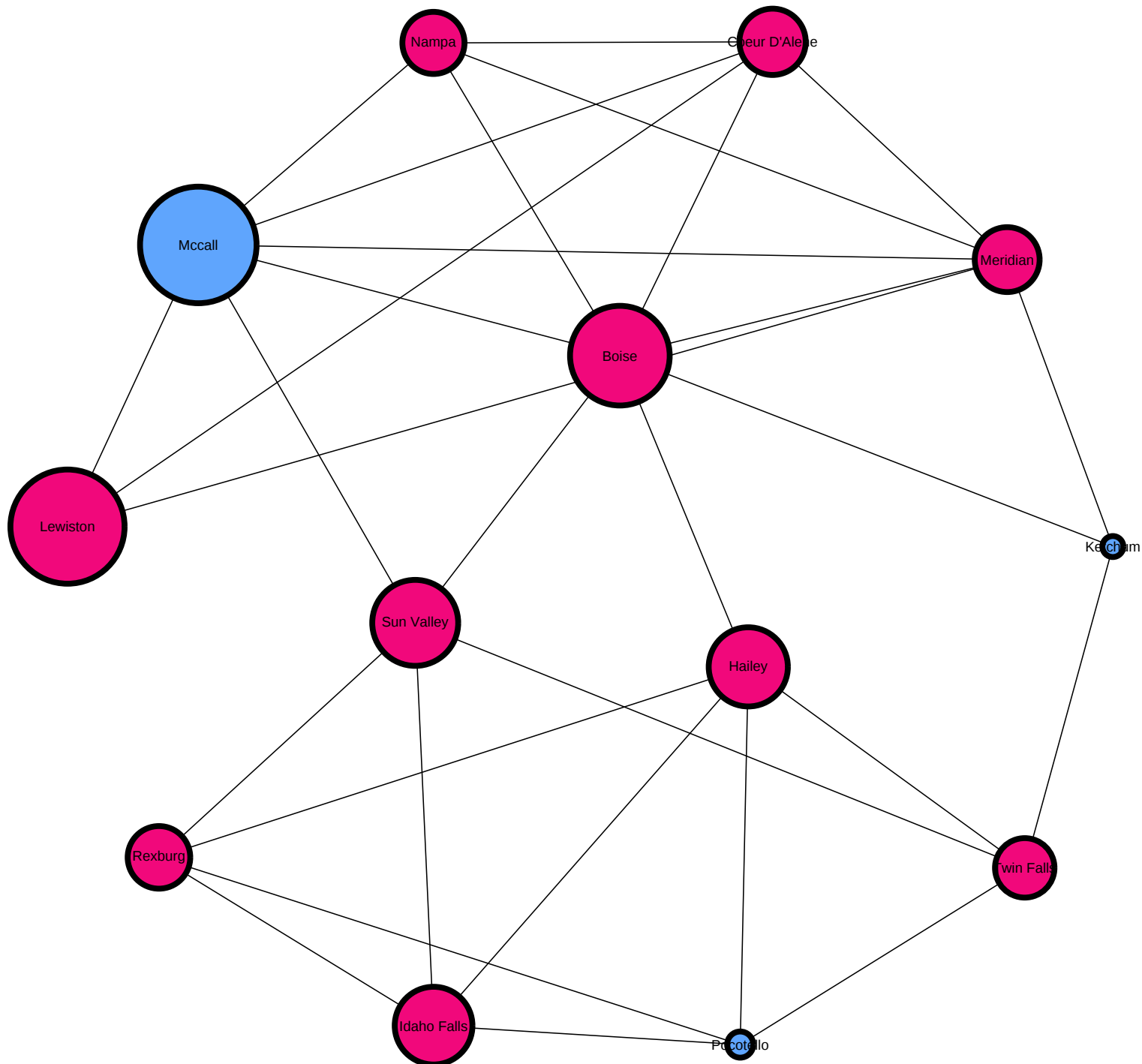# Network Comparison Between Different Corpus - Articles

Figure 1: This network was constructed using both proximity and and scores from sentiment analysis. Each node was sized according to score derived from sentiment analysis run on a corpus of newspaper articles written in that city regarding Renewable Energy. The larger the node the more positively the articles for that city discussed renewable energy. Each node was connected to its three nearest nodes.

Network Comparison Between Different Corpus - Legislation

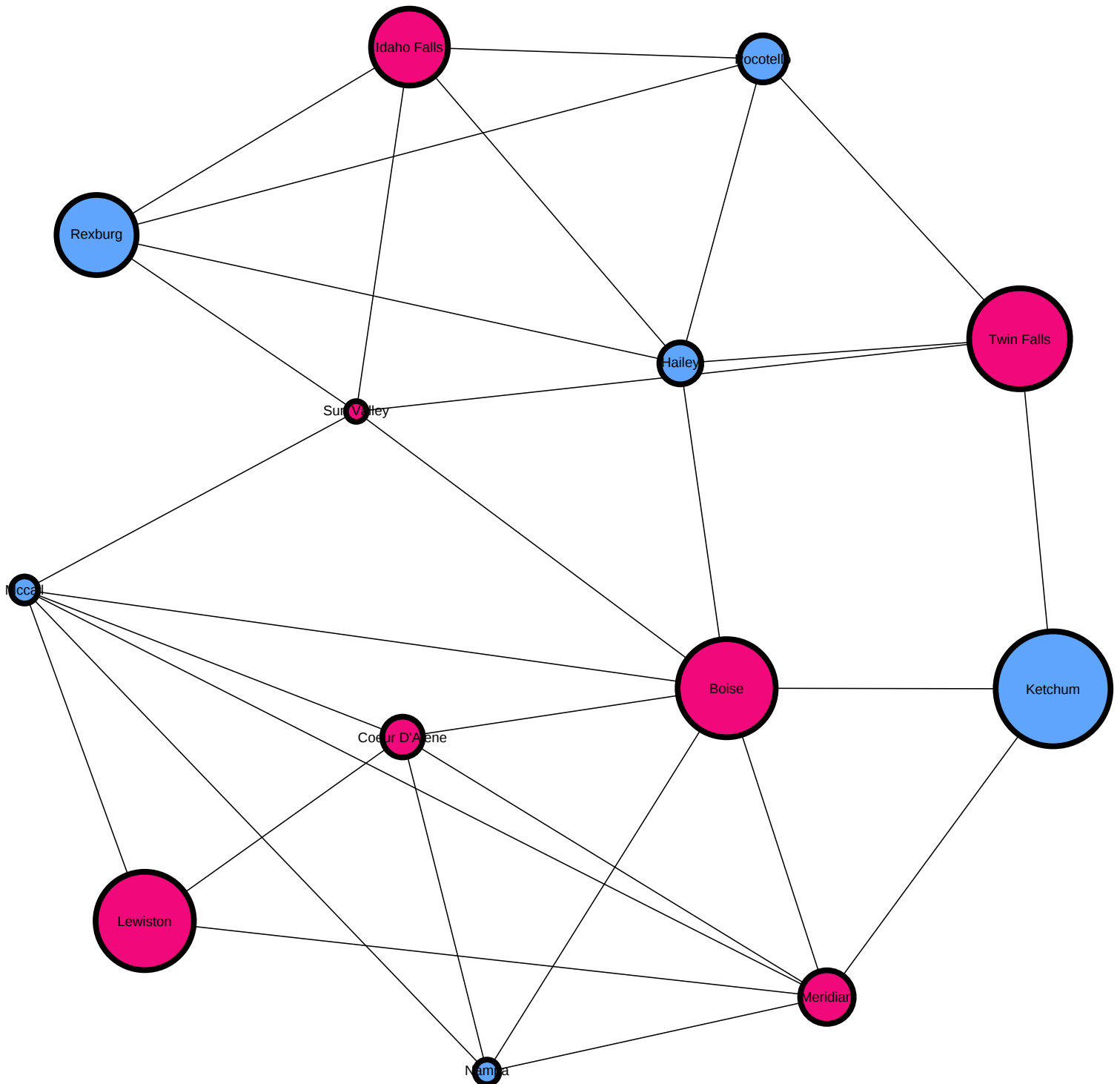■ Sentiment Calculated for Threshold
■ Random Value Assigned for Threshold

Figure 2: This network was constructed using both proximity and and scores from sentiment analysis. Each node was sized according to score derived from sentiment analysis run on a corpus of Legislation regarding the city and Renewable Energy. The larger the node the more positively the articles for that city discussed renewable energy. Each node was connected to its three nearest nodes.

Network Comparison Between Different Corpus - Reflection

After my initial attempts to create a Sentiment Analyzer using Machine Learning failed, I decided, for several reasons, to switch my search Database. I decided to use Nexis-Uni, and it allowed me to dramatically increase my corpus size and improve the Sentiment Analysis.

The first figure shows the use of Newspaper articles in Sentiment analysis. I used the same search patterns for each city to ensure fairness. I searched for "Renewable Energy Idaho", and then sub-search for the city name. If a search for a city returned less than 10 results, I randomized its score. However, I also capped the number of articles at 50. For example, a search on Boise returned over 170 results, but I sorted by relevance and took the top 50. As can be seen in the figure, I was able to calculate scores for 3 more cities than in my previous model, and calculate them with a significantly greater number of data-points. The total corpus size was 250.

The second figure used sentiment analysis on legislation regarding cities and renewable energy. I did not limit the Jurisdiction to just Idaho, I allowed US-Federal Legislation that met my criteria to be included in the corpus. In a similar fashion to the newspaper articles. If a city search returned less than five documents I randomized its score. I capped the number of documents for one city at 30. As can be seen in the figure, I had to randomize 5 cities. The total corpus size for the legislative approach was 103.

I wanted to compare how the scores differed based on the type of text being analyzed. I wanted to use the corpus whose analysis best represented the sentiment of the city. After running these experiments I decided to stick with Newspaper articles for my articles. The Sentiment Analysis dictionary works better on this less formal text, and I was able to generate a substantially bigger corpus size, and in turn randomize less cities. I will discuss how this all fits together more in my final report.