Multiplex Structure of Social Media and Financial Networks
(Extended Abstract)

Thársis T. P. Souza[1,⋆⋆], Giuseppe Pappalardo[4], Soong M. Kang[2], Tomaso Aste[1,3], and Guido Caldarelli[4]

[1] Department of Computer Science, UCL, Gower Street, London, WC1E 6BT, UK
[2] Department of Management Science and Innovation, UCL, Gower Street, London WC1E 6BT, UK
[3] Systemic Risk Centre, London School of Economics and Political Sciences, London, WC2A 2AE, UK
[4] IMT, School for Advanced Studies, Piazza S. Francesco 19, Lucca, IT

Abstract. We take advantage of an extensive data set composed of social media messages related to DJIA index components to investigate the relation between social media and financial markets network structure. Results suggest that social media attention persistence might lead stock market stability. Moreover, the social media structure has a considerable similarity with the stock market yet having higher persistence. In future work, we plan to use correlation-based graphs and network filtering tools to analyse the multiplex structure of social media and financial networks.

Keywords: social networks, financial networks, complex systems, social media, stock market

1 Introduction

Investors' decisions are modulated not only by companies' fundamentals but also by personal beliefs, peers influence and information generated from news and the Internet. Investors may react irrationally toward new information [3, 8], enabling opportunities to forecast movements in the market. However, it was only recently that the availability of vast amounts of data from online systems paved the way for the large-scale investigation of investor's collective behaviour in financial markets.

Previous studies tried to establish that online expressed opinions and measures of collective attention have predictive power over financial dynamics [10, 13, 11, 2, 5]. Nonetheless, common approaches focused on the investigation of individual links often neglecting joint dependences. In that way, only individual stocks are investigated and the multi-asset case is commonly neglected. Real-world datasets and, ultimately, human interactions are complex systems that need to be handled as such to explain financial dynamics in a realistic way. Moreover, recent research [9] has found that social media has a nonlinear impact on market prices therefore motivating the investigation of complex relationships in these systems.

Correlation-based graphs and network filtering tools have been successfully used in the literature to model the stock market as a complex system [12]. We aim to apply network filtering techniques [6, 12] to explain financial events via the analysis of the multilayer similarity and persistence of social media and financial networks. In this work, we show preliminary results obtained from a correlation-based structure derived from an extensive data set composed of social media messages related to DJIA index components. Similarity and persistence of social media and stock market structures are analyzed. Results suggest that periods of stability in the stock market might be preceded by periods of persistence in social media attention.

2 Data Analyzed

Our analysis is conducted on the 30 components of the Dow Jones Industrial Average (DJIA) index, which we monitored during the period from November 30, 2010 to August 17, 2015. The choice of these stocks was due to their representativeness of the stock market. We consider two streams of time series data: (i) market data, which are given at the daily stock price, and (ii) social media data

⋆⋆ Correspondence author: T.Souza@cs.ucl.ac.uk.

analytics based on more than 3 million Twitter messages. Let $P(t)$ be the closing price of an asset at day t, as financial variable we consider the stocks' daily log-returns: $R(t) = \log P(t) - \log P(t-1)$. As Twitter sentiment analysis [4] per se is out of scope of this study, we build our analysis on top of Twitter data analytics supplied by PsychSignal.com [1]. Similarly, naming $V(t)$ the total number of tweets, we defined the social media time series as $SM(t) = \log V(t) - \log V(t-1)$. A company is defined to be related to a given message if its ticker-id is mentioned as a *cashtag*, i.e., with its name preceded by a dollar symbol, e.g., $CSCO for the company CISCO SYSTEMS INC. In Twitter, a *cashtag* is a standard way to refer to a listed security.

3 Method

We followed procedure similar to [7] in the construction of a detrended stock log-returns structure for the correlation-based graph financial network.

Let $r_i(t)$ be the daily log-return of the stock i. Given a time-window T_k, we consider $c_i(t)$ as the detrended stock log-return of $r_i(t)$ in a time-window T_k:

$$r_i(t) = \alpha_i + \beta I(t) + c_i(t) \tag{1}$$

where $I(t)$ represents the market index, which we consider as the mean of log-returns of all stocks inside the window, i.e., $I(t) = \sum_{\gamma=1}^{N} r_\gamma(t), t \in T_k$.

We have selected a set of $n = 84$ overlapping time windows T_k with $k = 1, \ldots, n$ (each one of length $L = 60$ calendar days, with 20 calendar days shift between adjacent time windows) and computed the financial distance matrix $F_{ij}(T_k) = \sqrt{2(1 - \rho_{ij}(T_k))}$, where ρ_{ij} is the Kendall's rank correlation between the detrended log-returns of stocks i and j smoothed exponentially with $\theta = 20$ inside the window T_k. Analogous procedure was applied for the construction of the social media distance matrix SM_{ij}. Hence, each distance matrix have N dimensions representing the stocks analyzed. $F_ij(T_k)$ represents the distance matrix of stock's returns at window T_k while $SM_{ij}(T_k)$ represents the distance matrix of volume tweets related to the stocks investigated at window T_k.

We analyse persistence and similarity of the social media and financial distance matrices. Persistence is evaluated with the cross-correlation of the distance matrix for all windows considered. Hence, the persistence of social media is given by $corr(SM_{ij}(T_{k1}), SM_{ij}(T_{k2})), \forall k1, k2 \in [1, n]$ and the stock's returns persistence is analogously defined. The similarity is a measure of closeness between social media and financial matrices and are estimated as $corr(SM_{ij}(T_{k1}), F_{ij}(T_{k2})), \forall k1, k2 \in [1, n]$.

4 Results and Discussion

Fig. 1 shows the results of the persistence analysis. In A), we observe that social media presents long periods of high persistence. The first one (I) starts in November of 2010 and lasts until January of 2012 which are then followed by a new long period of high persistence (II). Interestingly, this new period of persistence is dissimilar with the previous structure. Then, a new long period of high persistence (III) started at the beginning of 2013 during approximately one year. This period is followed by two new blocks of persistence (IV and V): one assuming a different structure (IV) and the other (V) recovering a configuration similar to the period of a past block (II).

Fig. 1 B) shows the results of the same analysis for the financial distance matrix. Compared with the social media structure, the financial persistence is lower. Long blocks of medium to high persistence are still present, but less defined. Nonetheless, the financial structure presents blocks of high persistence that are similar to the ones identified in the social media structure. Surprisingly, we observe that periods of financial persistence follow periods of social media stability to some extent.

Acknowledgments

This work was supported by PsychSignal.com, which provided the social media analytics. T.A. acknowledges support of the UK Economic and Social Research Council (ESRC) in funding the Systemic Risk Centre (ES/K002309/1). T.T.P.S. acknowledges financial support from CNPq - The Brazilian National Council for Scientific and Technological Development.

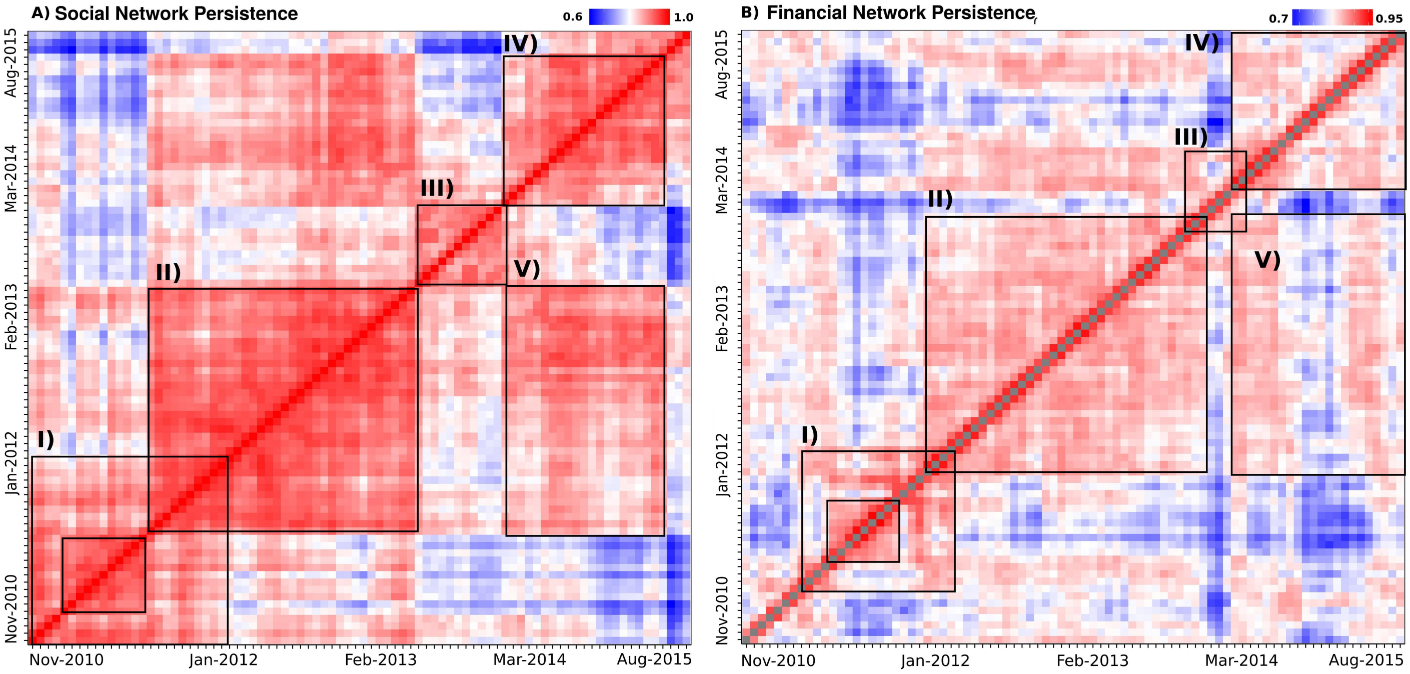

Fig. 1: **Evidence that periods of stock market stability (blocks highlighted) may be preceded by social media attention persistence.** Figure shows persistence analysis of social media layer (A) and financial layer (B). Persistence is measured as the lagged correlation between distance matrices, which were estimated in a rolling window of size of 60 calendar days with step-size of 20 days. Axes show the date of the beginning of each window. Blocks of periods with high persistence are highlighted.

References

1. The psychsignal website (Oct 2015), `https://www.psychsignal.com`
2. Alanyali, M., Moat, H.S., Preis, T.: Quantifying the relationship between financial news and the stock market. Sci. Rep. 3 (2013)
3. Bondt, W.F.M.D., Thaler, R.: Does the stock market overreact? The Journal of Finance 40(3), pp. 793–805 (1985), `http://www.jstor.org/stable/2327804`
4. Kolchyna, O., Souza, T.T.P., Treleaven, P., Aste, T.: Twitter sentiment analysis: Lexicon method, machine learning method and their combination. arXiv preprint. http://arxiv.org/abs/1507.00955 (2015)
5. Mao, H., Counts, S., Bollen, J.: Quantifying the effects of online bullishness on international financial markets. European Central Bank Workshop on Using Big Data for Forecasting and Statistics, Frankfurt, Germany (2014)
6. Massara, G.P., Di Matteo, T., Aste, T.: Network filtering for big data: Triangulated maximally filtered graph. arXiv preprint arXiv:1505.02445 (2015)
7. Musmeci, N., Aste, T., Di Matteo, T.: Risk diversification: a study of persistence with a filtered correlation-network approach. arXiv preprint arXiv:1410.5621 (2014)
8. Shleifer, A.: Inefficient Markets: An Introduction to Behavioral Finance. Clarendon Lectures in Economics, OUP Oxford (2000)
9. Souza, T.T.P., Aste, T.: A nonlinear impact: evidences of causal effects of social media on market prices. ArXiv e-prints (Jan 2016)
10. Souza, T.T.P., Kolchyna, O., Treleaven, P.C., Aste, T.: Twitter sentiment analysis applied to finance: A case study in the retail industry. arXiv preprint. http://arxiv.org/abs/1507.00784 (2015)
11. Tetlock, P.C.: Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance 62(3), 1139–1168 (2007)
12. Tumminello, M., Aste, T., Di Matteo, T., Mantegna, R.N.: A tool for filtering information in complex systems. Proceedings of the National Academy of Sciences of the United States of America 102(30), 10421–10426 (2005)
13. Zheludev, I., Smith, R., Aste, T.: When Can Social Media Lead Financial Markets? Scientific Reports 4 (Feb 2014)