Skill Barriers to a "Just Transition" of US Fossil Fuel Workers to Emerging Green Jobs, Final Project Report

Xizhi Wu
School of Computing and
Information, University of Pittsburgh
xiw183@pitt.edu

ABSTRACT

Nowadays, we are witnessing a "Just Transition" of fossil fuel workers into emerging green jobs. However, job polarization is widely observed to become a major concern in many societies. Occupations are differentiating into high wage occupations and low wage occupations, requiring distinct skill sets. As a result, mobility for workers between high wage occupations and low wage occupations is constrained. In our "Just Transitions", this type of difficulties could also be encountered by workers in fossil fuel jobs when adapting to green jobs. We examine this as skill barriers and try to analyze the barriers using the correlation between our derived occupations' similarity scores and the number of job transitions in a resume dataset. We present this as a novel way to study the skill barriers between jobs requiring distinct skill sets.

1. INTRODUCTION

With the initiative to develop a more economic and sustainable way of utilizing natural resources, the green energy industry is on the rise. While this can be beneficial to our society in the long run, some eminent re-employment issues for fossil fuel related workers demand attention and solutions from our society. The booming of the green energy industry brings with it the shrinking fossil fuel industry and its decreasing job positions. It would be ideal if our fossil fuel workers can be re-employed into green jobs with good working conditions and salaries. However, some of the green jobs require a different skill set compared to fossil fuel occupations, giving skill barriers as constraints on employers' mobility[1].

In this final project, we can quantify skill barriers as a distinction between two occupations by computing the difference between the skill sets they each required. A weighted occupation network can be constructed based on this, with each edge weight indicating the job complementarity[1]. In an occupation network, topologically closed nodes should be more similar than nodes far away. We used the node2vec[2] method to derive node embeddings for all the occupations in our network. The similarity betweens any pair of occupations can be computed with cosine similarity between their correspondent node embeddings. Higher occupation similarity between occupations indicates more overlap on the skill sets. We assumed that more overlap on skill sets leads to easier transition between jobs and vice versa. To verify this assumption, the correlations between similarity of jobs and the number of job transitions needs to be examined. However, our result indicates this assumption is not always tenable, so we offer some possible explanations in the discussion section and propose some future work that might overcome the limitation in this report to get a more comprehensive result.

The rest of the report is structured as follows. In section 2, we present datasets used and preprocessing on them. In section 3, we present technical details on our method. We conclude with a discussion in section 4 on our limitations and possible future work to overcome the limitations. In section 5, we present our acknowledgement. In section 6, we present supplementary materials that weren't present in the first 5 sections.

2. Data

2.1 O*NET dataset

In this project, two datasets were used. The first dataset is O*NET dataset[3], containing 52 csv files each describing a unique skill and how important it is for different jobs. Each csv file has 5 columns, an overview of the dataset describing the skill "Arm-Hand_Steadiness" is presented in Table 1.

Table 1. Arm-Hand_Steadiness

Importance	Level	Job Zone	Code	Occupation
81	71	5	29-1022.00	Oral and Maxillofacial Surgeons
81	57	2	51-6051.00	Sewers, Hand
78	66	5	29-1021.00	Dentists, General
…	…	…	…	…
10	Not relevant	4	11-9199.10	Wind Energy Development Managers
…	…	…	…	…

In our project, we mainly use three columns from the csv file, "Importance", "Level" and "Occupation". If the "Level" is "Not Relevant", then we will change "Importance" to 0.

We have selected 39 occupations containing 15 fossil fuel jobs and 24 green jobs. This is all the related occupations we can find so far. I can list some of the occupations here, they are "Rotary Drill Operators, Oil and Gas", "Service Unit Operators, Oil and Gas","Petroleum Engineers", these are fossil fuel jobs. "Solar Photovoltaic Installers","Wind Energy Operations Managers","Hydroelectric Production Managers", these are the green jobs. We have put the entire list as a column in section 6.

2.2 Resume Dataset

This dataset contains people's resume data describing their work experiences. In order to use this for our task of computing correlations between real life job transitions and job similarity, a lot of preprocessing and statistics need to be carried out. Thankfully, we can directly use another researcher's already well-processed dataset on this and only write a thank you in the

Acknowledgements. The well-processed dataset contains the number of job transitions from 1990 to 2019, an overview of the dataset is presented in Table 2.

Table 2. Processed Resume Data

Index	S_Soc	D_Soc	Date	Counts
0	15-1252	15-1252	2017	7800
1	15-1252	15-1252	2016	7513
2	15-1252	15-1252	2015	7290
…	…	…	…	…
1149367	25-1081	39-6012	2015	1
1149368	55-3019	55-3019	2019	1

The "S_Soc" and "D_Soc" correspond to the source of this job transition and the destination of this job transition. And "Counts" indicate how many transitions take place in this year. The first three rows are all about one job, and "15-1252" is the occupation code for Software Programmer. We can tell that Software Programmers really like switching companies.

3. Method and Results

3.1 Skill Complementarity

Occupations are distinguishable by their skill sets. Therefore, we try to identify an occupation by calculating revealed comparative advantage[1] to recognize its "revealed" skills according to

$$\mathrm{rca}(j,s) = \frac{\mathrm{onet}(j,s) / \sum_{s' \in S} \mathrm{onet}(j,s')}{\sum_{j' \in J} \mathrm{onet}(j',s) / \sum_{j' \in J, s' \in S} \mathrm{onet}(j',s')}$$

The revealed comparative advantage is an index used in international economics for calculating the relative advantage or disadvantage of a certain country in a certain class of goods or services as evidenced by trade flows. Here we use this as an indicator for each job's effective use of skills[1] to distinguish different occupations. A skill is revealed if rca>1, then e(j,s) = 1. Otherwise e(j,s) = 0.[1] Skill complementarity is then calculated as follows:

$$\theta(s,s') = \frac{\sum_{j \in J} e(j,s) \cdot e(j,s')}{\max\left(\sum_{j \in J} e(j,s), \sum_{j \in J} e(j,s')\right)}$$

With all these, we can construct a network of skills with skill complementarity as edges, and visualize all the edges with a skill complementarity greater than or equal to 0.5. We only consider 39 fossil fuel or green jobs on the job list we find. Below are some visualizations.

In Figure 1 it is a visualization on skill complementarity with a threshold of 0.5 after Louvain community detection. We observed that it falls into two polarized skill sets. Community on the left is social-cognitive skills, community on the right is sensory-physical skills. Skills frequently observed together in different skill sets of jobs are more likely to be inside the same community. They are sparsely connected outside the two communities and densely connected inside each community, indicating the 39 jobs have fallen into two separate communities requiring skill sets hardly overlapping between social-cognitive skills and sensory-physical skills.

In Figure 2 it is a visualization of 4 jobs by their skill sets. From top left to bottom right, they are "Rotary Drill Operators, Oil and Gas", "Petroleum Engineers", "Nuclear Technicians" and "Nuclear Engineers". We present the complete visualizations for 39 jobs in section 6. It is observed that most jobs have a skill set highly concentrated on the left or on the right, meaning that they highly rely on social-cognitive skills or sensory-physical skills.

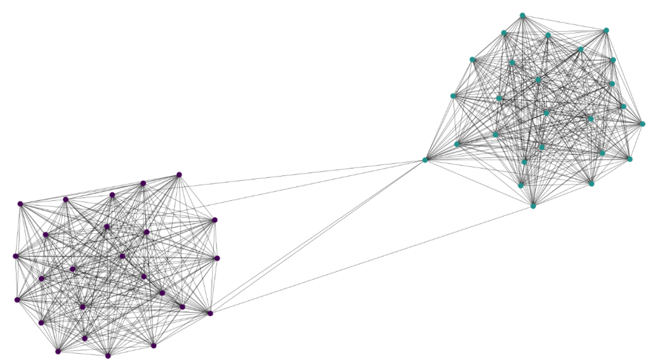


Figure 1. A visualization on skill complementarity with a threshold of 0.5 after Louvain community detection.

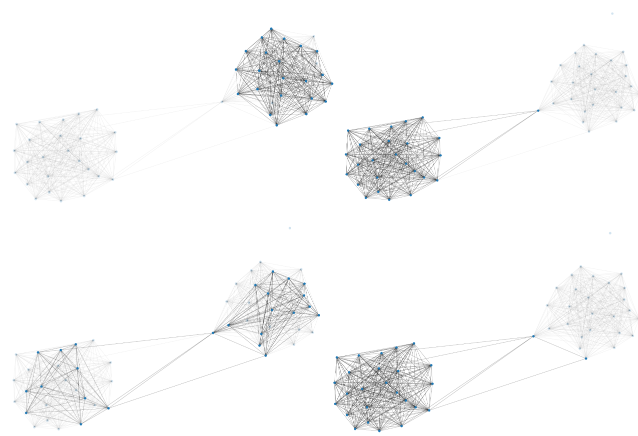


Figure 2. Visualization of 4 jobs by their skill sets. From top left to bottom right, they are "Rotary Drill Operators, Oil and Gas", "Petroleum Engineers", "Nuclear Technicians" and "Nuclear Engineers".

3.2 Job Complementarity

Similarly we derived the job complementarity with previously calculated comparative advantage, but we sum our e(j,s) over skills instead of jobs.

$$\Theta(j,j') = \frac{\sum_{s \in S} e(j,s) \times e(j',s)}{max(\sum_{s \in S} e(j,s), \sum_{s \in S} e(j',s))}$$

With this, we get our visualization of job complementarity with a visualizing threshold of 0.5 in Figure 3.

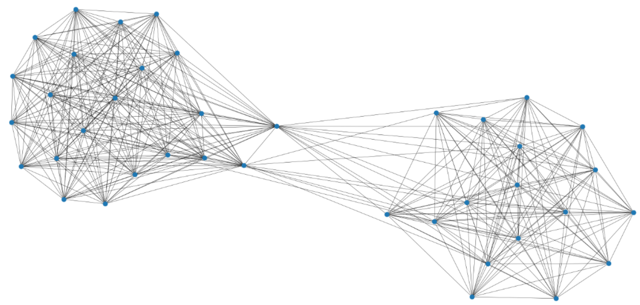


Figure 3. Job complementarity with a visualizing threshold of 0.5

In Figure 3, we can also observe two obvious communities with the left being jobs highly relying on sensory-physical skills and the right being jobs highly relying on social-cognitive skills, which work in concert with our skill complementarity. The most connected node in the middle between two communities is "Nuclear Technicians", which is the bottom-left visualization in Figure 2. It is observed in Figure 2 that "Nuclear Technicians" rely on both sensory-physical skills and social-cognitive skills. That could be the reason for its topological location in our Figure 3. We also highlight fossil fuel jobs and green jobs in our visualizations.

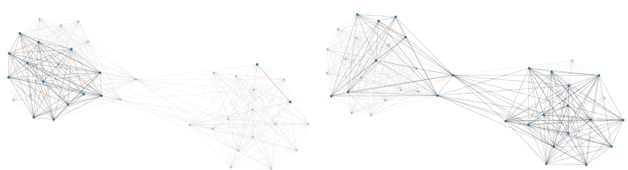


Figure 4. Fossil fuel jobs on the left and Green jobs on the right

In Figure 4 we can see that fossil fuel jobs mostly rely on sensory-physical skills while green jobs rely on both. Suppose a fossil fuel worker wants to make a transition to green jobs, it would be easier to transition to a green job with a similar skill set since the worker is already equipped with those skills.

Moreover, most fossil fuel jobs are on the left, therefore we assume that job transition is more likely to happen within the left community instead of across the two communities. And our skill barriers would be the different skill sets required by fossil fuel jobs and green jobs that make transition difficult. We try to verify this assumption in the next part.

3.2 Network Embedding with Node2Vec

We apply our job complementarity network with the node2vec algorithm to get node embeddings. Node2vec[2] uses a biased random walk to generate walk sequences. Nodes in the same walk sequence are considered similar because these walk sequences are going to be processed in a SkipGram[4] model to get embedding for the nodes. Therefore, topologically close nodes should have similar embeddings. By calculating the cosine similarity between any pair of nodes, and then studying its correlation with the number of job transitions between the same pair of nodes, we can verify our assumption.

Our random walk is based on the complete network of jobs instead of the visualization in Figure 3 because it only contains edges that have a job complementarity greater than 0.5. The job complementarity will be our edge weights. Our weighted biased random walk favors edges with higher weight. A visualization of the distribution of weights is presented in Figure 5.

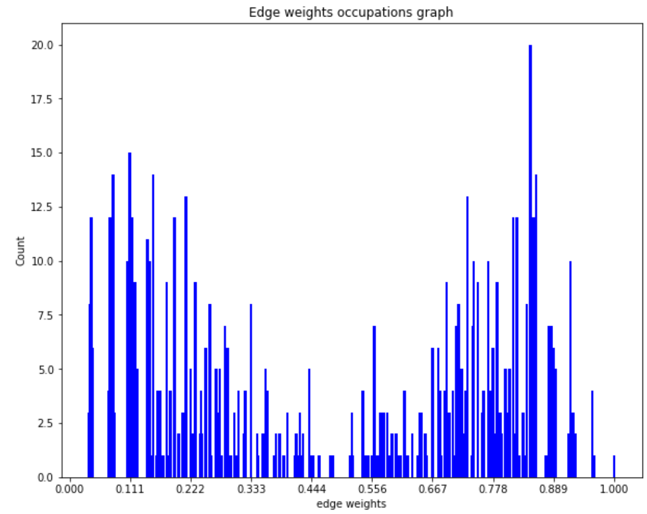


Figure 5. The weights are falling into two intervals, weights smaller or greater than 0.5. Our job complementarity network visualizes all the edges with weight greater than 0.5. The edges with a weight smaller than 0.5 are mainly across the two communities and weights greater than 0.5 are mainly inside those two communities.

Another factor influencing our random walk are the hyperparameters p and q. High p value means the walk is less likely to walk back to visited nodes and low p value will keep the walk local to its starting node. For hyperparameter q, setting it to q>1 causes the walk to be biased toward local nodes. And setting q to q<1 will lead the walk to visit further nodes. In our settings, we have p=1 and q=0.5.

We have a walk length of 100 and window size of 5. Our embedding is set to be a vector of length 1024. Now we are ready to get embeddings for all 39 jobs.

Then we collect the number of job transitions from the year 2009 to 2019 for the 39 fossil fuel jobs and green jobs. This is processed into a list of job transitions with the total numbers, then we calculate the pearson correlation scores between job transitions and cosine similarity of jobs. We have a correlation value of -0.1579 and a p-value of 0.24, indicating that they might not be correlated at all, which greatly contradict with our previous assumption of a positive correlation. We discuss possible explanations and future work that may overcome this limitation.

4. Discussion and Future Work

4.1 Result Discussion

There are a few possible explanations for our result. First, we used RCA value to construct our skill complementarity network and job complementarity. However, RCA emphasizes on the most distinct feature of a skill among all the jobs, but this most distinct part does not decide the job transitions, the common part or the majority part of the skills of a job decide the job transitions. This could be the major drawback in our research method.

There are other factors also influencing our result. We did not investigate the job market sizes for our job transitions. There could be jobs in the market that have a demand size big enough to allow jobs transition requiring different skill sets to happen. To avoid this, we need to standardize the transition number with their market size, penalizing the bigger markets. For example, there are 141 transitions between “Rotary Drill Operators, Oil and Gas” and “Petroleum Engineers”. Even though they require distinctive skill sets, but there might be huge demands for “Petroleum Engineers”

Another problem could be the granularity issues. As we can see from Table 2. that all the jobs have a code length of 6 digits, while in our job list of 39 jobs, a few of the jobs have 8 digits codes and share the same code on its first 6 digits. We remove the jobs that are sharing code with other jobs other than fossil fuel or green jobs.

Last but not least, we only considered 39 jobs. If we consider all the jobs, we could have more node combinations in our biased random walk, subsequently our embeddings could represent their relationships more accurately. Maybe two jobs are only different within the scope of 39 jobs we considered, but they are actually similar and there exists a bridge node between them outside our scope of 39 jobs.

4.2 Future Work

First we could find a substitute for RCA to better represent job transitions.. RCA emphasizes how that skill is being most effectively used by that job, while job transition does not always rely on the most effectively used skill.

Then we can further investigate their market size and standardize them. After that we can try to expand our scope of jobs to be greater than the 39 jobs to get a more comprehensive result.

5. Acknowledgments

6. REFERENCES

[1] Alabdulkareem, A., Frank, M. R., Sun, L., AlShebli, B., Hidalgo, C., & Rahwan, I. (2018). Unpacking the polarization of workplace skills. Science advances, 4(7), eaao6030.

[2] Grover, Aditya, and Jure Leskovec. "node2vec: Scalable feature learning for networks." Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. 2016.

[3] https://www.onetonline.org/find/descriptor/browse/2.A/2.A.1

[4] Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." Advances in neural information processing systems 26 (2013).

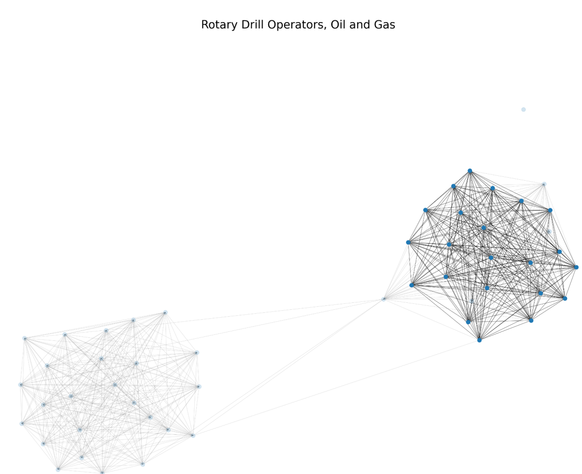
7. Supplement Material

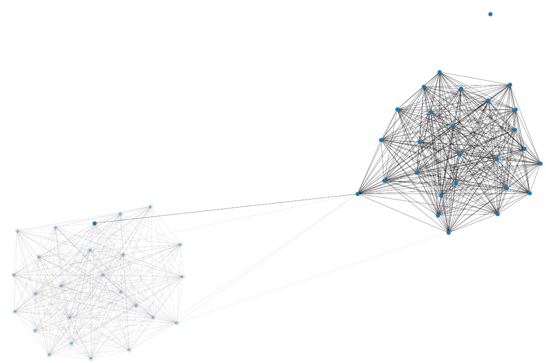
6.1 Job List

Complete Job List of 39 Fossil Fuel Related or Green Jobs:

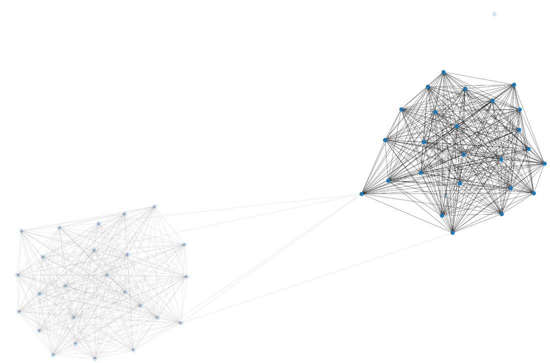
'Rotary Drill Operators, Oil and Gas',
'Roustabouts, Oil and Gas',
'Service Unit Operators, Oil and Gas',
'Petroleum Engineers',
'Derrick Operators, Oil and Gas',
'Petroleum Pump System Operators, Refinery Operators, and Gaugers',
'Wellhead Pumpers',
'Pump Operators, Except Wellhead Pumpers',
'Gas Plant Operators',
'Gas Compressor and Gas Pumping Station Operators',
'Energy Engineers, Except Wind and Solar',
'Crushing, Grinding, and Polishing Machine Setters, Operators, and Tenders',
'Continuous Mining Machine Operators',
'Loading and Moving Machine Operators, Underground Mining',
'Sailors and Marine Oilers',
'Geothermal Technicians',
'Geothermal Production Managers',
'Biomass Plant Technicians',
'Biomass Power Plant Managers',
'Biofuels Processing Technicians',
'Biofuels Production Managers',
'Biofuels/Biodiesel Technology and Product Development Managers',
'Nuclear Engineers',
'Nuclear Technicians',
'Nuclear Power Reactor Operators',
'Nuclear Monitoring Technicians',
'Environmental Economists',
'Solar Thermal Installers and Technicians',
'Solar Energy Systems Engineers',
'Solar Sales Representatives and Assessors',
'Solar Energy Installation Managers',
'Solar Photovoltaic Installers',
'Wind Energy Operations Managers',
'Wind Energy Development Managers',
'Wind Energy Engineers',
'Wind Turbine Service Technicians',
'Sustainability Specialists',
'Hydroelectric Production Managers',
'Hydroelectric Plant Technicians'

6.2 Skill Used by Occupations(39 Occupations)

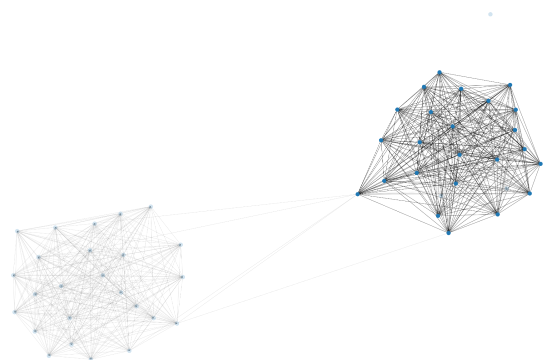




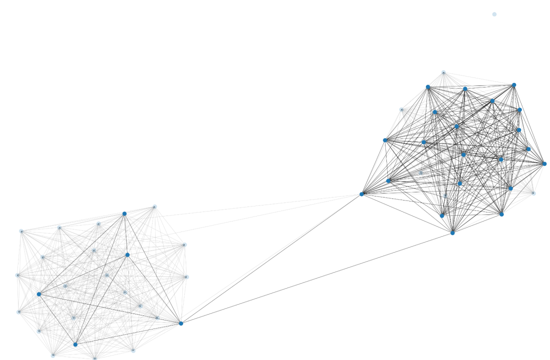
Roustabouts, Oil and Gas



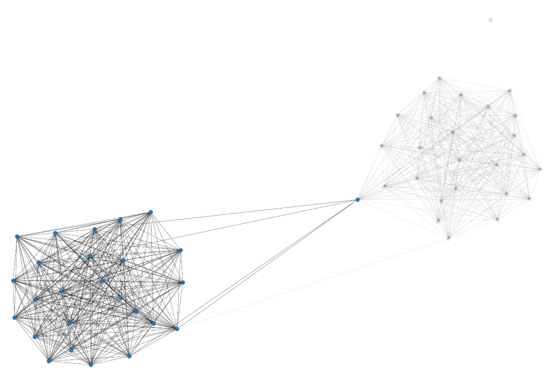
Derrick Operators, Oil and Gas



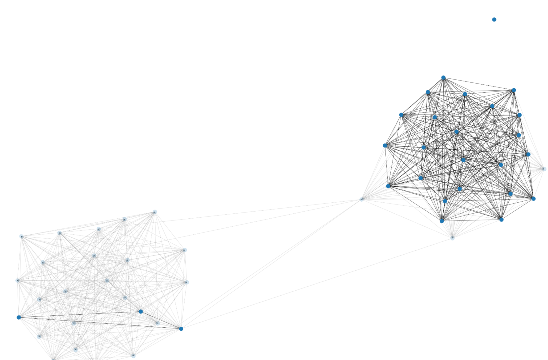
Service Unit Operators, Oil and Gas



Petroleum Pump System Operators, Refinery Operators, and Gaugers



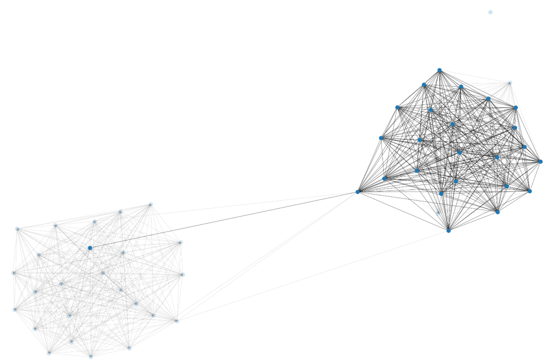
Petroleum Engineers



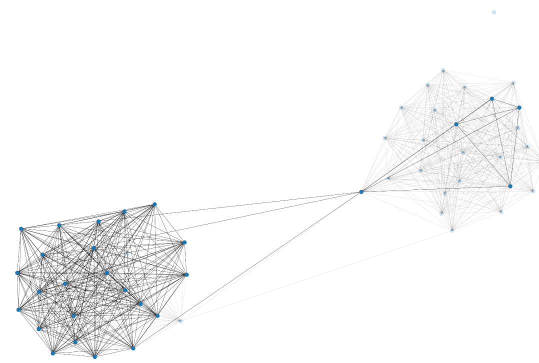
Wellhead Pumpers


Pump Operators, Except Wellhead Pumpers
Energy Engineers, Except Wind and Solar
Gas Plant Operators
Crushing, Grinding, and Polishing Machine Setters, Operators, and Tenders
Gas Compressor and Gas Pumping Station Operators
Continuous Mining Machine Operators

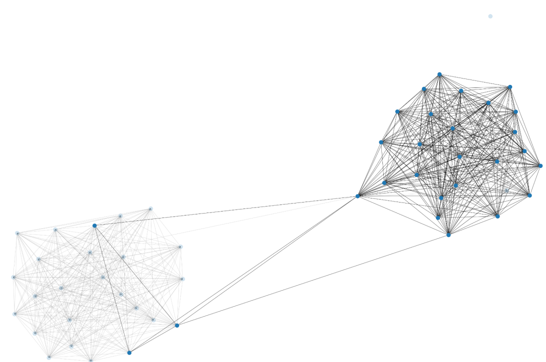
Loading and Moving Machine Operators, Underground Mining



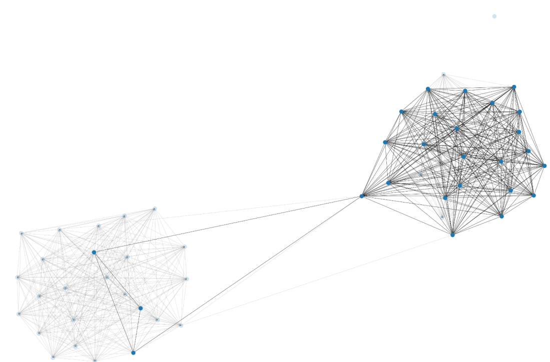
Geothermal Production Managers



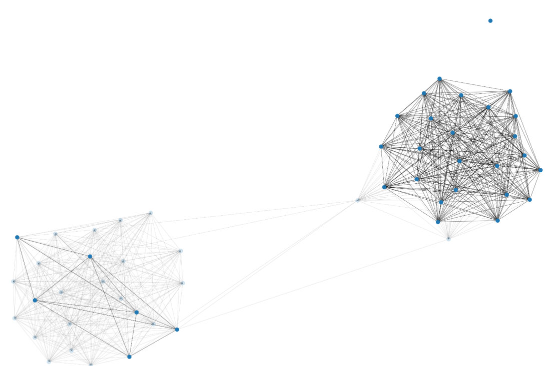
Sailors and Marine Oilers



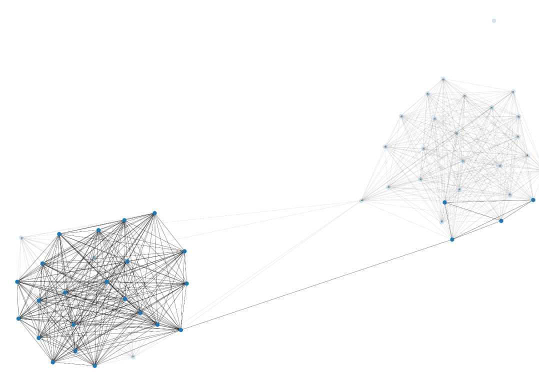
Biomass Plant Technicians



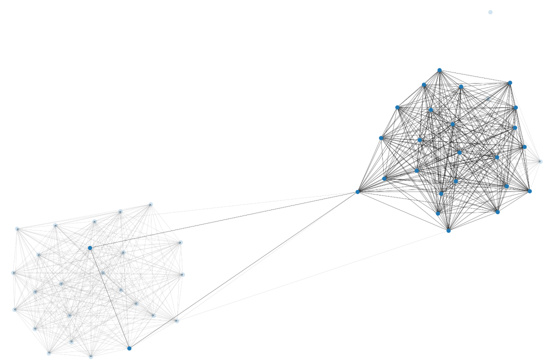
Geothermal Technicians



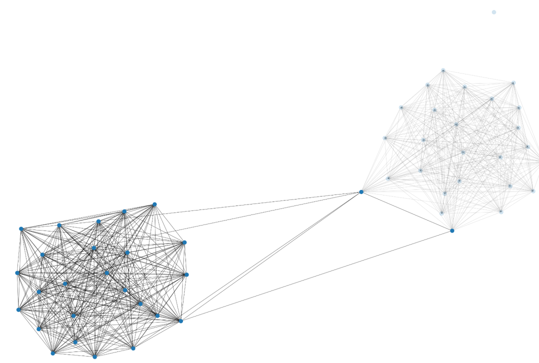
Biomass Power Plant Managers



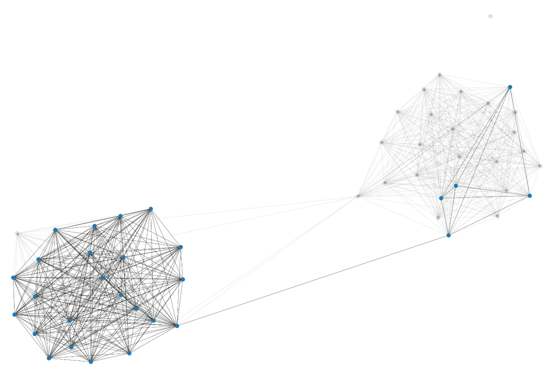
Biofuels Processing Technicians



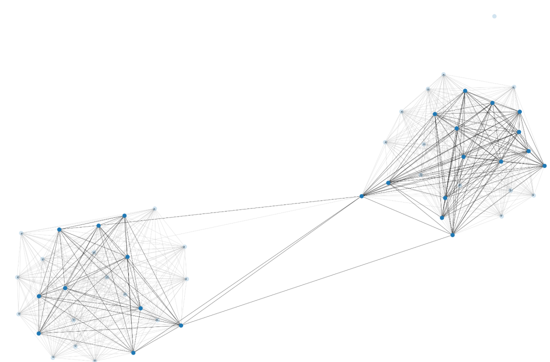
Nuclear Engineers



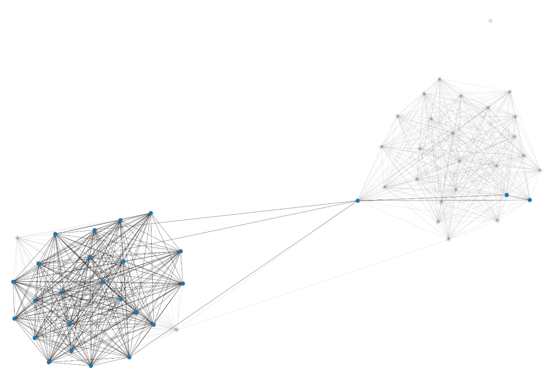
Biofuels Production Managers



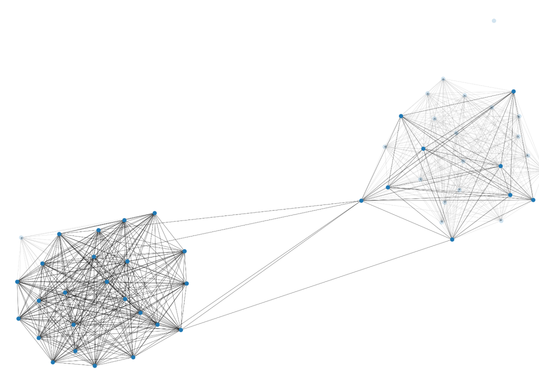
Nuclear Technicians



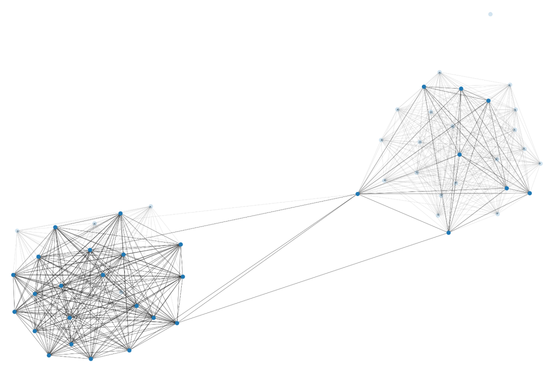
Biofuels/Biodiesel Technology and Product Development Managers



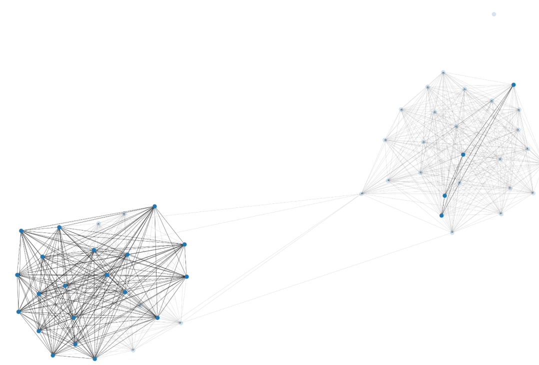
Nuclear Power Reactor Operators



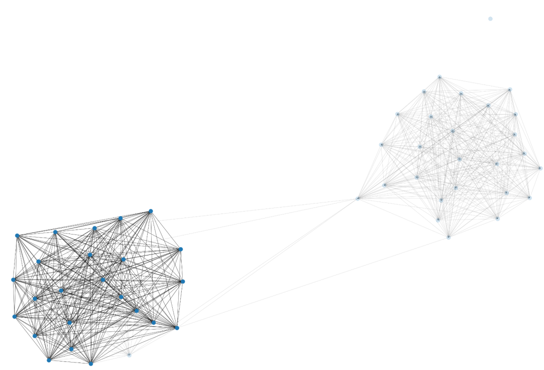
Nuclear Monitoring Technicians



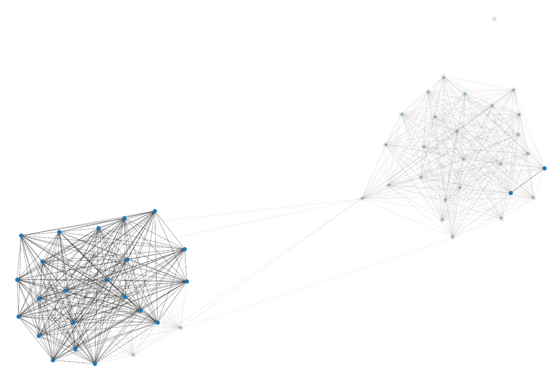
Solar Energy Systems Engineers



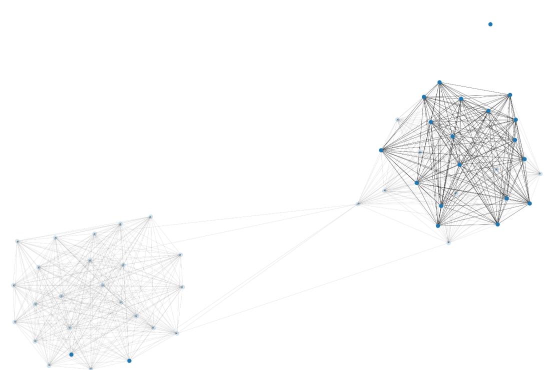
Environmental Economists



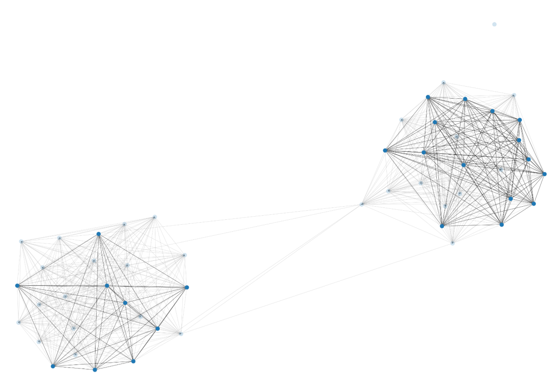
Solar Sales Representatives and Assessors



Solar Thermal Installers and Technicians

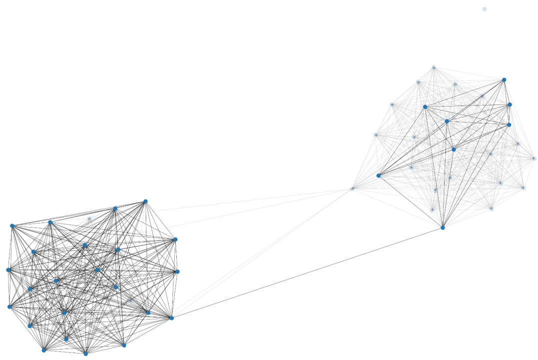


Solar Energy Installation Managers




Solar Photovoltaic Installers
Wind Energy Engineers
Wind Energy Operations Managers
Wind Turbine Service Technicians
Wind Energy Development Managers
Sustainability Specialists

Hydroelectric Production Managers



Hydroelectric Plant Technicians