read the articles to ensure they actually pertained to my subject matter. As such I was limited to a small number of articles per city. Then for each city I was able to retrieve appropriate data ran Sentiment analysis on all the articles. I just to tokenize (break the article down sentence by sentence) the sentiment analysis. I then normalized the data by the number of sentences. This proved more accurate than running Sentiment Analysis on the article as a whole. I averaged the scores for all the articles for a given city, giving me a preliminary threshold value for each city. For the six cities I was unable to gather enough data, I randomized the threshold value uniformly between 0 and 0.4 (a touch above the highest result for standard Sentiment Analysis.

Limitations

Two related major issues presented themselves. First and foremost was the sentiment analysis dictionary I used. As we discussed in class, there is no dictionary specific to our course of study. My initial plan was to use a machine learning algorithm to build a dictionary specific to our area of study. This would have been very beneficial to my project, but collecting the training data to train a model became too time consuming.

Idaho Cities' Thresholds for Renewable energy Adaptation (Explanation and Reflection)

Limitations (cont.)

This would have been very beneficial to my project, but collecting the training data to train a model became too time consuming. I was having to manually pull and convert files so I could not get enough data to train a model I was comfortable with applying to the sentiment analysis algorithm. As Such, I settled on an existing dictionary: VADER. IT has been proven to be fairly accurate across several different disciplines.

Next Steps

The next major step in this process is to collect more data. My primary focus is to collect 4 more recent articles for each of my existing working cities. This will serve to improve the overall accuracy of the Analysis. Then I want to expand the number of cities I am running S.A. on by collecting articles on these cities.

Ultimately, when I get this set to a point where I feel the model is generating a network where nodes all have appropriate thresholds, I will run an Influence Maximization Algorithm on it, to determine the seed sets in Idaho that, according to this model, would enable the maximum spread of renewable energy sources across the entire network.

Code Repository: https://github.com/maxtanous/influence_energy_problem

EBSCO Database: http://web.a.ebscohost.com.lili.idm.oclc.org/