Application of Social Network Analysis in the Estimation of Bank Financial Strength During the Financial Crisis

Michelle Morales[1], **David Guy Brizan**[2], **Hussein Ghaly**[1],
Thomas Hauner[3], **Min Ma**[2], **Syed Reza**[2], **Andrew Rosenberg**[2]
[1]Department of Linguistics, CUNY Graduate Center, USA
[2]Department of Computer Science, CUNY Graduate Center, USA
[3]Department of Economics, CUNY Graduate Center, USA
{dbrizan,mma,thauner}@gc.cuny.edu,
{hmghaly,ms.morales3,s.f.reza}@gmail.com, andrew@cs.qc.cuny.edu

Abstract

The Financial Crisis of 2007-2008 was a very complex and impactful global event. The goal of this research is to explore the possibility of using a bank's social relations to estimate a bank's financial strength. We apply Natural Language Processing techniques to a corpus of financial data released by the NLP Unshared Task in PoliInformatics in 2014 in order to explore and better understand this possibility. Our work begins with the extraction of named entities from the corpus to establish names of people involved in the crisis. We then aggregate the social histories of these individuals from an online collaborative knowledge base: Freebase. Accordingly, we use the social histories of entities to establish social connections between them. We end with a visualization of the connections we found: a presentation of a social financial crisis network.

1 Introduction

The financial crisis represents the near collapse of the financial system in the United States, where financial institutions were under threat of collapse, banks were bailed out by national governments, and stock markets plummeted around the world. It has been shown that various paticipants shared responsibility for creating or enabling the crisis. Mortgage lenders, borrowers, regulators, investors, rating agencies, and many others have been blamed (Bolton, 2009). At the center of the crisis were the financial institutions, and we focus on these key players. Specifically, we aim to discover how the strength of these players affected the outcome of the crisis. Whereas traditional financial models consider a bank's credit relations and financial strength using balance sheets, we explore using a bank's social relations to estimate a bank's financial strength in the aftermath of the financial crisis. Specifically, we measure the financial strength of a bank by the amount of bailout money and emergency lending to which it had access.

This work is part of the NLP Unshared Task in PoliInformatics 2014. The corpus includes reports, hearings, bills, and other transcripts related to the crisis and was provided by the organizers of task, a detailed breakdown of the data can be viewed at `https://sites.google.com/site/unsharedtask2014`. Our research is motivated by 2 questions (1) who was the financial crisis? (2) and, what was the financial crisis? We approach this task from the perspective of social network analysis (SNA) because it allows us to effectively analyze the interdependence and flows of influence among individuals, groups, and institutions. Social Network Analysis involves the investigation of networks, which are made of social entities linked by specific types of interdependencies. This analysis takes social connections as the primary building blocks of the social world (Pinheiro, 2011). Besides individual attributes, SNA considers all information about the relationships, represented as links, among the network members, represented as nodes. The information about the relations among the individuals within a social network is usually more relevant than the individual attributes of the individuals. Identifying nodes and links is key to whether a social network analysis will prove informative.

While there is a rich history of literature considering geographic networks and topological space, in economic analysis (Anselin 1998 ,Kelejian 1998, and Fingleton 2014) only a few applications consider a social network or social space independent of geography–despite the routine acknowledgement of the possibility (Hsieh, 2012). All rely on the first law of geography from Tobler (1970): "Everything is related to everything else, but near things are more related than distant things." However, social interactions also have an economic effect independent of spatial effects. Acemoglu et al. (2014) consider the scheduled interactions, social connections, and geographic proximity of banks in relation to Timothy Geithner, hypothesizing that banks socially closer to Geithner experienced higher stock returns in response to his nomination as Secretary of the Treasury. In the days after Geithner's nomination, they estimate a cumulative abnormal return of 15% for those banks connected to him. In the same vein, we also consider social connections in our analysis of the financial crisis. We begin in section 2 with a discussion of related work, section 3 details our data, section 4 explains our methods, section 5 describes our data visualization, in section 6 we sumamrize conclusions and results, and we end in section 7 with future work.

2 Related Work

Social network analysis has useful and practical interdisciplinary applications. When applied to telecommunications, SNA can help companies recognize the behavior of their customers and then predict the strength of links between customers and the possible impact of events among them [5]. A SNA perspective has also been employed to understand political, economic and social organizations and individuals (Christopoulos 2008). For example, it has been a useful approach for the study of terrorism and related fields. In the study of political violence SNA helped provide important information about the characteristics of a terrorist group structure, showing how a social structure influences members motives, behaviors, and the outcomes of their actions. Using SNA, Perliger (2011) provided insight into recruitment processes, evolution, and the division of political and social power among the members of these terrorist groups. The network perspective has very productive sub-fields within historical anthropology, development studies, epidemiology and other fields.

3 Data

We constrcut our social network around the Financial Crisis event of 2007-2008. Each member (node) in our network represents an individual or orgnaization involved in the crisis. Each connection (link) between members is established based on a social connection. Our data was provided to us as part of the NLP Unshared Task in PoliInformatics 2014, organized by the PoliInformatics Research Coordination Network. The data set includes Federal Open Market Committee (FOMC) meeting minutes and transcripts, Federal Crisis Inquiry Commission reports and transcripts,and Congressional bills, reports, and hearings. We also include data aggregated from Freebase, an online knowledge database (`http://www.freebase.com/`). We imported 2267 entities, including 954 people and 1207 organizations, most of which (77.5%) are financial institutions involved in the financial crisis.

We rely on two public journalistic sources for data on bank bailouts. For data on the TARP bailout we utilize the ProPublica *Bailout Tracker* (Kiel and Nguyen, 2014), an updated database of outgoing and incoming bailout funds (we exclude bailouts of the housing agencies Fannie Mae and Freddie Mac as they represent government sponsored entities). Data on the peak amount of bank debt through the Federal Reserve's myriad secret emergency lending facilities is available through Bloomberg (Friedman et. al 2014). Their reporting is based on the source documents of the Government Accountability Office's onetime audit of the Federal Reserve as legislated by DoddFrank Wall Street Reform and Consumer Protection Act.

4 Method

4.1 Named entity recognition

Named entities (NEs), especially person names (PER), location names (LOC), and organization names (ORG), deliver essential context and meaning in human languages (Chen et al. 2013). To investigate the behaviors of the NEs involved in the financial crisis and the relations between them, the first step is to extract the NEs from the datasets. We split all the transcripts in the Congressional Hearings and the Congressional Reports into seperate files by speaker to gain a better understanding of the personal concerns of NEs, one file per speaker. A named entity tagger (Li et al., 2012) based on the conditional random field (CRF) model is employed to highlight all the NEs in each file.

4.2 Relation extraction

We then aim to establish relations between two entities using the unstructured text data. To extract these relations we treat each sentence as an event, and target specifically the action in the event. Therefore, we focus on the verbs as they convey the action in the sentence. We explore the pair relationships between two NEs recognized in the previous stage, which enables us to determine the entities mentioned within the text of each statement by each speaker, as in the following example, NEs are marked by brackets: [Mr. Lockhart]: "The goal of these dual conservatorships is to help restore confidence in [Fannie Mae] and [Freedie Mac]."

Using simple regular expression functions, we are able to extract the intervening text between each pair of entities. The basic assumption here is that this intervening text would tell us something about the relationship between the two entities, as is the case:

Table 1: *Example Phrase*

Entity1	Intervening Text	Entity2
Fannie Mae	and	Freddie Mac

This indicates some degree of association between the two entities Fannie Mae and Freddie Mac, being mentioned together in many contexts and separated by "and". However, for larger intervening text chunks, it would be difficult to infer such relationships, as in the following:

Table 2: *Example Sentence*

Entity1	Intervening Text	Entity2
Mr. Lockhart	The goal of these dual conservatorships is to help restore confidence in	Fannie Mae

We then use Part of Speech Tagging (as implemented in the NLTK POS tagger), in order to identify the verbs within the sentence and the intervening text. We also consider verb frequency and ignore extremely frequent verbs, such as, "is" and "has". For the case of negation in text, we add "not" before the verb. We identify negation by searching for the trigger words "not" or "never". Our goal is to we reduce the intervening text to a single verb that indicates the relationship between the two entities.

Table 3: *Example Relation*

Entity1	Intervening Text	Entity2
Mr. Lockhart	restore	Fannie Mae

Although we find many good examples as is shown in Table 3, due to time constraints this task needs further development, before the events extracted from text can be included into our social network.

4.3 Aggregation of Freebase data

Given a list of extracted individuals from the task's data, we query the Freebase database (March, 2014) for a list of attributes. We focus on the attributes that would provide insight into the social relations of that person. Table 4 demonstrates the properties we consider for a person. For certain attributes we consider temporal information, such as the time frame employed by a company. Those relations are marked with an asterisk (*).

Table 4: *Freebase Relations*

profession
employment history*
education*
government positions held*
political party
organization member
shareholder
military service

For NEs that are organizations we only query Freebase for a list of employees, specifically board members. For this query we also consider the time frame these employees served as board members, noting board members appointed before or during the crisis as more relevant. All extracted data was entered into our SQL database.

5 Data Visualization

Lastly, we derive a visual presentation of our social network. Each entity from our database represents one node while links represent social relations between entities. The figure shows one subnetwork pulled for the larger network. Our online demo demonstrates the interactive version of our network where labels can be queried and observed.

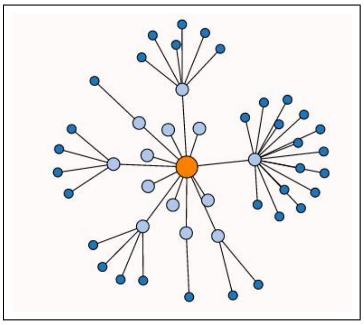

Figure 1: Social Network of Citigroup

The Citigroup Social Network demonstrates one small subnetwork. Citigroup represents the most central and largest node. The first connections we consider are board members of the organization (second largest nodes) appointed before/during the time of the crisis. Next, we establish social relations between board members and people from the task data (smallest nodes) by examining the social histories of the individuals, considering events like employment history, schools attended, etc. For each individual we consider the list of Freebase attributes extracted. For each individual, each attribute gets checked against the rest of the individuals, when a match is found, a social connection is made. In this small subnetwork we remove intervening institutions to simplify. However, in the larger we can see every individual, every financial institution, and every intervening financial institution (e.g. schools, employers,etc.) This network represents some of our findings. A detailed description and interactive visualization of both the subnetwork, as well as, our largest and most detailed network can be seen in our web demo: `http://lynx.cs.qc.cuny.edu/cunyfinancialcrisis/` Our preliminary results show that most social relations were established between individuals based on their educational institution. Due to the size of these institutions, and the unlikelihood of a relation forming based on attendance, moving forward we will fine tune which relations are considered in our network. Most importantly, we will continue to work on extracting relations from our unstructured data to incorporate into our network.

6 Conclusions

Although our work is still in its developing stages, it shows promise. Thus far, we have extracted relevant entities involved in the financial crisis from our corpus. We have also aggregated structured data from Freebase. From the collection of these data, we have then created social histories of all the individuals involved in the crisis. By analyzing each individual's history we then established social relations between entities and consequently constructed a visualization of a financial crisis social network. We also began exploring the use of unstructured data to establish relations between entities,and we will continue along this avenue as we believe it will help to establish more relevant relations between individuals.

7 Future Work

Moving forward, we will continue to focus on refining our analysis of unstructured data, using dependency parsing and semantic role labeling, to eventually incorporate these relations into our network. Future work will also include a comprehensive analysis of our network using a spatial econometric model. The underlying premise of a spatial econometric model assumes spatial dependence between observations. In our application, we would then consider spatial dependence through social topology, as derived from our social network analysis. Social interactions would then enter our economic model through a social spatial weights matrix, which would take into account the strength of each possible dyadic social interaction between observational units. Our economic model would consider the population of financial entities in the data, or banks, who received some financial rescue. By counting the total number of social connections between banks we can define the social connetion between banks. We hypothesize then that a bank's bailouts during the financial crisis were impacted by the average bailouts of banks it was socially connected to.

In conclusion, preliminary work shows promise that a social network constructed from the aggregation of data can be a useful approach to study the people and institutions involved in a global financial event, especially highlighting the interdependencies between them. Our work supports further exploration of the possibility of using a bank's social relations to estimate a bank's financial strength (Tobler, 1970).

References

Acemoglu, Daron and Johnson, Simon and Kermani, Amir and Kwak, James and Mitton, Todd. 2014. *The value of connections in turbulent times: Evidence from the united states*. National Bureau of Economic Research.

Anselin, Luc. 1988. *Spatial econometrics: methods and models*, volume 4.

Anselin, Luc and Bera, Anil K. 1998. *Spatial dependence in linear regression models with an introduction to spatial econometrics*. Statistics Textbooks and Monographs,volume 155, pp.237-290.

Bolton, Brian J., The U.S. Financial Crisis: A Summary of Causes and Consequences (October 21, 2009).

Christopoulos, Dimitris. 2013. *Network Analysis Intro 6/6 dc.* christopoulos@uwe.ac.uk

Fingleton, Bernard and Le Gallo, Julie. 2008. *Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties*. Papers in Regional Science, volume 87, pp.319-339.

Friedman, Robert and John Voskuhl and Otis Bilodeau. April 2014. *The Fed's Secret Liquidity Lifelines*. http://bloom.bg/o6GTTL

Hsieh, Chih-Sheng and Lee, Lung-Fei. 2012. *A Structural Modeling Approach for Network Formation and Social Interactions with Applications to Students' Friendship Choices and Selectivity on Unobservables*. Working Paper. Ohio State University

Kelejian, Harry H and Prucha, Ingmar R. 1998. *A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances*. The Journal of Real Estate Finance and Economics,volume 17, pp.99-121.

Paul Kiel and Dan Nguyen. April 10, 2014. *Bailout Tracker*. http://projects.propublica.org/bailout/

LeSage, James and Pace, Robert Kelley. 2008. *Introduction to Spatial Econometrics*. CRC press.

Qi, Li et al. 2012. *Joint bilingual name tagging for parallel corpora* Proceedings of the 21st ACM international conference on Information and knowledge management. ACM.

Peter Mika, Flink. 2005. *Semantic Web technology for the extraction and analysis of social networks, Web Semantics: Science, Services and Agents on the World Wide Web* Volume 3, Issues 23,Pages 211-223, ISSN 1570-8268.

ACL Workshop.2014.*NLP Unshared Task in PoliInformatics*.https://sites.google.com/site/unsharedtask2014/

Perliger, Arie, and Ami Pedahzur. 2011. *Social network analysis in the study of terrorism and political violence*. Political Science and Politics, 45-50.

Pinheiro, Carlos A.R. 2011. *Social Network Analysis in Telecommunications.* John Wiley & Sons. p. 4. ISBN 978-1-118-01094-5.

Chen, Yufeng, Chengqing Zong, and Keh-Yih Su. 2013. *A joint model to identify and align bilingual named entities*. Computational Linguistics 39.2: 229-266.

Tobler, Waldo R.1970.*A Computer Movie Simulating Urban Growth in the Detroit Region*. Economic Geography, pp.234-240.