

For the utilization of Graph Theory I would propose utilizing a web scraping application for job posting boards like linkedin, or indeed. With their APIs I would be able to collect the most recent job postings of a specific role (ex: Data Scientist, Data Engineer, Machine Learning Engineer) and utilize this information for sentiment analysis. Sentiment analysis would help in creating a word count for the most in demand skills, proficiencies and character traits that companies are looking for. In doing so, I would gain insights into the sentiments associated with different job descriptions. I could create categoricals such as entry level roles, mid-level roles, location, etc. This information can help identify positive or negative trends in the job market for specific roles. Most importantly we could convert this data into our needed network representation which could measure our degree centrality and help determine our most important factors in the network. We can then compare these centrality measures across various skills to further gain knowledge on in demand skills.

To actually load the data I would use a scrapy bot to comb through the different job postings and receive the position title as well as the position's job description. Moving forward we would cleanse the data and get it ready for analysis. We could then use tf-idf to obtain our most frequent terms per job posting and then utilize network x to math Company nodes to various job skills.