2017

Analysing the Behaviour of Online Investors in times of Geopolitical Distress: A Case Study on War Stocks

James Usher
Technological University Dublin

Pierpaolo Dondio
Technological University Dublin, pierpaolo.dondio@tudublin.ie

Recommended Citation

Usher, J. & Dondio, P. (2017). Analysing the Behaviour of Online Investors in times of Geopolitical Distress: A Case Study on War Stocks. *Proceeding WI '17 Proceedings of the International Conference on Web Intelligence*, pg. 275-283. doi:10.1145/3106426.3106510

Analysing the Behaviour of Online Investors in times of Geopolitical Distress: A Case Study on War Stocks

Pierpaolo Dondio
DIT School of Computing
Kevin Street, Dublin 8, Ireland
pierpaolo.dondio@dit.ie

James Usher
DIT School of Computing
Kevin Street, Dublin 8, Ireland
james.usher@student.dit.ie

ABSTRACT
In this paper we analyse how the behavior of an online financial community in time of geopolitical crises. In particular, we studied the behaviour, composition and communication patterns of online investors before and after a military geopolitical event. We selected a set of 23 key-events belonging to the 2003 US-led invasion of Iraq, the Arab Spring and the first period of the Ukraine crisis. We restricted our study to a set of eight so called military stocks, which are US-manufacturing companies active in the defence sector. We studied the resilience of the community to information shocks by comparing the community composition, its sentiment and users' communication networks *before* and after an *event* at different time intervals. We found how community reaction is governed by ordered patterns. Experimental evidence suggested how in the aftermath of an event the community does not lose its *information sharing* functionality. Communication networks show a higher in-degree Gini index, connectivity and a rich-club effect. Discussions tend to develop around central users acting as hubs. These backbone users correspond to rich-club users, present both before and after an event, whose sentiment is less volatile than other users and that were previously recognized as local experts of a specific stock. As further evidence of community resilience, the equilibrium of all the indicators analysed is restored after two weeks.

CCS Concepts
• **Human-centered computing → Collaborative and social computing → Collaborative and social computing theory, concepts and paradigms → Social media • Information systems → World Wide Web → Web mining**

Keywords
Behavioral Finance, Social Media, Web Mining, Content Analysis.

1. INTRODUCTION
The growing prominence of the Internet and social media has radically changed the way in which investors seek and share information (e.g., [1][2][3]). Apart from prices, official statistics, and reports, investors increasingly use e-communication platforms for investment ideas [4]. Recent estimates indicate that in 2011, there were about 21 million online investors in Europe and 30 million in the United States. While studies on the sociology of finance have mainly focused on large institutions and professionals [14][15], little is known about online investors. Recently, online communities about finance have received a growing attention in the professional and academic environment.

Such online communities operate on a large set of online platforms such as message boards, online fora and twitter-like specialized communities. By sharing and creating knowledge in peer-to-peer decentralized way, these virtual places nurture collective intelligence in complex and largely unpredictable ways.
Prior studies have focused on the predictive power of such communities, testing if the collective sharing of sentiment and information could be proficiently used to anticipate market movements. Despite findings reported in literature are partially conflicting, the consensus is that the messages exchange on media platforms are not just noise: in fact the number of messages could help to predict volatility, while disagreement is usually associated with higher trading volumes [4][6]. Few other studies reported interesting predictive results [7], with an accuracy of about 75-80% for short periods of time.
The study presented in this paper belongs to a smaller line of studies focused on the analysis of the behavior, composition and dynamics of such communities, and the quality and typology of the information shared.
A recent article by [10] analyzed the resilience of the largest Italian online community about trading during the recent financial crisis. Results show that while less expert user behavior changed during the crisis evolution becoming more sensitive to external shocks, expert users typically pondered their response to shocks both before and during the crisis. Findings indicate that these communities could be viewed as social laboratories where self-organization of decentralized collective intelligence systems takes place with relevant implications for financial markets.
The study presented in this paper aims to analyse the behavior of users of the popular message boards Yahoo! Finance during periods of geopolitical instability, providing experimental evidence about the resilience of the community. We wondered if a community of investors is resilient enough to maintain its functionalities in times of uncertainty and market turbulence.
Resilience is the ability of a system to maintain its functionalities despite disturbances and changes, and it is essential to system survival. [11] Discussed three main interpretations of resilience. The first one, called engineering resilience, refers to how fast a system returns to its previous equilibrium after a shock. Secondly, ecological resilience is meant as the magnitude of the disturbance a system can absorb before switching to a different form or state. Finally, adaptive resilience is the ability of a system to change its structures, modes of operation and components in order to maintain its functionalities.
Our study focused on a subset of Yahoo! Finance message boards represented by the discussions about the so-called "*war stocks*", a set of US manufacturing companies active in the defense and Aerospace sector. We analyzed the behavior of users around critical events belonging to recent political crisis, namely the US-led invasion of Iraqi of 2003, the 2010 Arab Spring and the 2014 Ukraine crisis outbreak.
We collected messages posted by users at different intervals before and after an event, and we analyzed the behavior of the online community by looking at the composition of the community, the communications patterns between users and the sentiment expressed by users.
We note how previous studies, such as [1,8, 9,10] used volatility indexes to identify periods of market shocks (therefore relying on an endogenous market indicator), while here we analyzed the community reaction to an exogenous shock represented by a geopolitical event. Indeed, the two situations might coincide, since geo-political events are potentially market-sensitive.

Interestingly, this is not the case. Data showed how our analysis is complementary to a volatility-driven approach, since in proximity of the critical events considered the market alternated few periods of uncertainty with the majority of periods of stability.

The paper is organized as follows. The next section describes the datasets used, while section 3 describes the experimental methodology followed. Analysis of the results is presented in section 4, while section 5 describes related works to date before our conclusions and future works section.

2. DATASET(s)

2.1 Stocks

We considered a set of eight stocks active in the aerospace and defence sector. They are medium and big capitalisation manufacturing companies with a strong R&D department. The stocks are listed in table 1. The column *capitalisation* is the capitalisation of each stock in billions of dollars, while the percentage in parenthesis is the relative size of each stock over the capitalisation of all the eight stocks. The war stocks had a total capitalisation of about $370 billion, which represents about 2.5% of the capitalisation of the S&P500 index, estimated at about $15 trillion at the end of the period covered by this study (July 2014).

Table 1 Stocks considered in the study

Stock	**Ticker**	**Capitalisation**
Honeywell International	HOC	74.7B $ (20.18%)
United Technologies	UTX	100.34B $ (27.11%)
L-3 Communication Holding	LLL	9.42B $ (2.54%)
Lockheed Martin Corporation	LMT	55.7B $ (15.05%)
Alliant Technologies	ATK	4.1B $ (1.10%)
Northrop Grumman Corp.	NOC	26.5B $ (7.16%)
Raytheon Co.	RTN	30.1B $ (8.13%)
Boeing	BA	69.55B $ (18.79)

2.2 Geo-political events

We collected a number of geopolitical events related to three geopolitical crises of recent years. Events are supposed to trigger a reaction on the online community of investors, especially for war stocks.

We have collected key events grouped in three major crisis: the Iraqi invasion started in 2003; the Arab spring (from December 2010 to the end of 2012) including events from Syria civil war, Libya, Tunisia and Egypt, and the outbreak of the Ukraine crisis (from November 2013 till August 2014). The inclusion criteria for the events were the following. We considered all the events reported on the timeline of events of the British Broadcasting Corporation (BBC) News about Iraqi invasion, Arab spring and Ukraine crisis. We then selected only events related to declaration of wars, terrorist attacks, and protests escalated in deaths or casualties. In order to be included events must represent the inception/outbreak of a geo-political crisis or an unexpected key events (such as the starting day of the Iraqi invasion, the capture of S. Hussein, the killing of M. Gaddafi or the *EuroMaidan* protest starting the Ukrainian crisis), since we assumed those news items would potentially generate the greatest impact on market and investor behavior. Appendix one contains the list of events considered in this study.

In the remaining of the paper, we refer to the *entire period of observation* as the period from 2003 to 2014, while we refer to the *war period* as a time interval including the three geopolitical crisis considered: the Iraqi war (October 2002 – December 2003), the Arab Spring (December 2010 – October 2012) and the Ukraine crisis (November 2013 – May 2014).

2.3 Online Communities Data

We collected data about users' activities on Yahoo! Finance Message Boards from 2003 till 2014. Yahoo! Finance keeps a message board for each stock quoted on the US market. Each message board is a stream of threads opened by registered users. Each thread is a stream of messages posted by users. A user can decide to add a new message to a thread, answer to an existing message or open a new thread.

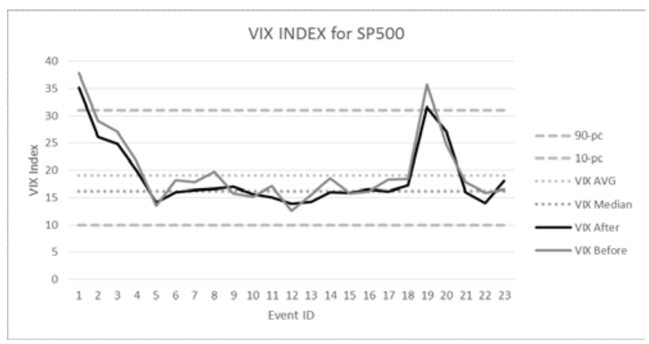
We collected data about discussions regarding the eight war stocks of interest from 2003 till July 2014. We gathered the list of threads, the list of messages for each thread, the content of each message, time of the message, users and the citations between users (i.e. if an user replied to another user). There were approximately 798,000 messages regarding the eight stocks examined, written in about 73,500 threads by about 18,500 users.

2.4 Market Data

We collected historical prices for the eight war stocks and the S&P500 index. The S&P500 was also used as the market benchmark. Closing prices adjusted by dividends and splits were collected via a Bloomberg terminal. In order to measure volatility, we collected the daily value of the VIX indicator.

We first wondered if the 23 events selected were associated with regimes of high volatility and if stocks returns in the immediate aftermath of an event were abnormal. If so, changes observed in users' behaviour could be attributed to an information shock (the geopolitical event) but to a market shock likely generated in response to the news shock. Data collected seems to corroborate the fact that the events selected are not associated with any specific volatile regime or high returns.

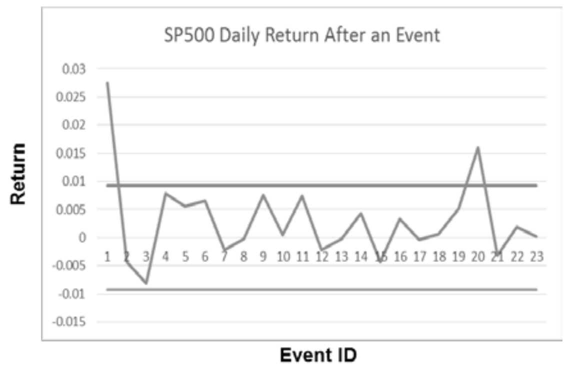
We first analysed the volatility. The market variance was measured considering the VIX implied volatility index. The VIX for the SP500, had a baseline average of 19.46 for all the period of observation, and 20.74 for the war period.



Graph 1. VIX index for the week after an event. Events are represented by an incremental ID. See Appendix A for a description of each event.

Graph 1 shows the implied volatility the day after an event (black line) and the day before the same event (grey line). The four horizontal dashed lines represent the 90-percentile and 10-percentile level, the average and the mean computed over the entire period of observation. Out of the 23 events, only 2 occurred in a situation of abnormally high VIX (above 90-pc level), 1 in the Iraqi war and 1 during the Arab spring. A further 4 events had a VIX in the first or last quartile, while the remaining 17 events showed an average VIX value. Therefore, the events collected for this study are not associated with high volatility but they happened in mixed volatility regimes. The Iraqi war had a higher than usual volatility, while the other two crises less than the usual but not significantly. The VIX graph for different periods is not displayed as it showed a similar pattern.

Interestingly, there is no statistical difference between the VIX before and after an event, meaning that on average the same regime was present both sides of the event. The VIX was higher before the event rather than after.



Graph 2 Returns one week after the event. The horizontal lines represent (from top to bottom) the 90-percentile and 10-percentile of the distribution of the returns over all the period of observation.

Regarding returns in proximity of an event, the following graph 2 shows how daily returns are distributed in the 10-90 percentile bands, except for two.

We can conclude that volatility and returns of the market around the events analysed in this study are usually not abnormal. Apart from two cases, the market did not consider the event a shock. The eight stocks considered had a behavior consistent with the market index (graphs omitted). Since we identified 23 events and 8 stocks, and since we analysed the status of the community at 8 different time intervals d around the time of the event (d=±1, ±5, ±10, ±20 days), we collected a total of 1472 observations. However, all the observations that are overlapping with observations associated to another event are discarded. The situation is possible since some events are less than 40 days apart, meaning that the observation after 20 days overlapped the observation 20 days before the subsequent event. This reduced the number of observation to 1358. In about 7.5% of these observations either the stock returns or the stock volatility for the period of observation were abnormal (in the 90-pc or 10-pc region). These are 102 observations distributed mainly in the aftermath of an event, as shown in table 1.

Table 1 Number of observations associated to abnormal volatility or returns. For each time interval there are 184 observations. One day after an event there were 35 observations associated with high volatility, representing 19.01% of the total number of observations.

d (days)	**-20**	**-10**	**-5**	**-1**	**1**	**5**	**10**	**20**
Abnormal Observation	2	5	12	15	35	23	10	5

Since previous studies provided evidence that market conditions – especially volatility regimes – do effect community behavior, we divided our observations in two groups: the first including the observations associated with *normal* returns, and the second smaller dataset composed by 102 observations associated with abnormal market conditions, called *high VIX* dataset. In the first dataset any changes in users' behaviour cannot be explained as a reaction to market movements, as usually referred in literature, but it could be attributed to the informative shock of the event. This makes the present study complementary to the mainstream analysis where different volatility regimes were identified as the main source of changes in users' behaviours. Moreover, in section 4 we show how in general the statistical significance of the results do not change in the two datasets.

3. Data Processing

3.1 Users' contribution, sentiment, stability

We introduced a set of features to describe the status of the online community. We measured the communication levels of the community, its sentiment and its composition. The features are computed for a specific time interval and associated to a specific event and stock. We considered:

- $M^{\pm d}$: The number of messages posted by all the users at a specific time interval, and $\overline{M^{\pm d}}$, the average number of messages by users. The indicators are a better representation of the overall contribution of each user to the discussion
- J^d: The jaccard similarity between the set U^d of active users in (i.e. posting a message) at day d and the set $U^{d+!}$ users active at day $d+1$. Jaccard similarity is defined as follows:

$$J^d = \frac{|U^{d+1} \cap U^d|}{|U^{d+1} \cup U^d|}$$

We use J^d to measure the stability of the community at a specific day. We call the users present in $U^{d+1} \cap U^d$ *persistent* users, and *volatile* users the complementary set $(U^{d+1} \cap U^d) \setminus (U^{d+1} \cup U^d)$.

- $p^{\pm d}, n^{\pm d}, s^{\pm d}$: p and n are the number of messages containing a positive or negative sentiment in a specific time period. We also defined an overall sentiment indicator as follows: $s = \frac{p-n}{p+n}$, representing the polarity of user sentiment in a period of time. The indicator s will be used to identify potential change in users' sentiment before and after an event.

3.2 Network of Investors: SNA metrics

Using Yahoo! Finance data we defined a network for each stock and a specific day d. The nodes of the networks were represented by users posting a message in the interval of time, while an edge is drawn from node a to node b if user a quoted at least one message written by user b. The notation is the following. We call $\aleph_e^{-d}(x)$ and $\aleph_e^{+d}(x)$ the networks of online users for the stock x built considering all the messages about stock x posted at day $t_{e_i} \pm d$ where t_{e_i} is the timestamp of event e_i and d is the number of days. $\aleph_e^{-d}(x)$ represents the network of investors *before* geopolitical event e_i, while $\aleph_e^{+d}(x)$ the network *after* the event.

In order to describe each network $\aleph$, we considered the following metrics. For each indicator we describe the rationale for considering it and its meaning in our context.

1. Number of nodes N. The number of nodes represents the number of distinct users active in the period of observation
2. Number of edges E. There is an edge from user a to user b if a replied to b (at least once). The number of edges is a

measure of the interactions between users. Note how this differs from the number of messages, since users can communicate by posting a new message rather than quoting a previous one.

3. As is the assortativity of the network. It measures the preference of nodes to link to other similar nodes. We compute the assortativity w.r.t. the in-degree centrality of nodes. Therefore, in our context a high level of assortativity means that highly cited users have a preference to answer to highly cited users and vice versa.

4. RC, the rich-club coefficient [23] measures the tendency of highly connected nodes to link between themselves. The rich-club coefficient can be seen as a more specific version of the assortativity considering only nodes above a certain value of in-degree centrality. In our context, a rich-club effect means the presence of users acting as focal points for discussion (high in-degree) and linking multiple discussions (ties with other high in-degree users). As described in [23], we use the normalized value for the rich club $\rho(d)$ parametrized for the node in-degree d. A value of $\rho(d) > 1$ indicates that nodes with a degree greater then d are linked more than a random network.

5. CG, the Gini index [22] of the in-degree centrality. The Gini index is a measure of dispersion of a random variable that takes values from 0 to 1. A Gini index of 0 is obtained by a uniform distribution (situation of perfect equality), while a Gini index of 1 is obtained by a distribution concentrated in one value. In our context, we apply the Gini index to the distribution of in-degree values as shown in [29]. A low value of the Gini index means a quasi-uniform distribution while a high value of the index suggests the presence of hubs.

Note how networks considered are directed, and the above indicators are computed for directed networks. In section 4 we perform a comparison between a pair of networks $\aleph_e^{-d}(x)$ and $\aleph_e^{+d}(x)$. The comparison of two networks based on the above indicators has to be controlled by the size of the network. The number of edges, for instance, it is proportional to the number of nodes, and we therefore consider the average number of edges for each node $\overline{E_d}$. The other indicators were compared only if the before and after network had comparable number of nodes (users). As a rule of thumbs, we only compare networks if the number of nodes of the before network (N^{-d}) and the after network (N^{+d}) satisfies the interval: $0.8 \leq \frac{N^{+d}}{N^{-d}} \leq 1.2$

4. EVALUATION

Aim of our experiments was to verify if the communication dynamics of online investors are significantly modified in proximity of a geopolitical event. Due to the time-locality of our interest and the presence of two distinct situations (pre and post an event), we follow a basic statistical hypothesis testing approach. We computed the features described in the previous section and analysed the difference between the set of values before and after an event at different time intervals. Since none of the features is normally distributed, we use the non-parametric Mann–Whitney U-test [18] to check statistical significance.

4.1 Analysis of Community Daily Retention

We analysed how the composition of the community changed before and after an event using the Jaccacd similarityJ^d. Table 2 shows the value of the Jaccard coefficient between the set of users at different time intervals (at d=1, 5, 10, 20 days) and the baseline average of the online community, computed over the entire period of observation (2003-2014).

A first result is that the members of Yahoo! discussions are highly volatile: only a minority of users is present both before and after an event. The baseline Jaccard values are 0.117 corresponding to a retention rate of about 21%. The table shows also an interesting pattern. Even if still highly volatile, around geopolitical events the community is more stable than usual. Almost 40% of users are the same after one day of the event. The gap gradually decreases until the effect is lost after 10 and 20 days where Jaccard values are comparable to the baseline ones. Therefore, there is a tendency of users to continue a discussion in the immediate aftermath of an event more than usual. If we consider the events associated to high VIX, the data shows a similar but stronger effect. One day after an even the retention is over 42.1% a value that is significantly higher than the military events dataset for $d = 1$ only.

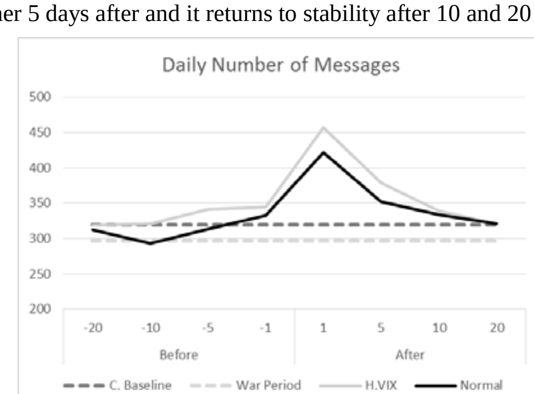
Table 2 Jaccard Similarity between the set of active users in a discussion about military stocks before and after a geopolitical event at different time intervals. The symbols ***,**,* refers to 0.99, 0.95 and 0.9 statistical significance.

d	J^d Normal	J^d Com. Baseline	J^d High VIX
1	0.244 (39.3%) ***	0.117 (20.92%)	0.266 (42.1%)
5	0.191 (32.07%) ***		0.207 (34.12%)
10	0.153 (26.6%) **		0.165 (28.4%)
20	0.122 (21.8%)		0.127 (22.57%)

4.2 Analysis of Users contribution

The analysis of users' stability revealed a high volatile community with a tendency to be more stable in the aftermath of a military event. We then measured the amount of users' contribution before and after an event, testing if the minority of persistent users is the major contributor to the forum. We computed the daily number of messages exchanged on the military stock boards. The community baseline was 319 messages per day against 296 of the war period.

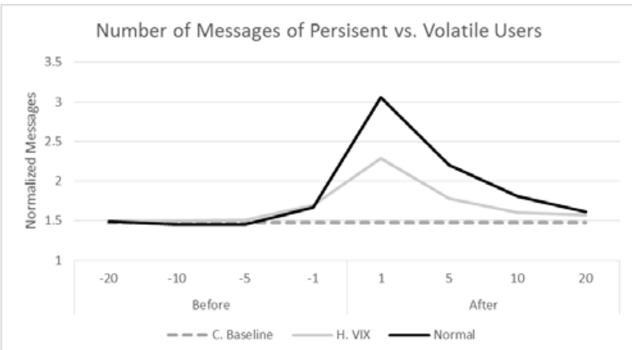
Graph 3 shows the number of messages before and after an event. Messages number spikes significantly one day after an event (a 40% increment w.r.t to the average of the war period); it is still higher 5 days after and it returns to stability after 10 and 20 days.



Graph 3 Number of messages posted by users *before* and *after* a geopolitical event at different time intervals. Note how messages spike in the immediate aftermath of an event.

We then focused on the contribution of *persistent* users. Graph 4 shows the number of messages by *persistent* versus *volatile* users.

On average, persistent users wrote more than volatile users (the baseline value is 1.5 message for each message of volatile users), After the event ($d = 1$ and $d = 5$), persistent users wrote respectively 3.06 and 2 times more than volatile users. For d=1, the persistent users contributed to about 68% of the total conversation. Persistent users are driving the post-event discussion. The effects return to stability after 20 days. Regarding the high volatility data, it is interesting to notice that the effect is in the same direction of the events dataset, but this time weaker. Persistent users wrote a similar number of messages in both the situations, however in the high VIX dataset volatile users are contributing more and the ratio is diminished.

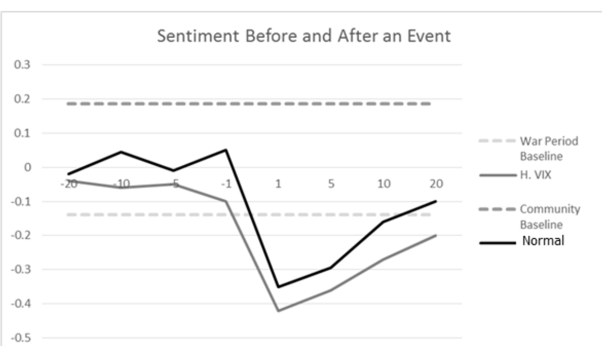


Graph 4 Number of persistent users messages over the number of messages of volatile users.

4.3 Analysis of Users sentiment

We computed a sentiment score for each message board's posts using the Sentiment API Standfore NLPCore, assigning each message to a positive neutral or negative category.

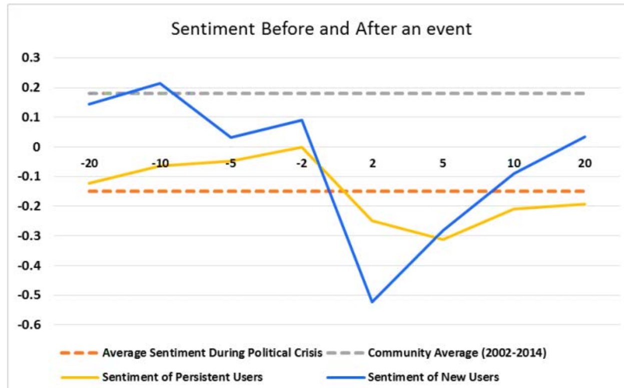
In general, the analysis of all the messages collected shows a bullish sentiment. We identified 45535 buy messages versus 31567 sell messages. The overall sentiment indicator has therefore a value of $s = \frac{s^+ - s^-}{s^+ + s^-} = 0.18$. This time the war period baseline had a significantly lower value of -0.14, meaning that our political events happened in situation of sentiment more negative than usual. The gap between the two baselines is due to a strongly positive community sentiment in the years 2006-2007 and 2009-2011.



Graph 5 Sentiment indicator before and after an event. There is a strong negative drop-down in the aftermath of an even that is quickly recovered in the following days

Graph 5 shows the sentiment before and after an event, at different time interval (d=±1, ±5, ±10, ±20 days). The graph shows how users' sentiment is slightly positive before an event, and it lowers afterwards. The drop-down is particularly violent in the immediate aftermath of an event (d=1) and it gradually becomes less negative in the following days. After 20 days the sentiment returns above the average of the *war period*. Users are affected by the event, but after a short period of strong negative sentiment they quickly recover to a more favourable outlook, that could suggest a community resilient to the shock.

If we compute the sentiment expressed by *persistent* users versus *volatile* users, we obtain the situation depicted in graph 6. Interestingly, persistent users have a less extreme sentiment than volatile users. Volatile users exhibited a highly polarized sentiment. There was a strong positive sentiment before the event, where they tended to be more optimistic than persistent users, while after the event new users entering the discussion had a stronger negative sentiment, which quickly recovers after 20 days.



Graph 6 Sentiment indicator before and after an event for persistent and volatile users. Persistent users' sentiment exhibits a lower deviation

Analysis of SNA metrics

We defined networks of investors as described in section 3.2, defining 679 before and after networks. We first considered the number of nodes and edges of the *before* and *after* the event. Table 2 shows how the average number of nodes and edges for the networks *before* and *after* an event at different days *d*. The column T is the result of a UW statistical test. The after network is generally a bigger network with more users involved in the discussion and with a higher number of edges. Note how the average number of edges is also significantly higher. One day after we have an average of 1.72 versus 1.39 in the before network and a baseline of 1.305.

Table 2 Average number of Nodes and Edges for the before and after networks at different intervals.

	d	After	Before	Diff.	T *n*=157	Baseline
Nodes	1	31.20	24.05	0.297	+	23.87
	5	28.10	24.30	0.156	+	
	10	25.40	24.20	0.050	=	
	20	24.40	24.01	0.016	=	
Edges	1	53.48	33.43	0.600	++	31.15
	5	45.96	32.81	0.401	+	
	10	39.10	32.43	0.206	=	
	20	34.10	31.45	0.084	=	

The degree of Persistent Users

We checked the role of the persistent users in the communication dynamics by computing their average in-degree. Data shows how not only persistent users contribute more to the discussion, but they are also the most central nodes of the network. The VIX dataset has similar result for persistent users, while a slightly higher values for volatile users. The two columns WU is the result of a WU statistical test between the average degree of persistent versus volatile users of the normal dataset.

Interestingly, volatile users are more cited in the regime of high volatility, while persistent users are cited less (but not significantly) than the situation of low volatility

Table 3 The average in-degree centrality of volatile and persistent users before and after an event at different time intervals. The column WU is the result of a statistical test between the average degree of persistent versus volatile users for the *normal* dataset.

	Normal			Hvix	
d	Persistent	Volatile	WU	Persistent	Volatile
1	3.01	0.87	***	2.46	1.17
5	2.72	1.12	***	2.32	1.28
10	1.76	1.35	***	1.68	1.37
20	1.41	1.39		1.35	1.41

Before vs. After Networks

We performed a more in-depth analysis of a subset of networks. We measures the rich-club effect, assortativity and the Gini index of the in-degree distribution/. However, since some of the daily network resulted too small for those metrics to have a meaning, we decided to filter the set of networks by only considering the three most discussed stocks: BA, LLL and UTX. As explained in the previous section, the SNA metrics should be compared only if the two networks have a similar number of nodes. Following a rule of thumb we compare only pair of networks with less than 20% difference in the number of nodes, decreasing the number of networks from 552 to 476.

Table 4

Feature	*d*	After	Before	% Diff	T $n = 59$
nodes	1	45.36	35.83	0.297	+
	5	39.58	33.71	0.063	=
	10	39.13	31.39	0.037	=
	20	34.5	32.71	0.023	=
edges	1	104.78	71.66	0.462	++
	5	87.47	64.40	0.334	+
	10	75.52	56.51	0.264	+
	20	62.1	57.25	0.151	=
Gini index of in-degree distribution	2	0.587	0.49	0.136	=
	5	0.553	0.45	0.25	+
	10	0.48	0.423	0.046	=
	20	0.42	0.388	-0.02	=
Assorativity	2	0.102	0.225	0.055	=
	5	0.154	0.236	0.235	+
	10	0.18	0.242	0.121	=
	20	0.198	0.221	0.118	=
Rich Club (5)	2	1.792	1.46	0.070	+
	5	1.803	1.236	0.147	+
	10	1.34	1.305	0.153	+
	20	1.25	1.267	-0.040	=
Rich Club (6)	2	1.815	1.449	0.016	=
	5	1.733	1.288	0.210	+
	10	1.405	1.265	0.189	+
	20	1.212	1.300	-0.095	=

Table 4 reports, for each SNA feature and each time interval, the value of the before and after network along with the increment (in percentage) and the output of the statistical test. The absolute value of each indicator is important in order to understand not only if the difference is significant, but also if the absolute value of the indicator suggests a not negligible effect.

Overall, there are several statistical differences, and in general they are positive, meaning that the value of the SNA metrics increased in the after network. Statistical significance disappeared 10 or 20 days after the event. Assortativity is the only metrics to show a different trend, smaller in the after network. The value is still positive even if negligible.

For the rich club coefficient we have reported the value corresponding to a value of the in-degree equal to 5 and 6. A value greater than one suggests the presence of a more-then-random rich club effect. The rich club effect is present in all the periods of observation, and it is stronger in the after network for d=1 and d=5.

Regarding the analysis of the networks in regime of high VIX, the small number of networks is not enough to generalize the results. In general, the networks associated to high volatility showed the same behaviour. We only report the metrics for which the networks have values different from the *normal* dataset. In the VIX networks the assortativity increased to 0.281 for d=1 and therefore resulted significantly higher than the ME case, while the Gini index had smaller values compared to the ME, (0.55 for d=1 compared to 0.587 for the normal dataset).

What is in the Rich club?

The networks after the event (d=1 and d=5) shows a high rich club effect, meaning that the top-cited nodes have a tendency to cite themselves. We tried to understand better who the members of the Rich Club are. On average, networks have about 45 nodes and 7-8 users are in the rich club for dg=5 and 6-7 for dg=6.

The following table 5 shows how 85% of Rich club members in the after network for d=1 are persistent users (76% for d=5), higher than the baseline of 68%. Note how in the VIX dataset persistent users are only 69% in the after network for d=1, meaning that there is a higher chance for a new entrant to enter the rich-club.

Table 5 Proportion of persistent users in the rich club (degree = 6).

d	Normal	High VIX	BASELINE
1	85%	69%	68.1
5	76%	66%	64.1

In order to better understand the features of the member of the Rich-club in the aftermath of an event, we performed a logistic regression to describe the binary variable$rich$, that is 1 when a member is part of the rich club and 0 otherwise. The dependent variables selected for the regression – listed in table 6 – describes the past behaviour of the users in terms of activity sentiment and impact.

Table 6. Independent variables used in the regression model.

Var	Description
p	Boolean: 0=persistent user, 1= non persistent user
$m_{y!}$	Total number of messages on Yahoo! MB. u = 100 messages
m_s	Total number of messages on stock s. 1u = 100 messages
c	Percentage of citations over total number of messages. 1u = 1%
l	Longevity. Number of days since registration. 1u=100 days
s	Average sentiment of the user. 1u = 10*s

The results of the regression are presented below:

Chi Square= 600.8081; df=6; p= 0.0000				
Variable	Coefficient	Std. Err	P	Odds Ratio
p	-0.5309	0.464	0.2526	0.588075
$m_{y!}$	0.0786	0.0578	0.1737	1.081772
m_s	0.4111	0.1259	0.0011	1.508476
c	0.7425	0.0808	0	2.101182
l	0.0843	0.1305	0.5182	1.087955
s	-0.3045	0.2222	0.1705	0.737492

The overall model fits significantly. By looking at the coefficients, sentiment and persistent has a negative effect. The two significant coefficients are number of messages on the stock and proportions of citation, both with positive value. Holding all the other coefficient fixed, there is an increase of 110% of the odds of being into the rich club, while 100 messages more on the stock will increase it by 50% (versus 8% of increase for 100 extra messages on Yahoo! Finance. The results seem to suggest how rich club member are users expert of the stock, with high previous impact on discussions, and with a lower sentiment. Being persistent looks like a necessary but not sufficient conditions to enter the rich club.

Analysis of Results

Results show some constant trends. First, the large majority of the statistical significance between the community before and after and event is obtained immediately after a geopolitical event (d=1 or 5 days), and such difference returms to stability after 10 trading days.

In the aftermath of an event, the community is characterized by the following:

- More online users are talking about the war stocks, as evidenced by the increase in the number of nodes in the *after* network and users. New users are attracted to the discussion.
- The community retains users more than the baseline behavior, meaning that online users consider worthy to still look for information on the online community
- These users are interacting more, as evidenced by the high increase in the number of edges in the after network. This behaviour could be interpreted as an increased collective effort of the users to join strengths together and try to discuss and make sense of the consequences of the recent event. A similar effect was found in [10]
- There is an increase in the Gini index of the in-degree distribution and the rich club coefficient of the *after* networks. This means that in the *after* networks there are a group of actors with high importance in the network, acting as hubs and central point of reference during the discussion. This hubs communicates often among themselves, as suggested by the presence of rich-club effect.
- The assortativity of the networks, on average positive (weak effect around 0.2-0.25) is null after an event. This data suggest that there is more communications between hubs and nodes with low degree in the after network. This could suggest a pattern of question-answer from peripheral nodes to hubs. The presence of hubs, joint to the rich-club effect and a null assortativity suggested the presence of local conversations (sub-discussions) linked by the rest of the discussion by the presence of one or few hubs

The above results suggest the following summarization: the after network is usually a larger networks with more users interacting and more users contributing to the message boards, where multiple discussions are taking place around few focal users acting as hubs. Those hubs communicate among themselves.

In order to unveil more information about the knowledgeable users, we have introduced the concept of persistent and volatile users.

It turned out that this way of discriminating among users captured a clear difference in behaviour and communication patterns.

Persistent users exhibited a higher in-degree (meaning that they are posed questions); they contributed more and they kept the network connected. The analysis of user sentiment revealed how persistent users had a quite stable sentiment, meaning how the impact of the event did not alter significantly their opinions. The number of persistent users is bigger around war events than usual, and their individual contribution can reach 3 times the contribution of volatile users. On the other side, noisy volatile users leave the discussion before the event – implicit evidence of lack of interest – or join the discussion after an event – implicitly attracted by the consequences of such event. In any case, this set of volatile users do not drive the discussion but they join it by connecting to one of the persistent users, acting as the backbone of the discussion.

Regarding the composition of hub nodes, being a persistent user resulted an almost necessary but not sufficient reason to guarantee a place in the rich club. What counted more was the fact of being a local expert of the stock, with a large number of previous messages on the stock, with high impact on the discussion (measured by the number of citations) and a lower-than-average sentiment.

The large majority of our observations were not in a regime of market uncertainty. The very presence of statistical significant difference is a result on its own, and evidence that stock market movements are not the only stimulus triggering a reaction on online users. Even if we did not have a large enough number of observations associated to high volatility, results seem to suggest a similar reaction but generally stronger. It is easier for new users to become hub of the network, and networks show different pattern of communication (positive versus null assortativity, lower Gini index). However, we did not have enough observations to generalize our results and a comparison between situations of high versus normal market uncertainty is a line for future research.

In summary, the evidence collected depicts the online community as point of reference for information sharing. The community resulted sensitive to external events but it showed resilience and it kept its information sharing functionality by relying on local experts.

5. RELATED WORKS

One of the earliest studies investigating online financial community predictive power is the one by Antweiler and Frank [5]. The dataset used was 1.5 million posts from Yahoo Finance and RagBull, and the study covered 45 stocks of the Dow Jones Industrial Average. The authors' key conclusion was the following: the effect of stock messages helps predict market volatility, but the effect on stock return is statistically significant but economically moderate. Disagreement among posted messages is correlated with increasing trading volume. Further studies have been performed by Spiegel et al. [12] over the effect of rumours on stock return. Similar to our methodology, rumours are represented by press news, recommendation and indications coming from financial portal. The dataset was composed by 958 Israeli stocks monitored for 27 months using a set of about 2000 rumours. The work by [7] investigates the predictive power of Twitter's messages. Authors extracted from about 10 million tweets' text 7 indicators of mood using *OpinionFinder* and *GPMOS*. Using a Granger causality analysis, authors correlate DJIA values to GPOMs and OF values of the past n days to obtain 83% accuracy. The author reports that calm, other than sentiment polarity better predicts the market.

Behavior of online investors has been previously investigated in [16], where author suggested how these new investors are largely non-professional or casual investors more inclined to trade more often and aggressively. This would be a consequence of the fact that they have less information about markets than professionals, are more subject to over-confidence bias [17], and take more sub-optimal decisions, such as *buying high and selling low* [15].

[10] Measured the daily activity of users of an online financial forum from 2003 to 2012, therefore including a period with high market uncertainty and applied data mining techniques to measure user expertise in order to capture both population aggregate dynamics and more micro-based events. Results showed that, thanks to high variety of population dynamics, the forum played different roles depending on the stages of the crisis: sharing news during crisis outbreak, sharing more technical analysis during crisis progression. [9] Analysed the impact of bad news on a network of users exchanging information on *Unicredit* stock, and found that the network reacted differently to bad news depending on the volatility regime. Furthermore, knowledge generated by online investors helped to predict Unicredit stock volatility.

[8] Presented a survey on a sample of online investors in a financial virtual community. Authors found that knowledge sharing and learning in virtual communities cannot compensate for the financial education-gap of these investors. Results also showed that online exposure tends to increase investors' propensity for risk, which in turn does not guarantee better portfolio performance.

Even if the effect of information shocks has not been deeply investigated, there is a plethora of literature on the effect of geo-political instabilities on financial market. We report some interesting references to the reader. [24] Investigated the economic impact of conflict, using the terrorist conflict in the Basque Country as a case study. The research shows that Basque equities perform significantly better than their non-Basque equity counterparts once the 1998 cease file started. However overall Basque equities showed a negative performance metric in comparison to non-Basque stocks over this period. [25] views terrorism and military events as a form of economic warfare. Their research focuses on market reaction to Palestinian assassinations and reactions to terrorist attacks. Equity percentage changes are monitored on the Tel Aviv 25 index. The exchange shows a drop of 1.1% where an assassination attempt is made on a Palestinian political or high ranking terrorist figure. The research also shows that the negative reaction does not reverse itself for the next two weeks. Interestingly this point concurs with our findings regarding online financial community behavior where statistical effects begin to fade after the end of the second week.

[27] Reveals that stocks prices with fall dramatically in countries where there is a highly level of democracy and disposable income. [28] Looked at the effect of terrorism and military events had on the Pakistan's Karachi stock exchange. Three internal markets were observed, namely the money market, the stock market and the FX market from the period dating December 2005 to June 2008. Each military and terrorist event was categorized. The findings showed that Pakistan's economy suffered badly from the terrorist activity and each internal market performed poorly during said incidents.

6. CONCLUSIONS

In this study we have analyzed how Yahoo! Message Board users behave in the aftermath of geo-political military events. By comparing the community composition, its sentiment and the communications networks among users before and after an event at different time interval we studied the resilience of the community to information shocks. We found how community reactions are governed by ordered patterns. Experimental evidence suggested how in the aftermath of an event the community does not lose its functionalities of information sharing. Discussions tend to develop around central users acting as hubs. Communications networks show a higher rich-club effect and in-degree are less uniformly distributed. These backbone users correspond to users present both before and after an event, whose sentiment is less volatile than other users and their contributions to the discussion is much larger than the remaining part of the community. Those users are usually local expert of the stock that contributed in the past with a large number of messages of high impact. The equilibrium of all the indicators analysed is restored after two weeks.

Future studies will include the analysis of a different topology of online communities, such as twitter-based platform, a more detailed text-mining analysis of users' contributions in order to understand the quality of the information shared. Our analysis, methodologically simple, turned out to be effective to identify a set of results large enough to reveal interesting patterns in community behaviour around an event. We believe the present methodology has provided a valuable set of results that further studies employing more sophisticated techniques could progress.

7. Appendix – Geopolitical Events

2003 Iraqi Invasion

ID	Day	News
1	11/10/2002	Congress votes resolution for attacking Iraq'
2	21/12/2002	President Bush approves the deployment of U.S. troops to the Gulf region
3	28/01/2003	State of the union speech: attack even without UN
4	19/03/2003	The USA begins invasion of Iraq
5	10/04/2003	Fall of Baghdad
6	01/05/2003	Mission accomplished
7	14/12/2003	Saddam captured

Arab Spring

ID	Day	News
1	17/12/2010	Tunisia. Mohammed Bouazizi sets himself on fire outside the local municipal
2	14/01/2011	Tunisia. Ben Ali resigns
3	25/01/2011	Egypt's Day of Revolt. First coordinated mass protests in Cairo as Egyptians demand Mubarak to resign.
4	11/02/2011	Friday of Departure. Vice President Omar Suleiman announces that the president has resigned and the army council will run the country
5	17/02/2011	Libya's Day of Rage. Dozens are killed as demonstrations erupt in cities across the country,
6	15/03/2011	Activists call for a Day of Rage across Syria, inspired by other popular uprisings across the Arab world
7	24/03/2011	Libyan plane destroyed by jets French fighter jets have

		destroyed a Libyan plane in the coastal city of Misrata in the first enforcement of the no-fly zone imposed by the UN to try to halt Muammar Gaddafi anti-rebel offensive.
8	9/10/2011	Maspero Massacre. Dozens die in Egypt, as violence errupts between protesters and the army during Coptic Christians' protests against the destruction of a church.
9	20/10/2011	Gheddafi killed
10	11/09/2012	US consulate in Libya attacked and ambassador is killed

Ukraine Crisis

ID	Day	News
1	30/11/2013	Police attacks on Euromaidan protesters
2	18/02/2014	Clashes erupt, with reasons unclear: 18 dead, followed by 88 people two days after on the 20th of Febraury
3	01/03/2014	Russia's parliament approves Putin's request to use force in Ukraine to protect Russian interests. Ukraine's interim PM Yatsenyuk says Russia has effectively declared war.
4	18/03/2014	Russia annexed Crimea
5	22/04/2014	Ukraine's acting president orders the relaunch of military operations against pro-Russians in the east.
6	02/05/2014	Clashes in the Black Sea city of Odessa, leave 42 people dead, most of them pro-Russian activists.

8. REFERENCES

[1] Sassen S. (2005) ‘The Embeddedness of Electronic Markets: The Case of Global Capital Markets’, in The Sociology of Financial Markets, edited by Knorr Cetina K. and Preda A., New York, USA: Oxford University Press.

[2] Preda A. (2009a) ‘Brief Encounters: Calculation and the Interaction Order of Anonymous Electronic Markets’, Accounting, Organizations and Society, 34, pp. 675-693.

[3] Knorr Cetina K. (2005) ‘How Are Global Markets Global? The Architecture of a Flow World’, in The Sociology of Financial Markets, edited by Knorr Cetina K. and Preda A., New York, USA: Oxford University Press.

[4] Tetlock P.C. (2007) ‘Giving Content to Investor Sentiment: the Role of Media in the Stock Market’, The Journal of Finance, 62, pp. 1139-1168

[5] W. Antweiler and M.Z. Frank, 2004. Is all that talk just noise? The information content of internet stock message boards. Journal of Finance, 59: 1259-1295

[6] P. Dondio, Predicting Stock Market Using Online Communities Raw Traffic, in the proceedings of IEEE/ACM/WI International Conference on Web Intelligence 2012, Macau, China

[7] J. Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. Journal of Computational Science, In press, 2011.

[8] R. Casarin, N. Casnici, P. Dondio, F. Squazzoni, Back to basics! the educational gap of online investors and the conundrum of virtual communities, Journal of Financial Management, Markets and Institutions 3 (1) (2015).

[9] N. Casnici, P. Dondio, R. Casarin, F. Squazzoni, Decrypting financial markets through e-joint attention efforts: On-line adaptive networks of investors in periods of market uncertainty, PloS ONE 10 (8) (2015).

[10] P. Racca, R. Casarin, F. Squazzoni, P. Dondio, Resilience of an online financial community to market uncertainty shocks during the recent financial crisis, Journal of Computational Science, Volume 16, September 2016, Pages 190-199, ISSN 1877-7503

[11] R. Martin, P. Sunley, On the notion of regional economic resilience: conceptualization and explanation, Journal of Economic Geography.

[12] Spiegel U., T. Tavor, J. Templeman, 2010. "The effects of rumours on financial market efficiency," Applied Economics Letters, Taylor and Francis Journals, vol. 17(15), 1461-1464.

[13] T. O. Sprenger, A. Tumasjan, P. G. Sandner, I. M. Welpe, Tweets and trades: The information content of stock microblogs, European Financial Management 20 (5) (2014)

[14] Beunza D. and Stark D. (2003) ‘The Organization of Responsiveness: Innovation and Recovery in the Trading Rooms of Lower Manhattan’, Socio-Economic Review, 1, pp. 135-164.

[15] Bloomfield R., Libby R. and Nelson M.W. (1999) ‘Confidence and the Welfare of Less-Informed Investors’, Accounting, Organizations and Society, 24, pp. 623-647.

[16] Barber B.M. and Odean T. (2002) ‘Online Investors: do the Slow Die First?’, Review of Financial Studies, 15, pp. 455-487

[17] Griffin D. and Tversky A. (1992) ‘The Weighing of Evidence and the Determinants of Confidence’, Cognitive Psychology, 24, pp. 411-435.

[18] Mann, Henry B.; Whitney, Donald R. (1947). "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other". Annals of Mathematical Statistics 18 (1): 50 -60. doi:10.1214/aoms/1177730491.

[19] Newman, Mark EJ. "Modularity and community structure in networks."Proceedings of the national academy of sciences 103.23 (2006): 8577-8582.

[20] Freeman, Linton C. "Centrality in social networks conceptual clarification."Social networks 1.3 (1978): 215-239.

[21] Borgatti, Stephen P., and Martin G. Everett. "A graph-theoretic perspective on centrality." Social networks 28.4 (2006): 466-484.

[22] Gini, Corrado. "Variabilità e mutabilità." Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi 1 (1912).

[23] Zhou, Shi & Mondragón, Raúl J. (2004). "The Rich-Club Phenomenon In The Internet Topology". IEEE Communications Letters. 8: 180-182. doi:10.1109/lcomm.2004.823426.

[24] Abadie, A. and Gardeazabal, J. (2003) The economic costs of conflict: A case study of the basque country.

[25] Zussman, A. and Zussman, N. (2006) Assassinations: Evaluating the effectiveness of an Israeli counterterrorism policy using stock market data. The Journal of Economic Perspectives 20(2) 193–206.Research and Documentation Online. (2014).

[26] Arin, K.P., Ciferri, D. and Spagnolo, N. (2008) The price of terror: The effects of terrorism on stock market returns and volatility.

[27] Karolyi, G.A. and Martell, R., 2005, Terrorism and the stock market, Ohio State University

[28] Gul, T., Hussain, A., Bangash, S. and Khattak, S.W. (2010) Impact of terrorism on the financial markets of Pakistan. European Journal of Social Sciences 18(1) 98–108.

[29] Badham, Jennifer M. "Commentary: Measuring the shape of degree distributions." Network Science 1.02 (2013): 213-225.