

# Financial Distress of U.S. Banking Industry Viewed through Web Data

Esteban Romero-Frías<sup>1</sup> and Liwen Vaughan<sup>2</sup>

<sup>1</sup> *erf@ugr.es*

Department of Accounting and Finance, University of Granada, Paseo Cartuja s/n, CP 18071, Granada (Spain)

<sup>2</sup> *lvaughan@uwo.ca*

Faculty of Information and Media Studies, University of Western Ontario, London, Ontario, N6A 5B7 (Canada)

## Abstract

Building on previous studies on inlink and co-link research on commercial websites, the current study attempted to apply and combine both methods to investigate the recent financial crisis in the U.S. banking industry. The methods combine Web content mining and Web structure mining. Two sets of inlink and co-link data were collected, one with the keywords “crisis”, “bailout” and “subprime” and one without any keyword. The one with the keywords was meant to include only inlink webpages that are likely to be about the financial crisis. The number of inlink pages that contained the keywords correlates significantly with the degree of a bank’s crisis measured by the amount of government bailout money. Both sets of co-link data were analysed using multidimensional scaling to generate maps of business competition. The comparison between the maps with and without keywords shows that the map generated from the co-link data with keywords depicts a more accurate image of the industry by clustering banks with more financial problems together.

## Introduction

This paper reports an ongoing study that explores the possibility of using Web data to discover timely business information. Many economic and financial data such as revenue and profit are not timely as they are calculated quarterly or annually. In contrast, the Web is constantly changing. The Web data thus have the potential to provide more timely information. Unlike the established economic and financial variables, however, the meanings of usefulness of Web data are not very clear. Many studies have examined various types of Web data in terms of their relationship with economic and financial data. For example, Tumarkin & Whitelaw (2001) studied the relationship between Internet message board activity and abnormal stock returns and trading volume. Using social network analysis, Das and Sisk (2005) analyzed messages posted to stock boards. Jin, Matsuo & Ishizuka (2009) also used social network analysis but collected data from the general Web to rank companies. Findings from these studies contributed to our understanding on how to use Web data to gain business information but many more studies are needed for a clearer and firmer understanding. Toward this end, we are studying how Web data can be used to analyze the recent crisis of the U.S. banking industry. Specifically, we try to find out if the inlink and keyword data can be a measure of the degree of crisis, if the data can be used to visualize the industry in that banks with more distress are grouped together. We also collected various financial and economic data to triangulate the findings from the Web data. The research is currently in progress. This paper reports preliminary findings.

Previous research in Webometrics has analyzed commercial websites in order to explore new sources of business information that could be useful for data mining and business intelligence purposes (Thelwall, Vaughan & Björneborn, 2005). Link analysis has been used specifically for competitive intelligence (Chau, Shiu, Chan & Chen, 2007; Reid, 2003).

A recent study (Vaughan & You, 2008) proposed a method that combines page content with co-link data to achieve a more detailed picture of the competitive landscape of a sector within an industry. We apply this content assisted method to the banking industry in order to study the impact of the current financial and economic crisis. The keywords selected to refine the search are “crisis”, “bailout” and “subprime”, which are used frequently to describe the current worldwide financial problems. These keywords are intended to filter out pages that

link the banks, whether they are inlinks or co-inlinks, for reasons other than the financial crisis. From a methodological point of view, this method provides new possibilities to use Webometric techniques in the field of commercial websites to research specific issues. We supported our co-link analysis with correlation test between inlinks and data on the degree of financial crisis. Definitions of inlink and co-link follows that by Björneborn & Ingwersen (2004).

## **Methodology**

The bank sector was selected because it is at the origin of the financial crisis worldwide, specially in the United States. The crisis in this industry has shown particular characteristics, coming from the subprime mortgage crisis to the government rescue plans to save the industry. Moreover, statistics about the use of the Web by companies usually rank financial industry at the top, only below the information technology industry, which has been already studied. These features make the banking industry an appropriate and relevant subject to combine the aforementioned Webometric techniques.

Our first attempt to study the crisis focused on a set of international top banks. However, this approach faced some limitations, such as the different home languages of the companies that could bias the collection of data using specific keywords. The varied performance of national economies was also a significant deterrent because the heterogeneous conditions made it more complicated to analyse the impact of financial crisis in the companies. Many countries have also established rescue plans to support banks, but the conditions applied for the allocation of funds and the amounts were not comparable and these data were not publicly available for all countries.

Therefore, we decided to focus in the U.S. banking sector and more specifically in the banks listed in the New York Stock Exchange (NYSE). In the U.S., the Federal Government designed an unprecedented and expensive rescue plan to buy financial assets of the banks in order to protect a general bankruptcy of the financial system. In order to have more homogeneous data and to select a collection of major U.S. banks, we took the 46 companies included in the NYSE ([www.nyse.com](http://www.nyse.com)). These companies are listed on NYSE webpage under the following labels: Industry-Financials, Supersector-Banks, Sector-Banks, and Subsector-Banks. Only U.S. national companies with common stock traded were taken into account (the list of companies was taken on 16<sup>th</sup> January 2009, see Appendix 1). Additionally, Wachovia Corporation, that was not in the list due to the recent merge with Wells Fargo, was included.

Different measures of the impact of the financial crisis in the banks were considered, e.g. the change in the banks' stock prices over a one-year period. After exploring various measures with publically available data, we decided that the most reliable measure of a bank's of financial distress is the amounts of money it received under the Troubled Asset Relief Program (TARP) established by the U.S. Federal Government. Data were collected from a Special Report by CNN Money (2009). Only 22 out of the 47 banks in the study had received money under the TARP.

Yahoo! is used for data collection as the other two major search engines in the market, Google and Live Search (MSN) do not allow to perform the type of queries shown in Table 1.

“Linkdomain” command searches for webpages that link to all pages of a site, while “link” command searches for pages linking only to a particular URL. Both commands were tested in the study. Data collected with the linkdomain command generated MDS maps that clustered troubled banks closer. Four banks had to be excluded from the co-link analysis because they did not have any co-link with other banks in the study. If we used the “link” command instead of the “linkdomain” command, more banks may have to be excluded because of the lack of

co-links with other banks. Appendix 1 shows a sample of banks included in the study. The complete list of banks is omitted due to the space limitation.

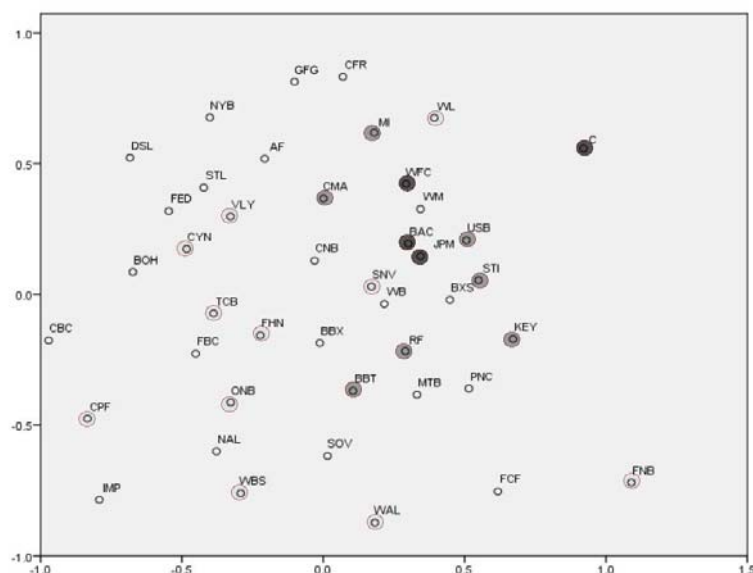
**Table 1. Yahoo! Query Syntax for inlink and co-link data.**

Type of data collected	Yahoo! Query
Inlinks without keywords	linkdomain:boh.com –site:boh.com
Inlinks with keywords	linkdomain:boh.com –site:boh.com (crisis OR bailout OR subprime)
Co-links without keywords	(linkdomain:boh.com -site:boh.com) AND (linkdomain:cnb.com -site:cnb.com)
Co-links with keywords	(linkdomain:boh.com -site:boh.com) AND (linkdomain:cnb.com -site:cnb.com) (crisis OR bailout OR subprime)

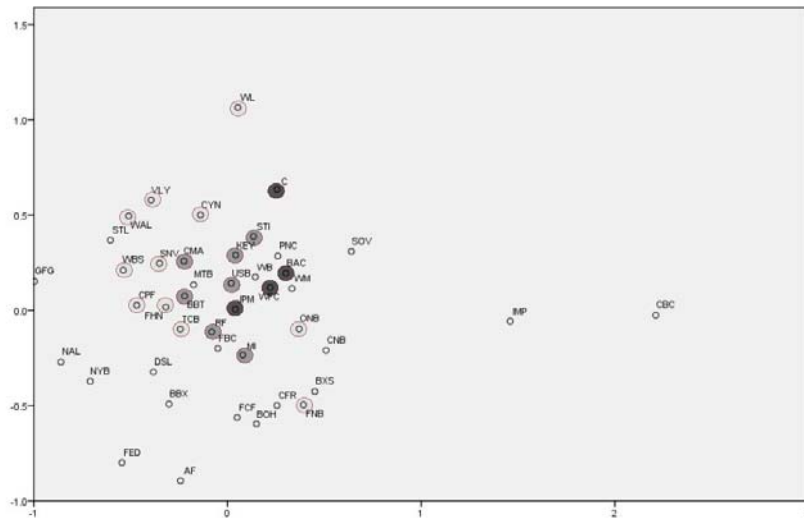
## Preliminary Results / Findings

There are significant ( $p < 0.01$ ) correlations between the number of inlinks to company websites and the amount of bailout money received. Correlation coefficients are 0.72 and 0.78 for the queries without and with keywords respectively. Both correlation coefficients are high but the one with the keywords is higher, suggesting that using the keywords in the search did add some information about the crisis. However, the difference between the two is not very large. Further research is needed to explore how the search can be revised to hopefully retrieve data with more crisis information.

MDS maps obtained from the two sets of co-link data are shown in Figure 1 and Figure 2. Figure 1 was generated from the data without the keywords while Figure 2 was from the data with the keywords. Small circles in black colour are used to mark the banks that received money. The strength of the black colour represents the amounts received.



**Figure 1. MDS map without keywords.**



**Figure 2. MDS map with keywords.**

Four banks in dark black colour received more than ten billion dollars, the seven median coloured banks received between one and ten billion dollars and, finally, banks in light black colour received less than one billion dollars. Comparing the two MDS maps, we can see that map with keywords (Figure 2) clusters the banks in approximately three layers with darker coloured ones in the centre, followed by median coloured in the outer layer and then the light coloured ones near the edge. Assuming that the government bailout money is approximately proportional to the degree of crisis of the affected banks, the map positioned banks according to their crisis level with the more affected ones in the centre and less affected ones on the outside. In contrast, MDS map without keywords (Figure 1) does not present a clear pattern that reflects the degree of crisis (banks more affected by the crisis are distributed in different parts of the map). This suggests that incorporating keywords in the inlink search helps to obtain specific information related to the keyword, in this case, the crisis. It makes sense that more co-links with the keywords would appear among banks that are in a similar degree of crisis because the more a bank is in crisis, the more pages discussing the crisis would link to it, i.e. the bank would receive more inlink pages that contained the crisis related keywords.

## Future work

Currently, we have data on the amount of government bailout money for 22 banks. We will try to find data for other banks through other sources. Alternatively, we will try to find a measure of the degree of crisis for which data are available for more banks. In order to verify the nature of the high correlations reported in this paper, we will collect other financial and accounting data about these banks. Correlations among all these variables will be examined together to gain a more accurate understanding of the correlations. We will also test other possibilities of collecting Web data, e.g. try different keywords or limit search to particular domains such as .org. We also plan to do a qualitative content analysis of inlink pages that contain the keywords to find out if the keywords are used in the context of the financial crisis. Finally, we plan to monitor the development of the financial crisis to find out if results of our data analysis will reflect changes of the crisis; if and how our methodology needs to be modified to better reflect the economic condition.

## Acknowledgments

This study is partly supported by a research grant from the Social Sciences and Humanities Research Council of Canada (SSHRC).

## References

- Björneborn, L. & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, 55(14), 1216-1227.
- Chau, M., Shiu, B., Chan, I. & Chen, H. (2007). Redips: Backlink search and analysis on the Web for business intelligence analysis, *Journal of the American Society for Information Science and Technology*, 58(3), 351-365.
- CNN Money (2009), *Economy rescue: Adding up the dollars*. Retrieved Jan. 21, 2009 from [http://money.cnn.com/news/specials/storysupplement/bailout\\_scorecard/index.html](http://money.cnn.com/news/specials/storysupplement/bailout_scorecard/index.html).
- Das, S. R. & Sisk, J. (2005). Financial communities. *Journal of Portfolio Management*, 31(4), 112-123.
- Jin, Y., Matsuo, Y. & Ishizuka, M. (2009). Ranking companies on the Web using Social Network Mining. In *Ting, I. H. & Wu, H. J. (Eds). Web Mining Application in E-commerce & E-services*. Berlin: Springer-Verlag, pp. 137-151.
- Reid, E. (2003) 'Using Web link analysis to detect and analyze hidden Web communities', in: *Vriens, D. (Ed.). Information and Communications Technology for Competitive Intelligence*, pp 57-84, Hilliard, Ohio: Ideal Group Inc.
- Thelwall, M., Vaughan, L. and Björneborn, L. (2005) 'Webometrics', in Cronin B. (ed.), *Annual review of information science and technology*, Vol. 39, pp 81-135, Medford, NJ: Information today.
- Tumarkin, R. & Whitelaw, R. F. (2001). News or noise? Internet postings and stock prices. *Financial Analysts Journal*, 57(3), 41-51.
- Vaughan, L. and You, J. (2008) 'Content assisted web co-link analysis for competitive intelligence', *Scientometrics*, 77 (3), 433-444.

## Appendix 1. Sample entries of banks included in the study

Company	Labels in the figures	Main URL
Bank of America Corporation	BAC	<a href="http://www.bankofamerica.com">http://www.bankofamerica.com</a>
BB&T Corporation	BBT	<a href="http://www.bbt.com">http://www.bbt.com</a>
Central Pacific Financial Corp.	CPF	<a href="http://www.centralpacificbank.com">http://www.centralpacificbank.com</a>
Citigroup Inc.	C	<a href="http://www.citigroup.com">http://www.citigroup.com</a>
City National Corporation	CYN	<a href="http://www.cnb.com">http://www.cnb.com</a>
Comerica Inc.	CMA	<a href="http://www.comerica.com">http://www.comerica.com</a>
F.N.B. Corporation	FNB	<a href="http://www.fnb-online.com">http://www.fnb-online.com</a>
First Horizon National Corporation	FHN	<a href="http://www.firsthorizon.com">http://www.firsthorizon.com</a>
JPMorgan Chase & Co.	JPM	<a href="http://www.chase.com">http://www.chase.com</a>
KeyCorp	KEY	<a href="http://www.key.com">http://www.key.com</a>
Marshall & Ilsley Corporation	MI	<a href="http://www.mibank.com">http://www.mibank.com</a>
Old National Bancorp	ONB	<a href="http://www.oldnational.com">http://www.oldnational.com</a>
Regions Financial Corporation	RF	<a href="http://www.regions.com">http://www.regions.com</a>
Suntrust Banks Inc.	STI	<a href="http://www.suntrust.com">http://www.suntrust.com</a>
Synovus Financial Corp.	SNV	<a href="http://www.synovus.com">http://www.synovus.com</a>
TCF Financial Corporation	TCB	<a href="http://www.tcfbank.com">http://www.tcfbank.com</a>
US Bancorp	USB	<a href="http://www.usbank.com">http://www.usbank.com</a>