Cohesiveness in Financial News and its Relation to Market Volatility

Matija Piškorec[1], Nino Antulov-Fantulin[1], Petra Kralj-Novak[2]
Igor Mozetič[2], Miha Grčar[2], Irena Vodenska[3]
Tomislav Šmuc[1]

[1]Ruđer Bošković Institute, Zagreb, Croatia
[2]Jožef Stefan Institute, Ljubljana, Slovenia
[3]Metropolitan College, Boston University, USA

September 18, 2014

Part I

Cohesiveness in a corpus of documents

Financial document - an example

BBC | News | Sport | Weather | Capital | Culture | Autos

NEWS BUSINESS

Home | UK | Africa | Asia | Europe | Latin America | Mid-East | US & Canada | Business | Health | Sci/Environm

Market Data | Economy | Entrepreneurship | Business of Sport | Companies | Technology of Business | Kn

27 June 2012 Last updated at 22:31 GMT

Barclays fined for attempts to manipulate Libor rates

Barclays has been fined £290m ($450m) for trying to manipulate a key bank interest rate which influences the cost of loans and mortgages.

Its traders lied to make the bank look more secure during the financial crisis and, sometimes - working with traders at other banks - to make a profit.

Barclays said the actions "fell well short of standards". Chief executive Bob Diamond is to give up his bonus.

PA

Chief executive Bob Diamond will give up his bonus for this year - last year it was £2.7m

The Financial Services Authority is now looking into other banks.

The matter is also being investigated in the US, where the Department of Justice said criminal investigations into "other financial institutions and individuals is ongoing".

Figure: Entities in a news article: institutions (green), financial glossary terms (blue) and negative sentiment words (red).

Cohesiveness in a corpus of documents

Intuition: Large cohesion in a collection of financial news documents indicates a form of *herding effect* that either reflects on important event in the financial markets or can potentially elicit a response on financial market behavior.

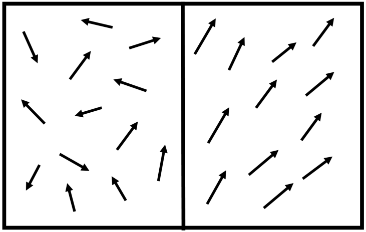
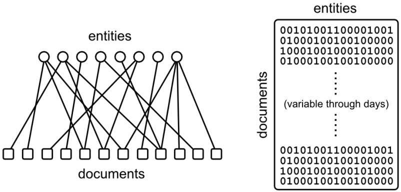


Figure: Financial documents on the Web represented as vectors of entities. "Normal" state on the left and "cohesive" state on the right.

Document-entity matrix

Document-entity matrix A - biadjacency matrix of a **bipartite-graph** between documents and entities



A is a boolean matrix and it records whether each entity is present or not in the document. Its size is $m \times n$ where m is number of documents and n is number of entities:

$$A_{i,j} = \begin{cases} 1 & \text{if entity } e_j \text{ is in document } d_i \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Document-entity matrix

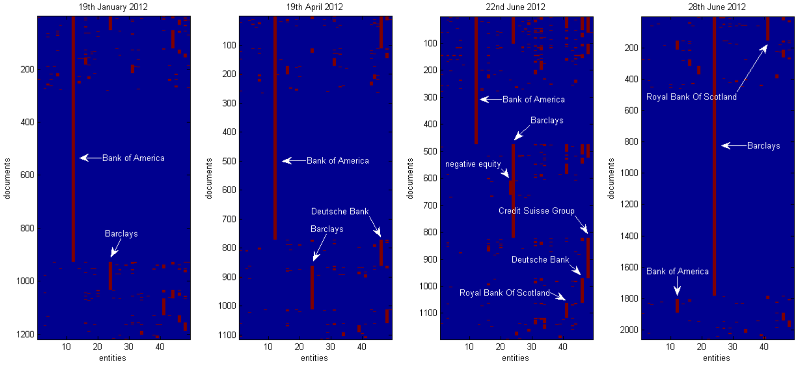


Figure: Document-entity matrix from four distinct days in 2012. Vocabulary consists of 13 banks listed on NYSE and 36 financial glossary terms.

Document-document similarity matrix

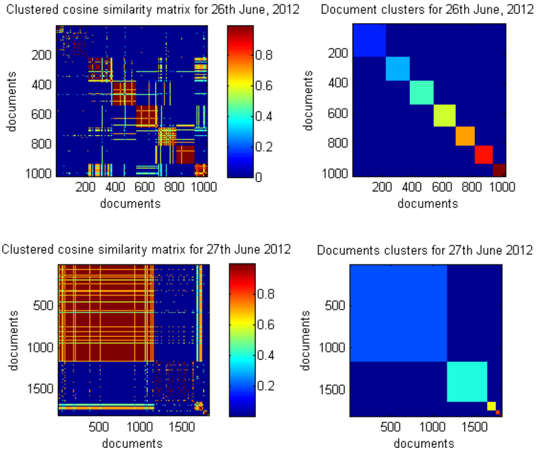


Figure: Document-document similarity matrix for two distinct days in 2012. Vocabulary consists of 13 banks listed on NYSE and 36 financial glossary terms.

News cohesiveness index (NCI)

We define *News cohesiveness index* (NCI) in two equivalent ways...

Definition through Frobenius norm

We define the NCI as the Frobenius norm of the scalar similarity matrix between all pairs of documents $C^d_{ij} = \langle d_i, d_j \rangle$ or pairs of entities $C^e_{ij} = \langle e_i, e_j \rangle$:

$$NCI = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{m} \| C^d_{ij} \|^2} = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} \| C^e_{ij} \|^2}. \qquad (2)$$

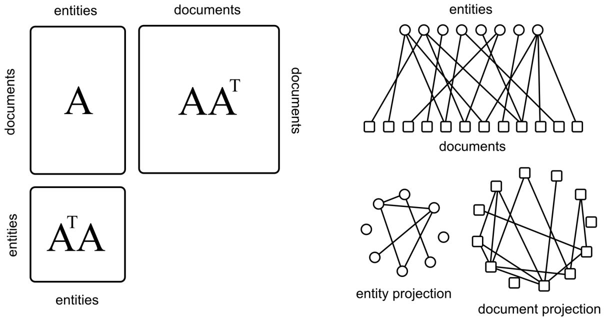
Definition through singular values

Function of singular values σ_i of matrix A:

$$NCI = \sqrt{\sum_{i=1}^{k} \sigma_i^4} \qquad (3)$$

NCI - properties of scalar similarity

If we measure the similarity between document $\vec{x}_1$ and $\vec{x}_2$ as a scalar product $\langle \vec{x}_1, \vec{x}_2 \rangle$ then bipartite projection to documents AA^T and projection to entities $A^T A$ have the same NCI: $\| AA^T \|_F = \| A^T A \|_F$.



NCI measure captures the **intrinsic property of the document-entity bipartite graph** that is invariant to projection!

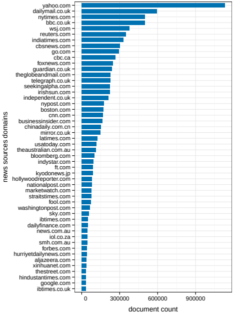
Part II

Results

Document collection pipeline

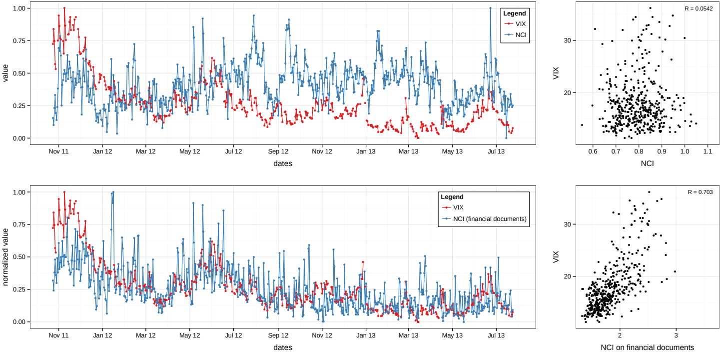
Corpus used in this work: Over 1,400,000 finance-related texts from 24^{th} October 2011 until 24^{th} July 2013 collected as a part of FIRST[1] and FOC[2] projects.

Filtering financial documents: We use 3500 manually labeled documents to build a machine learning model for classifying documents as financial, non-financial and neutral.

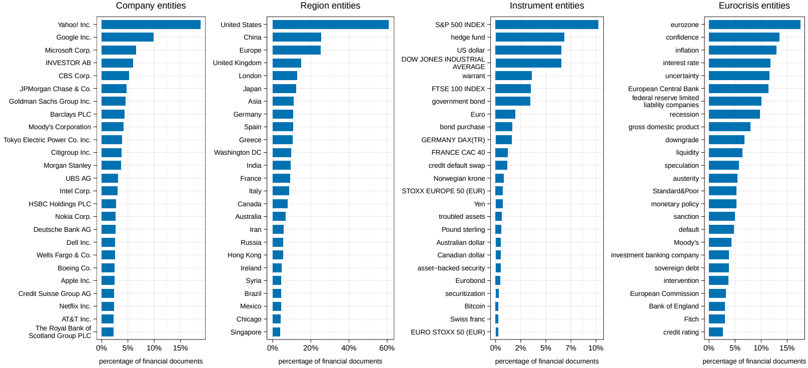


[1] http://project-first.eu/
[2] http://www.focproject.eu/

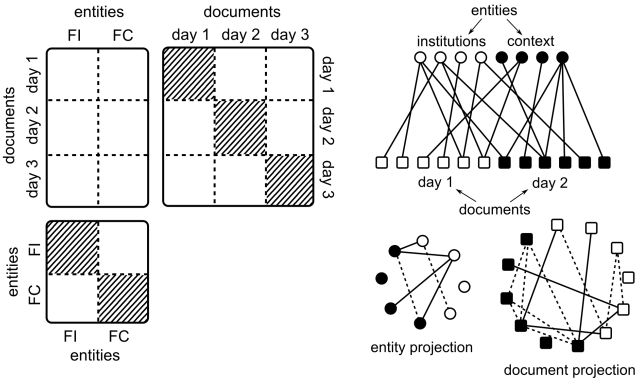
NCI and VIX


Legend
VIX
NCI
value
dates
Nov 11
Jan 12
Mar 12
May 12
Jul 12
Sep 12
Nov 12
Jan 13
Mar 13
May 13
Jul 13
1.00
0.75
0.50
0.25
0.00
R = 0.0542
VIX
NCI
R = 0.703
Legend
VIX
NCI (financial documents)
normalized value
NCI on financial documents

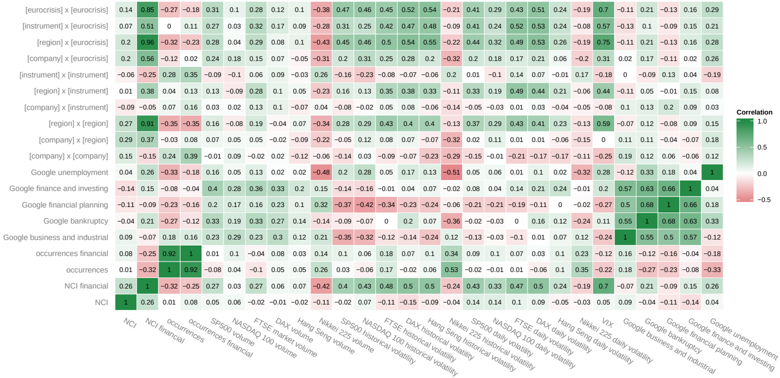
Structure of semantic components



Structure of semantic components



Correlations



Article

SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
INFORMATION TECHNOLOGY
COMPUTER SCIENCE
COMPLEX NETWORKS

Received
17 February 2014

Accepted
2 May 2014

Published
22 May 2014

Correspondence and requests for materials should be addressed to T.Š. (tomislav.smuc@irb.hr)

Cohesiveness in Financial News and its Relation to Market Volatility

Matija Piškorec[1], Nino Antulov-Fantulin[1], Petra Kralj Novak[2], Igor Mozetič[2], Miha Grčar[2], Irena Vodenska[3] & Tomislav Šmuc[1]

[1]Laboratory for Information Systems, Division of Electronics, Ruđer Bošković Institute, Croatia, [2]Department of Knowledge Technologies, Jožef Stefan Institute, Slovenia, [3]Department of Administrative Sciences, Metropolitan College, Boston University, USA.

Motivated by recent financial crises, significant research efforts have been put into studying contagion effects and herding behaviour in financial markets. Much less has been said regarding the influence of financial news on financial markets. We propose a novel measure of collective behaviour based on financial news on the Web, the News Cohesiveness Index (NCI), and we demonstrate that the index can be used as a financial market volatility indicator. We evaluate the NCI using financial documents from large Web news sources on a daily basis from October 2011 to July 2013 and analyse the interplay between financial markets and finance-related news. We hypothesise that strong cohesion in financial news reflects movements in the financial markets. Our results indicate that cohesiveness in financial news is highly correlated with and driven by volatility in financial markets.

The exponential growth of online media, expansion of communication and mobility-tracking capabilities have spawned research regarding the utility of the big data available from these sources. Big-data analytics aims to provide tools for better understanding large techno-social systems[1,2], improve predictions of different socio-economic outcomes and optimise processes. For example, Gonzales et al.[3] use 100,000 trajectories of mobile phone users to explain human mobility patterns. Ginsberg et al.[4] use Google search queries to help detect outbreaks of influenza epidemics in areas with a large population of web-search users. Whereas the aforementioned work estimates the current state of disease spread, other works focus on the predictive value of online information. For example, Goel et al.[5] demonstrate that Google search query volumes significantly improve predictions for the revenue of featured movies, video game sales and rank of songs. Similar to the above studies, our work explores the relationship between large corpora of online news and financial markets.

In this context, previous studies have analysed the relationship of search query volumes of specific terms with movements in financial markets of related items[6]. Bordino et al.[7] demonstrate that daily trading volumes of stocks traded on the NASDAQ 100 are correlated with the daily volumes of Yahoo queries related to the same stocks and that query volumes can anticipate peaks of trading by one or more days. Dimpfl et al.[8] report that Internet search queries for the term "dow" obtained from Google Trends can help predict the Dow Jones Industrial Average

Thank you for you attention!

Questions?