Detecting Communities in the Fossil Fuel Supply Chain

Aaron Vitarana

Host Supervisor: Lucas Kruitwagen

This Summer, I interned (remotely) with the Economics of Sustainability group at the Institute of New Economic Thinking, University of Oxford. In a project conceived by my supervisor, a model of the fossil fuel supply chain had been created from publicly available data into a network structure and, broadly speaking, I was tasked with finding out whether meaningful information could be extracted from this model. Although this project seems unrelated to physics, many of the methods of obtaining information from a network are derived from concepts in statistical physics and many of the associated techniques had been devised by physicists.

Having had my Summer term cut short by COVID-19 and anticipating the difficulty of undertaking a remote internship, I did a lot of pre-reading in the few weeks before my internship officially started. During this time, I mainly read academic papers written by the group and reports on global usage of the three types of fossil fuel: coal, oil and gas. The overarching aim of the group is to find 'sensitive intervention points' that will mitigate climate change and its associated risks. Sensitive intervention points can be understood as points, in a wide-variety of areas, where a small change will lead to a large impact (astute readers may notice the similarity between this idea to the butterfly effect and more generally chaos theory). In the case of this project, through understanding the cost of transporting fossil fuels from where they sourced, i.e. coal mines, oil fields etc., to where they are used, i.e. power stations, cities etc., by railways, shipping and pipelines, regions that are more susceptible to transitioning to non-carbon-based energy sources may be identifiable. Understanding this on a global scale could guide future policy decisions as we transition to a post carbon-economy.

I focused my attention on the coal supply chain and more specifically finding communities within the network created by minimising the cost of transport. A heavily simplified version of this network can be understood from Figure 1, showing the 'flow' of coal from the coal mine to where it is used by railways and (possibly) shipping routes. The complex coal network that was produced as a result of minimising the cost of transporting coal is shown in Figure 2, with the thickness of connection indicating the quantity of coal transported along the connection. Communities in this network can be understood as pieces of infrastructure that can be considered as operating in conjunction with each other. To the human eye, it is often quite clear where communities have been formed, but for a very large network and for more difficult to discern communities, an algorithm should be used to determine where communities are formed.

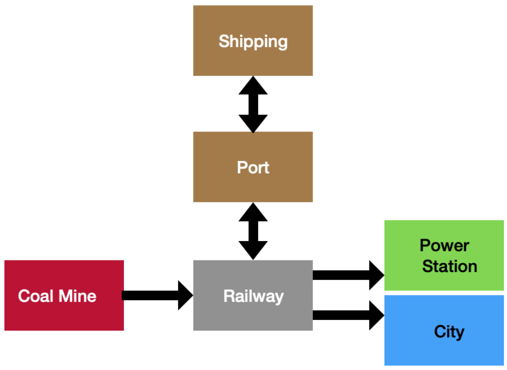


Figure 1: Simplified Coal Network

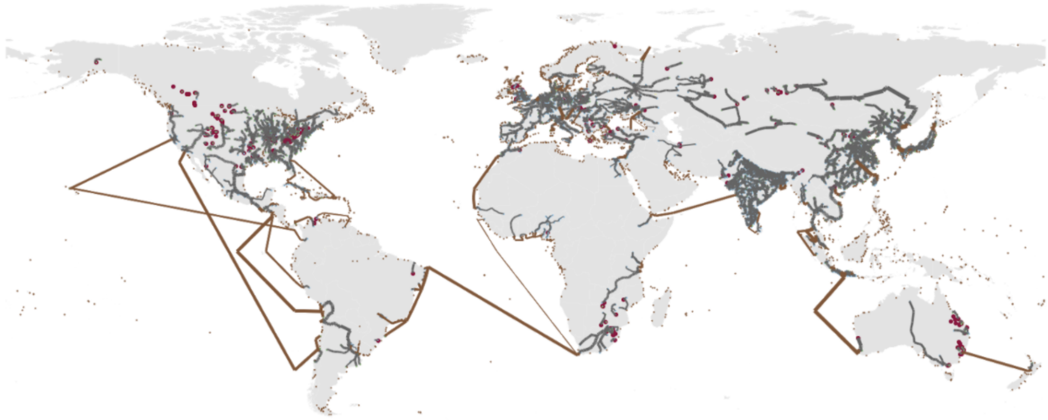


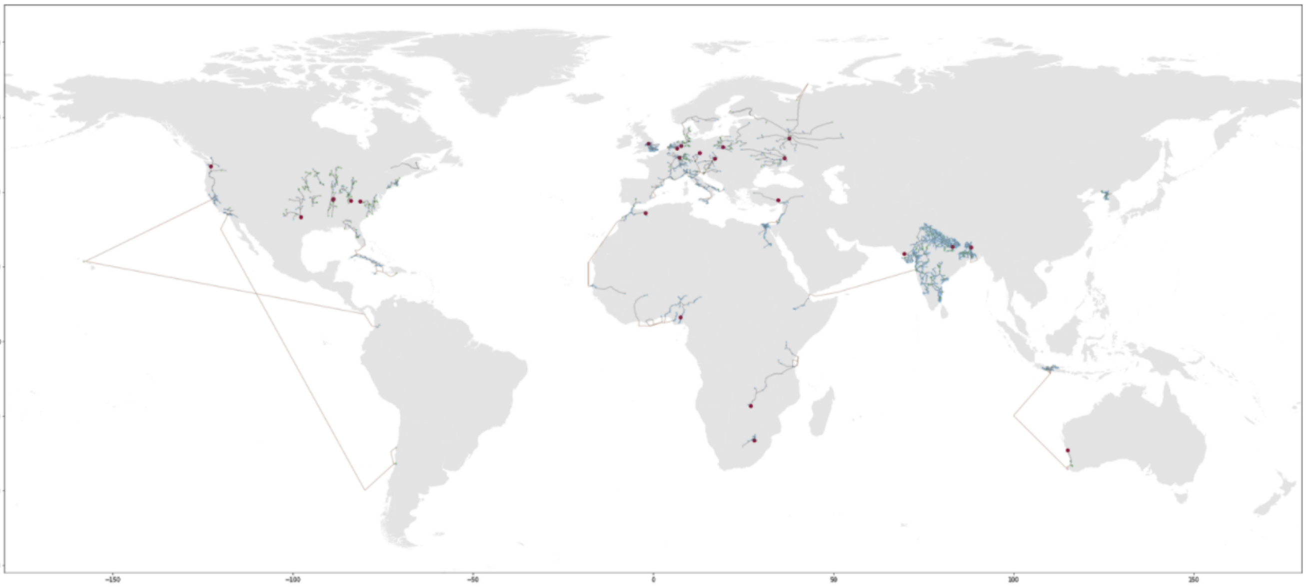
Figure 2: Simulated Coal Network. The amount of coal transported is indicated by the thickness of the connection between pieces of infrastructure. Types of infrastructure: shipping routes (brown edges); railways (grey edges); ports (brown nodes); coal mines (burgundy nodes); cities (sky blue nodes) and power stations (lime green nodes)

There are a wide range of 'community detection' algorithms available and how useful they are is largely dependent on the use case. For this specific graph, there were three key features that I required for the algorithm that I would use: the ability to handle a large graph; the ability to handle a weighted graph and the ability to handle a directed graph. The network structure that had been created contained over 100,000 connected pieces of infrastructure and understanding how much time the algorithm would take to run was a factor that must be considered. In addition, different amounts of coal were being transported through each connection, so the algorithm should take this into account (by favouring communities to be formed between pieces of infrastructure where a large amount of coal was being transported between). Finally, from Figure 1, it is clear that coal could only flow from a coal mine to a city/ power station and not the other way around; the idea of direction is often ignored when finding communities and so this was the most difficult criteria to fill in finding an appropriate algorithm. After reading multiple papers and consulting professors in this field via email, I decided that the Directed Louvain algorithm was the most suitable.

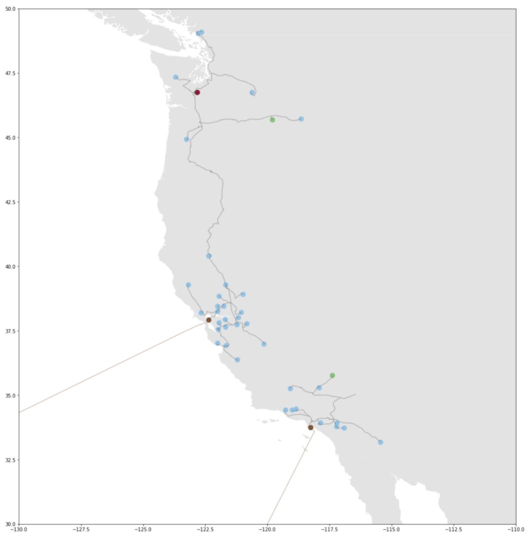
Using this algorithm, separated the graph into communities of different sizes (see Figure 3 and 4). In the short-term, these communities can be used to refine the network in terms of both the data that is used and the method of creating the graph. In the long-term, once improvements to the network have been made, the communities found may highlight regions to target for intervention.

My overall experience of interning was overwhelmingly positive and it was a valuable insight into research. Spending a large proportion of my time coding in Python, my technical skills have improved dramatically. A large proportion of my time was spend doing 'data science', since I needed to manipulate the data into a form that could be analysed; I found visualising the analysis that I had been done to be harder than expected. Working remotely was an interesting experience and I feel this was a shame, since I was not able to gain the full experience of research in a group and my internship was very close to turning into being non-remote in the final couple of weeks. However, I received ample guidance by my supervisor through Slack and video calls and I was also able to jointly present the work I did to other members of the group. In the future, when possible, I hope to meet the group in-person and possibly continue the research further. I had also expected that I would struggle more than I did to complete the project without being at the University. I really enjoyed the academic freedom that I was allowed during this project and the self-teaching that I was required to do. The flexibility of working hours and not requiring to commute was great. I was able to establish a healthy work routine and was more productive overall than if I was required to work strictly from 9-5. One thing that I was also glad to see in this internship is how widely applicable ideas from physics are and how the mathematical and technical skills that I have gained from my degree can be used to help solve real-world problems.

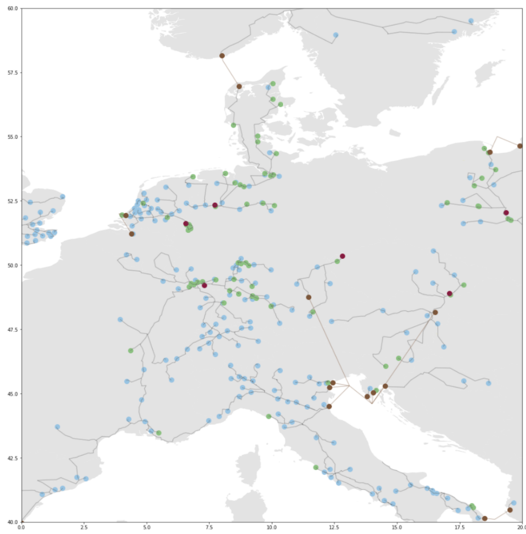
Word Count: 1009



(a) Largest 50 Communities



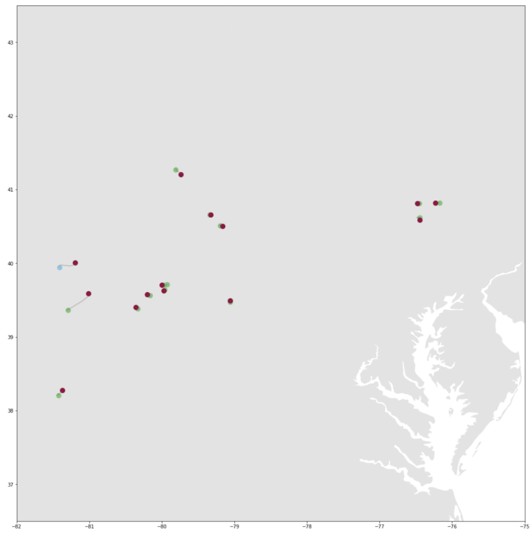
(b) Large Communities in USA



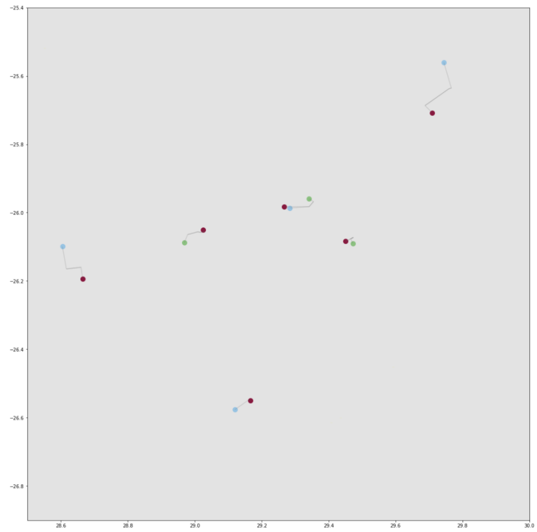
(c) Large Communities in Europe

Figure 3: Largest Communities detected by the Directed Louvain Algorithm

(a) Smallest 50 Communities



(b) Small Communities in USA



(c) Small Communities in South Africa

Figure 4: Smallest Communities detected by the Directed Louvain Algorithm