Experiment for analysing the impact of financial events on Twitter

Ana Fernández Vilas[1[0000-0003-1047-2143]], Lewis Evans[2], Majdi Owda[2], Rebeca P. Díaz Redondo[1] and Keeley Crockett[2]

[1] I&C Lab. AtlantTIC Research Centre. University of Vigo. 36310 Spain.
avilas@det.uvigo.es, rebeca@det.uvigo.es
[2] School of Computing, Mathematics & Digital Technology.
Manchester Metropolitan University, M1 5GD UK.
L.Evans@mmu.ac.uk, m.owda@mmu.ac.uk, k.crockett@mmu.ac.uk

Abstract. Twitter, as the heart of publicly accessible Social Media, is one of the currently used platforms to share financial information and is a valuable source of information for different roles in the financial market. For all these roles, the quality analysis of Twitter as a source of financial information is essential to take decisions. The work in this paper is aligned with the ongoing work of the authors to a solution for irregularity monitoring in the financial market by harnessing data in online social media. To do so, the permeability of a variety of social media data feeders to financial irregularities should be analysed. That is the case of the experiment in this paper by putting the focus on Twitter microblogging platform and checking if this general purpose social media is permeable to a specific financial event. For this, we detail the analysis of Twitter permeability to a specific event in the past few months: the announcement about the merge of Tesco and Booker to create a UK's Leading Food Business on the 27th January 2017. Both companies Tesco PLC and Booking Group PLC are listed in the main market of LSE (London Stock Exchange). Our findings provide promising evidences to address the problem of real-time detection of irregularities in the financial market via Twitter according to the volume (as a sign of the importance of the irregularity) and to other features (as signs of the potential origin causing the irregularity).

Keywords: Twitter, Stock Market, Financial Irregularities, Permeability.

1 Introduction

As the heart of publicly accessible Social Media, Twitter has become a vital source for open source intelligence in natural disasters, politics, consumers' opinion, etc. Also, Twitter is one of the currently used platforms to share financial information from businesses, brokers, news agencies or through individual investors tweets. As Twitter usage to share financial information is definitively increasing [1]; it is important to stress that, according to [2], stock microblogs exhibit three distinct characteristics above stock message boards: (i) Twitter's public timeline may capture the natural market conversation more accurately and reflect up to date developments; (ii) Twitter reflects a more ticker-like live conversation which allows micro-bloggers to be exposed to the most recent information of all stocks and does not require users to actively enter the forum for a particular stock; and (iii) micro-bloggers have a strong incentive to publish valuable information to maintain reputation (increase mentions, the rate of retweets, and their followership), meanwhile financial bloggers can be indifferent to their reputation in the forum. Providing sensing, harvesting and analysing methods and tools of such information could be very useful for many stakeholders such as businesses and individuals making decisions to invest, stock market analysts and law enforcement agencies.

Our medium-term objective is a collaboration project among the University of Vigo and the Manchester Metropolitan University to deploy an architecture for real-time monitoring of irregularities in the stock market. That architecture will apply data mining and fusion technologies from a pool of social feeders related with the stock market. In order to design the architecture, the permeability of the different feeders should be analysed, that means, to what extent a specific financial information feeder is permeable to fraudulent and common irregularities in the financial market. That is the case of the experiment in this paper by putting the focus on Twitter microblogging platform.

This paper states the following research question: Is Twitter permeable to specific actions ion the financial market? We hypothesize that Twittersphere, the total universe of Twitter users and their habits, is permeable towards relevant actions in the financial market and that the impact of this permeability can be measure according to (1) the disturbance of Twitter behaviour in terms of volume, tweets features and geographical distribution; and (2) the rapidness of this permeable layer between the financial market and the social media (Twitter in our experiment). Showing that a general purpose social media is permeable to financial-specific events is the first step to consider Twitter as a relevant feeder for taking decisions regarding the financial market and event fraudulent activities in that market. For this, we detail the analysis of Twitter permeability to a specific event in the past few months: the announcement about the merger of Tesco PLC (hereinafter Tesco) and Booker Group PLC (hereinafter Booker) to create UK's Leading Food Business on the 27th January 2017. Both companies Tesco PLC and Booking Group PLC are listed in the main market of LSE (London Stock Exchange).

This paper is structured as follows. Section 2 introduces the Twitter efforts to accommodate financial information in a general-purpose microblogging platform as well as related work in the area of the use of Twitter data for financial analysis where researchers capture data by using the APIs provided by Twitter, which are discussed in Section 2.1 along with the selection of Twitter features we consider during our experiment. Section 3 describes the scenario, the extraction strategy and the resulting datasets for the experiment, which aims to analyse permeability for the Tesco & Booker merger. After cleaning the data and conducting the analysis, the paper reports the impact of the merger on 27th January 2017 in terms of tweets volume and features (Section 4), in terms of geographical distribution (Section 5) and in terms of its rapidness to react to the action (Section 6). Finally, Section 7 discusses our findings and introduces our on-going work in the study of permeability of Twitter to financial events.

2 Twitter & Financial information

It is fair to say that it was Twitter that popularised the term hashtag as well as its # symbol to index keywords or topics so that people can easily follow topics they are interested in. Also, in 2012 Twitter unveiled a new clicking & tracking feature for stock symbols (known as Cashtags). Cashtags are stock market symbols that can be included in tweets and when preceded with a dollar sign (for example $VOD in regards to Vodafone) become clickable. [3] reported an exploratory analysis of public tweets in English, extracted via Firehose, which should contain at least one Cashtag from NASDAQ or NYSE. The analysis concludes that the use of Cashtag is higher in the technologic sector, which seems to be related with the technological profile of most of the Twitter users; and the top 10 Twitter accounts according to the usage of cash-tags are companies or news agencies (i.e. automatic or semi-automatic Twitter accounts). The analysis also highlights the existence of relevant information behind the co-occurrence of Cashtags (revealing main competitors of companies) and the co-occurrence of Cashtags with Hashtags (allowing to group companies into clusters). Some other works research on the possible connections between Twitter information and market performance, that is the predictive value of information gathered form social media [4] [2]. Most of these works, based on the twitter data volume, also apply some sentiment analysis technique in order to distinguish the polarity of the impact [5], [6] [7] [8].

2.1 Twitter Mining

There are three different ways to catch Twitter data: Search API, Streaming API and Firehose. The Twitter Search API provides the endpoints to recover tweets that were published in the previous two weeks, with the possibility of filtering according to several criteria. On the other hand, Twitter Streaming API returns 1% of the tweets that match some search parameters in real time. Finally, Twitter Firehose provide access to 100% of the tweets, but it is not a free-access API. Twitter APIs are constructed around four main “objects”: Tweets, Users, Entities (hashtags, URLs, mentions and media in a tweet) and Places. Then, users construct API queries (combining object fields and query operators) to retrieve information posted from specific users, containing a particular combination of keywords, including particular entities, etc. With regard to this work, the experiment does not include the analysis of the spreading perspective of information on Twitter so we select the following features (existing in both Search and Streaming APIs under different field names) for the analysis, all of them accessible from a Tweet object:

- Content perspective: the status update (`Tweet:text`) and the entities (`Tweet:entities`), specifically `hashtags` (including cashtags) and `urls`.
- Context perspective: the post time of the status update (`Tweet:created_at`) and, if available, also the place (`Tweet:coordinates; Tweet:place:bounding_box`).
- Social Perspective: User (`Tweet:user`, specifically the field `verified`).

There are highly relevant differences between the Searching API and the Streaming API, time direction being the most apparent and functionally-impacting one. Search API goes back in time meanwhile Streaming API goes forward. Moreover, there are other differences related to mainly the format and the rate limit rules. Regarding their extracting capacity, Twitter forums contain plenty of discussion about this issue which has not ever made enough clear from Twitter officially.

3 The experiment and the data

As mentioned, the aim of this experiment is analysing the permeability of Twitter to the occurrence of specific events in the day to day of financial market. For that, we perceive TESCO on Twitter with the pair (cashtag, keyword), that is ($TSCO, “tesco”), representing the financial perspective of TESCO on Twitter ($TSCO) and general references to TESCO on Twitter (“tesco”). According to this representation, we respond to our research question. Our hypothesis is that Twitter (although not a specific financial forum) is permeable to financial events and this permeability can be analysed by monitoring the name of companies as a keyword (“tesco” in this case) and the Cashtag of the company ($TSCO). Also, we hypothesize that the permeability and the impact is not alike in the two perspectives. Meanwhile the cashtag is invariably linked to financial news of a company, the general content, or ‘Tesco’ content, have some completely different dynamics which is generally driven by company decisions, marketing campaigns, consumer opinions, etc. Presumably, financial events should have a bigger impact on cashtag tweets (according to volume and features) than on tweets containing the keyword ‘Tesco’. Nevertheless, this presumably different behaviour should be inspected. Taking this merge action as our first experiment to a general measure of permeability, while taking into account that we are reporting a single event, we analyse the impact of this financial event on Twitter $cashtag-content and on Twitter keyword-content related with the company, separately. The impact on both data sources ($TSCO and ‘tesco’) is measured in terms of Twitter volume (Section 4), in terms of geographical distribution (Section 5) and in terms to their response to the announcement by the RNS (Regulatory News Service) of LSE[1] (Section 6).

3.1 Data extraction

We prepared the experiment according to the following extraction strategy for the query ($TSCO, “tesco”). Once we selected the event, we used the Search API to recover the information backwards before the announcement on 27th of January 7:00 a.m. and the streaming API to recover information forwards. The aim of streaming data just after the announcement was to visualize the impact of the announcement and analyse the time Tesco Twitter behaviour returns to a regular pattern. The results of the combination of the search and streaming results is shown in **Fig. 1**. Once the behaviour becomes stable, we used the Search API again to obtain a regular dataset as a reference for the experiment.

---

[1] http://www.londonstockexchange.com/exchange/news/market-news/market-news-home.html

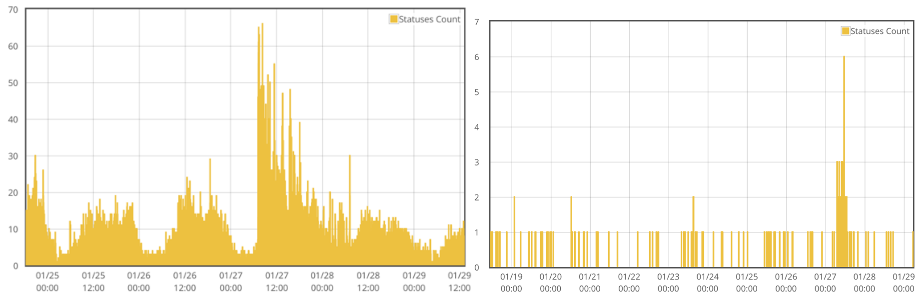


Fig. 1. Total Twitter volume for 'tesco' (left) and $TSCO (right) by merging (without duplicates) the retrieved data from queries to Search API (backwards) and Steaming API (backwards).

Clearly, the Twitter Search API is not appropriate for continuous analytical monitoring and as a data source to taking decisions in real time. It is not intended and does not fully support the repeated constant searches that would be required to deliver 100% coverage. However, the experiment in this paper is limited to one individual company, 2 keywords and timelines in the scale of weeks. In such conditions, Search API provide a better coverage than de Streaming API (1% according to the Twitter official information) if we use the superior filtering characteristics of the Search API. Nevertheless, as the Search API has a limit on the number of returning Tweets, to get the whole data, we repeatedly ask Twitter for the most recent results backwards by windowing the searches according to the publication date and merging results according to the post Id. Apart from that, the Search API guarantees a fair comparison according to the volume of data, in any manner we should compare Search results with Streaming results. According to that, and to give response to the research questions, we use Search API queries to cover the time periods in **Table 1**.

Table 1. Time Periods (UK Time) extracted with the Search API.

Name/Period	'tesco'		$TSCO	
	Total	**Per/hour**	**Total**	**Per/hour**
Pre-announcement 25th Jan 00:00- 27th Jan 06:59	11,817	214.85	12	0.218
Post-announcement 27th Jan 07:00 - 29th Jan 23:59	25,547	393.03	91	1.400
Regular 2-weeks-after 8th Feb 00:00-10th Feb 06:59	13,417	243.94	26	0.473
Regular 2-weeks-after 10th Feb 07:00 - 12th Feb 23:59	20,012	307.88	22	0.338

4 Impact on Twitter volume

In this section, we detail the impact of the event by analysing the variation in the number of tweets (volume) with respect to the regular behaviour, which provides a quantitative measure of Twitter permeability to the Tesco & Booker merger. During this part of the analysis some irregularities appeared which uncovered an inconsistency in the named scheme of tickers in Twitter. In particular, to our knowledge, Twitter has not promoted the specific distinction among markets so that the uniqueness of ticker symbols inside a market disappear in the Twittersphere. That is the case of $TSCO cashtag which corresponds to 'Tesco PLC' in LSE and to 'Tractor Supply Company' in NASDAQ (National Association of Securities Dealers Automated Quotation), the second stock exchange in USA. So, the returned results to a $TSCO query include tweets related to Tesco Plc and also to Tractor Supply Company. If cashtags are the Twitter vehicle to aggregate and allow the spreading of financial information about companies, some kind of market prefix should be used, specially in the times when companies are becoming increasingly global.

Fig. 2 shows the temporal series in a tweets-per-hour (TPH) scale. Although it is quite obvious that the number of TPH in 'tesco' dataset is up several orders of magnitude higher than those of $TSCO dataset, the peak behaviour is more acute in the $TSCO one. As it is shown in **Table 2**, considering the hourly volume of 'tesco' dataset on the 27th January, there are not outliers during the day, with a peak value of 2,057 tweets in the sample from 8:00 to 9:00. Nevertheless, there are 3 outliers in the $TSCO dataset: samples 8:00-9:00, 9:00-10:00 and 12:00-13:00, corresponding to the time just after the announcement and lunch time in the UK, the latter being consistent with previous studies about social timing, i.e. [9].

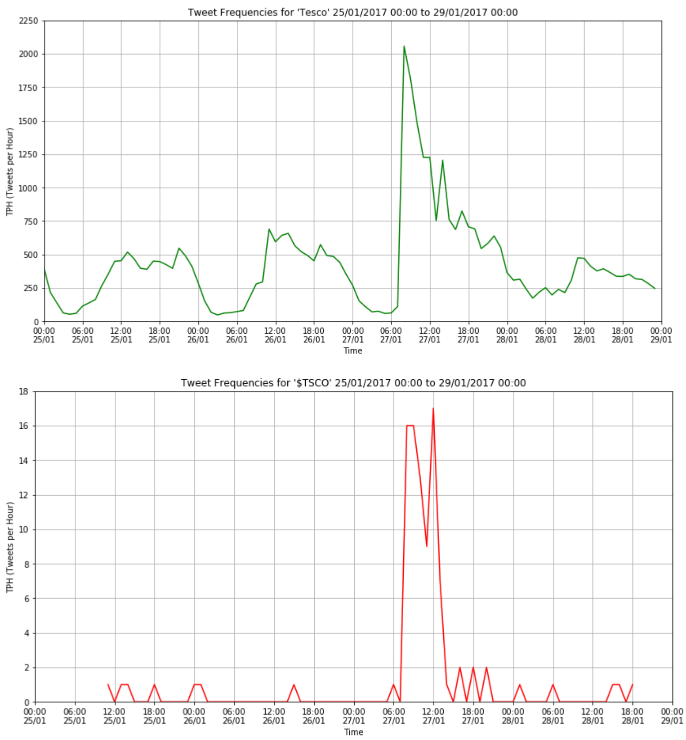


Fig. 2. Time series of the 'tesco' and $TSCO dataset from 25th January to 29th January

Table 2. Peak behaviour on the 27th January for ‘tesco’ and $TSCO.

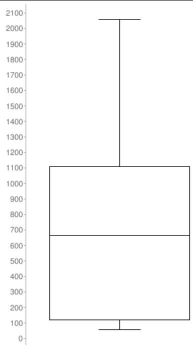
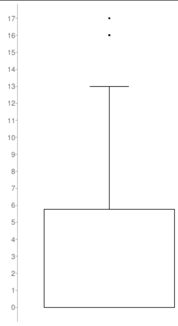
1-hour scale ‘tesco’		1-hour scale $TSCO	
Tweets: 40567 Median: 662.5 Minimum: 58 Maximum: 2057 1st Q: 121.5 3rd Q: 1110 Inter Q: 988.5 Outliers: none		Tweets: 98 Median: 0 Minimum: 0 Maximum: 17 1st Q: 0 3rd Q: 5.75 Inter Q: 5.75 Outliers: 17 16 16	

Table 3. Variability of features in ‘tesco’ and $TSCO datasets (Green bars correspond to the variation in the ‘tesco’ dataset and blue bars correspond to the variation in the $TSCO dataset).

Periods	Pre-Announcement (25th-27th Jan 06:55) TUESDAY-WEDNESDAY-THURSDAY				Post-Announcement (27th 07:00- 29 Jan) FRIDAY- SATURDAY-SUNDAY			
Counting & percentages	**"tesco"**	**%**	**$TSCO**	**%**	**"tesco"**	**%**	**$TSCO**	**%**
Tweets	**17,154**		**8**		**25,547**		**91**	
Tweets per hour	**311.89**		**0.15**		**393.03**		**1.40**	
Tweets from verified users	2,560	14.92%	0	0.00%	2,696	10.55%	2	2.20%
Tweets with URL	5,383	31.38%	6	75.00%	12,367	48.41%	64	70.33%
Tweets being RT	7,522	43.85%	0	0.00%	8,070	31.59%	18	19.78%
Different users	12,141	43.85%	7	0.00%	15,757	31.59%	47	19.78%
Different erified users	155	1.28%	0	0.00%	336	1.32%	2	4.26%
Periods	**2 Weeks after (7th - 9th Feb 06:55) TUESDAY-WEDNESDAY-THURSDAY**				**2 weeks after (10th 07:00 - 12th Feb) FRIDAY-SATURDAY-SUNDAY**			
Counting & percentages	**"tesco"**	**%**	**$TSCO**	**%**	**"tesco"**	**%**	**$TSCO**	**%**
Total Tweets	**16,878**		**23**		**20,011**	**100.00%**	**22**	
Tweets per hour	**306.87**		**0.42**		**307.86**		**0.34**	
Tweets from verified users	2,364	14.01%	0	0.00%	2,650	13.24%	0	0.00%
Tweets with URL	4,980	29.51%	19	82.61%	6,971	34.84%	19	86.36%
Tweets being RT	6,530	38.69%	3	13.04%	5,676	28.36%	5	22.73%
Different users	10,749	38.69%	10	13.04%	11,374	28.36%	13	59.09%
Different verified users	164	0.97%	0	0.00%	150	0.75%	0	0.00%

Apart from the peak comparison, we also inspected the potential disturbances on other dataset features before and after the announcement, also comparing these dates with the regular behaviour 2 weeks later (see **Table 3**). We highlight the invariability on the number of verified users either along all the periods and along the two datasets. Secondly, the percentage of tweets which contain URL are significantly higher in the $TSCO dataset with respect to the ‘tesco’ one, which is a result of the professional and financial orientation of the $TSCO data as a channel to spread facts and news rather than opinions and sentiments. Finally, the retweeting activity is higher in the announcement periods (pre- and post-) compared to the regular periods in both datasets. The increase of retweeting is, by nature, linked to the need or desire of spreading a piece of content but, the reason behind may be different as, in fact, it is in our case study: retweeting in the ‘tesco’ keyword dataset is mainly related with a Tesco campaign for wining a voucher, meanwhile retweeting in $TSCO data is mainly linked to spreading the information about the merge (post-announcement) and about other financial news.

5 Impact on geographical distribution

Although Twitter is one of the most used data source in data mining, the geo-location component of Twitter is not comparable to other data sources which we can refer to as Location-based social networks. In fact, according to [10], the geo-located tweets returned by the Streaming API cover up to the 90% of the geo-located tweets extracted from Firehose API. However, [10] also reveals that the number of geo-located tweets is low, being only a 1.45% of the tweets obtained from Firehose API and 3.17% of the tweets obtained from Streaming API. The total percentage of geo-located tweets for the 'tesco' dataset is consistent with this previous study [10], with a percentage of 4.3% for all the periods in the experiment. Although the number of tweets in the $TSCO dataset may be not representative enough, we should remark that the percentage of geo-located tweets in the $TSCO dataset is almost 0%, 1 tweet out of a total of 199, so bellow the 4.3% in 'tesco' dataset. Also, there is not variability of those percentages throughout the periods considered (pre- post- and regular). Although these data should be interpreted with caution, we may consider the possibility of accessing from a desktop device or corporate mobile in the case of financial professionals (supposedly devices without location feature or whit this feature disabled).

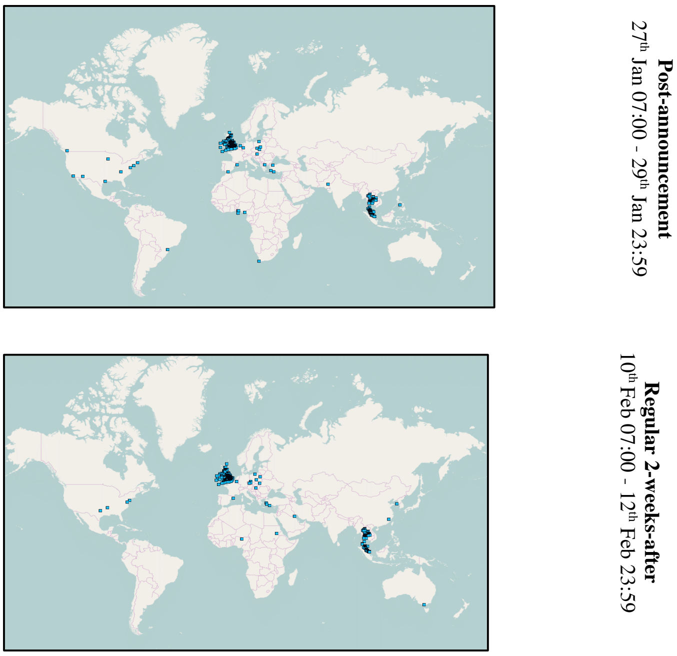


Fig. 3. Geographical distribution of 'tesco' dataset after the announcement and during a regular period.

Beyond the percentage of geo-located tweets that the Twitter APIs return, the variation of the geographical distribution of the tweets due to the financial event deserves to be analysed. **Fig. 3** shows this distribution and, apparently, there is not much variation if we compare post-announcement with the regular period for the same days of the week.

A deeper inspection of the tweets per country in **Table 4** confirms that most of tweets come from the countries where Tesco deploy its main business either under Tesco trademark or thorough subsidiary local companies. Apart from UK and Republic of Ireland, the main retail locations of Tesco PLC all over the world are the Czech Republic, Hungary, Poland, Slovakia, Turkey, Malaysia and Thailand. According to the results in the table, before the announcement, the bigger contribution to Twitter volume corresponded to the UK market which is consistent with the historical roots of the company in this country where its retailing business is fully integrated in the society. Nevertheless, after the announcement, this percentage decreases in favour of other locations over the world, which is a sign of the global impact of the action so that twitter users outside UK are not so linked to Tesco PLC main business campaigns during regular period but they are reactive to a relevant event related with a company with presence in their countries. Nigeria is highlighted in Table 4 as a country with a definitely high position during the post-announcement despite the fact that Tesco does not have business in this country. 42 of the 43 tweets in Nigeria has the same content but they are tweeted from 42 different users, not being retweets, so that it may be a violation of the spam terms in Twitter rules.

Table 4. Geographical distribution of tweets in 'tesco' dataset

Pre-Announcement			Post-Announcement		
Country	Geolocated Tweets	%	Country	Geolocated Tweets	%
UK	443	61.27%	UK	585	51.32%
Malysia	196	27.11%	Malysia	341	29.91%
Thailand	53	7.33%	Thailand	125	10.96%
			Nigeria	42	3.68%
			Ireland	15	1.32%
Rest of the wor	31	4.29%		32	2.81%
TOTAL	723		TOTAL	1140	
2 Weeks after (8th - 10th Feb)			**2 weeks after (10th- 12th Feb)**		
Country	Geolocated Tweets	%	Country	Geolocated Tweets	%
UK	333	58.12%	UK	503	55.76%
Malysia	162	28.27%	Malysia	242	26.83%
Thailand	33	5.76%	Thailand	100	11.09%
Ireland	18	3.14%	Ireland	22	2.44%
	27	4.71%		35	3.88%
TOTAL	573		TOTAL	902	

6 Rapidness

Although our analysis focuses on the permeability of Twitter to financial events our long-term objective is the use of Twitter as a sensor of irregularities in the stock market. So, this section includes our findings related to the rapidness and synchronization of Twitter as a channel to the activity in the stock market: rapidness in its response to the RNSs of LSE (London Stock Exchange) and synchronization with the share prices also in LSE. Regarding the rapidness, the experiment definitively shows the good characteristics of Twitter. The first tweet referring to the RNS was at 7:03 a.m. on 27th, just 3 minutes before the RNS announcement about the Tesco and Booker merge. Beyond the very first tweet, it is remarkable the rapidness of the peak response to the announcement in both datasets, so that the 27th Twitter time series (‘tesco’ and $TSCO) can be considered abnormal time series when a regular Friday is taken as a reference. We highlight that the peak starts form 7:00 to 8:00 both in the #TSCO and $TSCO dataset (see **Fig. 4**).

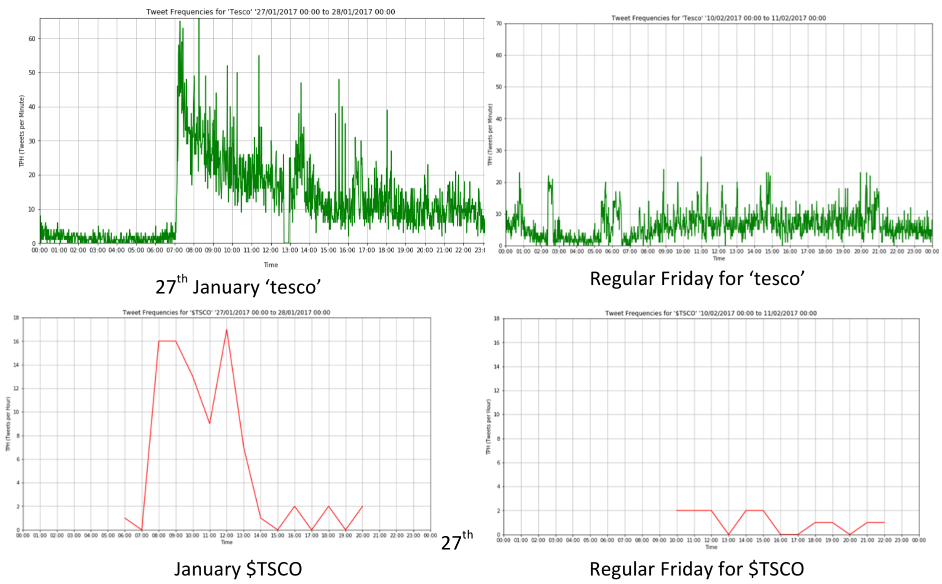


Fig. 4. Time series at hour scale on the 27th January in comparison with a regular Friday.

Regarding the synchronization with the share prices at LSE (**Fig. 5**), it is fair to mention that although the share prices were abnormally low the day before the announcement, we haven’t found any reference to a potential Tesco & Booker merger in tweets before the announcement in our dataset, neither by manual inspection of Twitter Web Site.

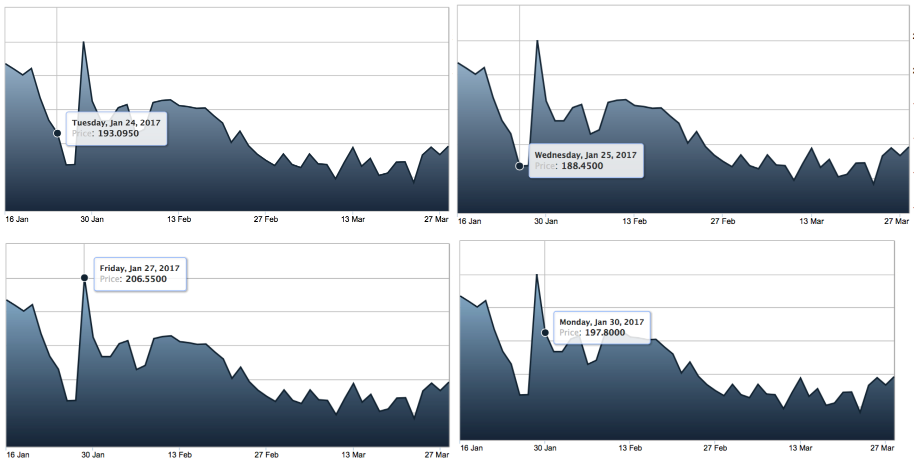


Fig. 5. Main observational points in the Evolution of the Tesco PLC share price (16th January to 27th March 2017)

7 Discussion

This paper inspects the permeability of Twitter to financial events in order to provide evidences which allows Twitter to be used as social sensor for the financial and stock market. Bearing in mind that this a single experiment for a single financial event and also that the event was fully covered by traditional social media, we can conclude that the event in the financial market invaded the Twittersphere on the 27th January, just after the RNS announcement at 7:00, and that the behaviour of ($TSCO, "tesco") was altered in comparison with the regular behaviour around the company involved in the financial event. Nevertheless, the experiment had a little success in predicting the irregularity, that is, identifying some rumour or sign of the announcement. Even considering that the experiment was not deployed over the whole Firehose Twitter data, uncovering rumours before the announcement turns definitively into a hard task if the human spreading of rumours is not mimic inside Twitter, that means, if the rumour is not there. At this respect, and according to [11], social media data can only be generalized to human behaviour if social media provides a representative description of human activity. Twitter is a social media which, at least, exhibit some demographic bias. Moreover, Twitter may be providing a skewed representation of their content. Although well-known rumour detection algorithms [12] [13] can be applied to Twitter, an alternative approach can be the fusion of financial information from different data sources in a way that we can mitigate the inevitable bias in a single source, and, at the same time, combine their weaknesses and strengthens in a proper representation of the real financial activity.

Meanwhile this paper analyses the quantitative and objective permeability of a financial event on Twitter, our ongoing work has conducted and initial analysis the qualitative characteristics of that permeability: in terms of topic modelling and information provenance, but also considering the polarity of twitter financial information. The experiment in this paper provides promising results to address our final objective: a real-time monitoring system which would detect irregularities according to the volume (as a sign of the importance of the irregularity) and to other features (as signs of the potential origin causing the irregularity). Such a system implies prime benefits for individuals, specially uninformed traders, and for regulatory and low enforcement agencies as a sign which may trigger further actions. Unfortunately, the need to influence social media for different purposes is often linked to the propagation of information with low credibility level or definitely false. Also, as it is shown in [14], firms strategically disseminate information in social media, that is, they decide to use or not to use certain channels depending on the piece of news. Even worse, information may be automatically disseminated by artificial agents in order to influence the community in a deceptive way.

Acknowledgement

This work was funded by Spanish Ministry of Education Culture and Sports, National Plan for Scientific and Technical Research and Innovation (Sub-Programme for Mobility) under the research stay grant PRIX16/00368. We thank the Manchester Metropolitan University (School of Computing Mathematics and Digital Technology) for its support during the research stay.

References

[1] L. Cazzoli, R. Sharma, M. Treccani and F. Lillo, "A Large Scale Study to Understand the Relation between Twitter and Financial Market," in *2016 Third European Network Intelligence Conference*, 2016.

[2] T. O. Sprenger, A. Tumasjan, P. G. Sandner and I. M. Welpe, "Tweets and Trades: the Information Content of Stock Microblogs.," *Eur Financial Management,* vol. 20, p. 926–957, 2014.

[3] M. A. O. Hentschel, "Follow the money: A study of cashtags on Twitter," *First Monday,* vol. 19, no. 8, 2014.

[4] E. J. Ruiz, V. Hristidis, C. Castillo, A. Gionis and A. Jaimes, "Correlating financial time series with microblogging activity.," in *Proceedings of the fifth ACM international conference on Web search and data mining*, 2012.

[5] N. Oliveira, P. Cortez and N. Areal, "The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices," *Expert Systems with Applications,,* p. 125–144, 2017.

[6] J. K.-S. Liew and T. Budavári, " Do Tweet Sentiments Still Predict the Stock Market?," 2016.

[7] N. Rajesh and L. Gandy, "CashTagNN: Using sentiment of tweets with CashTags to predict stock market prices," in *11th International Conference on Intelligent Systems: Theories and Applications (SITA)*, 2016.

[8] P. Cortez, N. M. R. Oliveira and J. C. P. Ferreira, "Measuring user influence in financial microblogs: experiments using stocktwits data," in *WIMS'16 Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics*, 2016.

[9] M. Adnan, A. Leak and P. L. Geo-Spatial, "A geocomputational analysis of Twitter activity around different world cities," *Information Science,* vol. 17, no. 3, 2014.

[10] F. Morstatter, J. Pfeffer, H. Liu and K. M. Carley, "Is the sample good enough? Comparing data from twitter's streaming API with Twitter's firehose.," in *In Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013*, 2013.

[11] H. Liu, F. Morstatter, J. Tang and R. Zafarani, "The good, the bad, and the ugly: uncovering novel research opportunities in social media mining.," *International Journal of Data Science and Analytics,* vol. 1, no. 3-4, pp. 137-143., 2016.

[12] S. Vosoughi, *Automatic Detection and Verification of Rumors on Twitter,* Massachusetts Institute of Technology, 2015.

[13] A. Tafti, R. Zotti and W. Jank, "Real-Time Diffusion of Information on Twitter and the Financial Markets," *PLoS ONE,* vol. 11, no. 8, 2016.

[14] M. J. Jung, J. P. Naughton, A. Tahoun and C. W. :, "Do Firms Strategically Disseminate? Evidence from Corporate Use of Social Media," *SSRN,* 2016.