

73-449: Social, Economic, and Information Networks — Final Project Proposal

Anderson, Section A

Michael Rosenberg

mmrosenb@andrew.cmu.edu

For my final project, I will be doing an analysis of the Offshore Leaks Database. This database shows relationships and networks among people and companies in offshore entities that are traditionally designated as “tax havens”. Thus, to some degree, this database provides strong insights into how the world’s rich conduct offshore business. This database was constructed from three leaks: the International Consortium of Investigative Journalists’ (ICIJ) China Leaks investigation in 2013, the Panama Papers leak in April 2016, and the Bahamas Leak in September 2016.

Due to the size of the original network, I decided to start out my analysis by only looking at behaviors within certain tax havens and certain Eastern European countries. The visualization I have provided (see Figure 1) Contains all agents associated with Russia, Ukraine, Poland, Panama, and The Bahamas and all relationships among those agents.

This portion of the dataset contains 51336 nodes (agents) and 39120 edges (relationships). The distribution of agents across country features 47.66% from Panama, 31.48% from Russia, 17.06% from the Bahamas, 2.91% from Ukraine, and 0.88% from Poland. When we look at the types of agents featured in this subnetwork, 49.63% of agents were entities (company created in a tax haven), 27.33% (person who plays a role in an entity), 20.35% were addresses (locations associated with agents), and 2.69% were intermediaries (go-betweens for entities and officers). Interestingly, despite the very small portion of nodes being intermediaries, they play an extremely key role in terms of degree in the network (see Figure 1). We see that the average degree in this network is 1.524, which suggests that this is a relatively sparsely connected network. There are 16666 weakly connected components in the network, which suggests to me that to do further meaningful analyses on this database, I will need to study all countries within the network due to the apparent fragmentation that comes from studying only a couple of countries. The clustering in this network is also relatively weak, as the average clustering coefficient is 0.015. That being said, the network has an average path length of 2.088 in the largest connected component, which suggests a potential small-world effect in the areas of the network that are not severely fragmented.

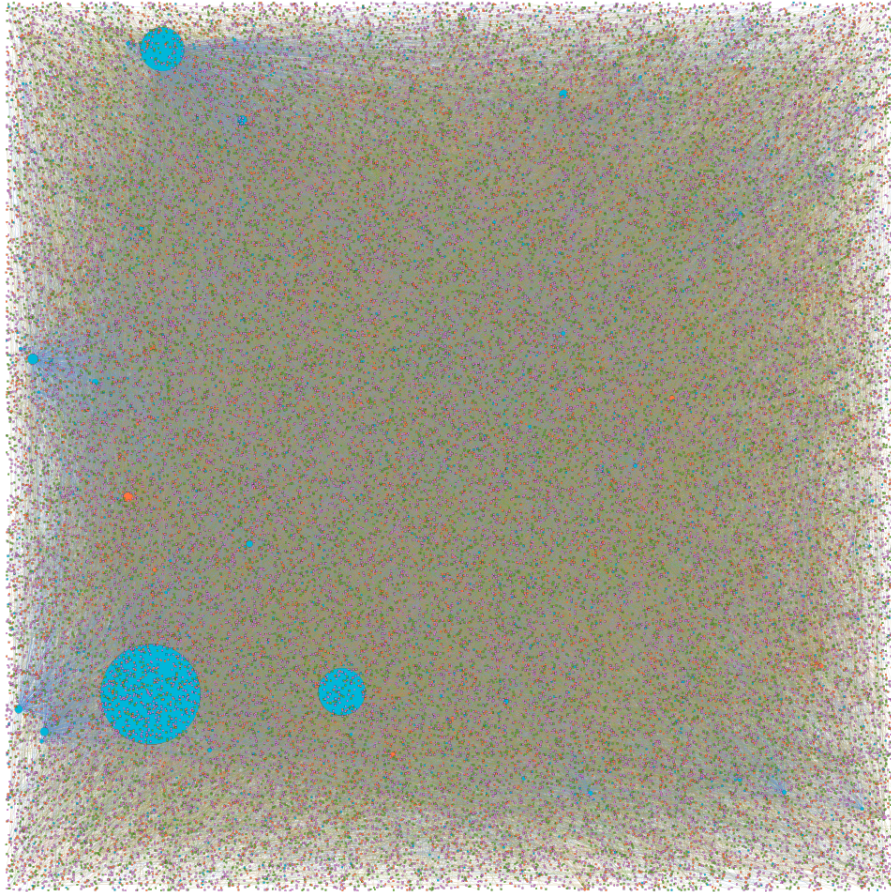


Figure 1: Graph of our Eastern Europe network. The nodes are colored by agent type, where pink is for entities, green is for officers, orange is for addresses, and light blue is for intermediaries. Nodes are sized by degree.