GNIP | Whitepaper

Social Media in Financial Markets:
The Coming of Age...

CONTENTS

ABSTRACT

This whitepaper is an update to our previous paper "Social Media in Markets: The New Frontier", which provided an overview of the emerging use of social media data as part of the investment process and outlined specific use cases in research and trading. This paper "Social Media in Financial Markets: The Coming of Age..." provides an updated perspective on the industry lifecycle and rapid advances in social media analytics in markets. In this whitepaper, we look at updates in academic research, consider the lag between the growth of social media analytics in the financial services industry compared to the brand/marketing industry, we highlight use cases for potential alpha generation, and look at emerging trends for 2014.

SOCIAL MEDIA IN FINANCE – A BRIEF HISTORY

Social Media continues to expand at breakneck pace. According to Adobe, there are now 12 social media platforms with more than 100 million users[1]. Since Facebook was founded in 2004, the number of social profiles across all networks has grown to a staggering 5.7 billion plus[1]. People, businesses, governments and brands are using multiple social networks from micro-blogging platforms like Twitter and Google+, to rich blogging platforms like Tumblr, WordPress and Disqus, and location based platforms like Foursquare.

According to Adobe, there are now 12 social media platforms with more than 100 million users.

Since Facebook was founded in 2004, the number of social profiles across all networks has grown to a staggering 5.7 billion plus.[1]

Following Twitter's launch in 2006, progressive, entrepreneurial firms rapidly emerged and created insightful social data analytics. In addition to simply monitoring mentions directly through the social media platform, external data analytics proved to be a more robust and valuable way of extracting insights from social media conversations. The early use cases were around marketing and brand management. A plethora of technology start-ups began serving big brands like Coca-Cola, Walmart and Verizon and providing dashboard analytics tools for executives in marketing and PR. The new field of Social Media Monitoring and Social Media Analytics (hereafter "brand analytics") emerged to help brands understand what consumers were thinking in real-time and to complement the monitoring that brands were doing themselves. Some seven years later, use cases and analytics in the brand management space are far more advanced, and mainstream. Firms like IBM, Adobe and Salesforce are all now big players in the space.

The finance industry was slower to develop since professionals and pundits were less likely to share their investment opinions on social media. While many financial professionals started using their own Twitter accounts for following breaking news or sharing of articles, true data analytics use cases were initially few and far between. The early adopters of social data analytics in finance were a small band of hedge funds and high frequency traders (HFTs) – most very privately so. Derwent Capital was a notable exception, employing a Twitter only analytics trading strategy – but the firm folded after a few months.

The finance industry was slower to develop since professionals and pundits were less likely to share their investment opinions on social media.

However, in 2013 things changed. Notable news events and information continued to be released on Twitter, grabbing the interest of the financial community. Significant events in 2013 included:

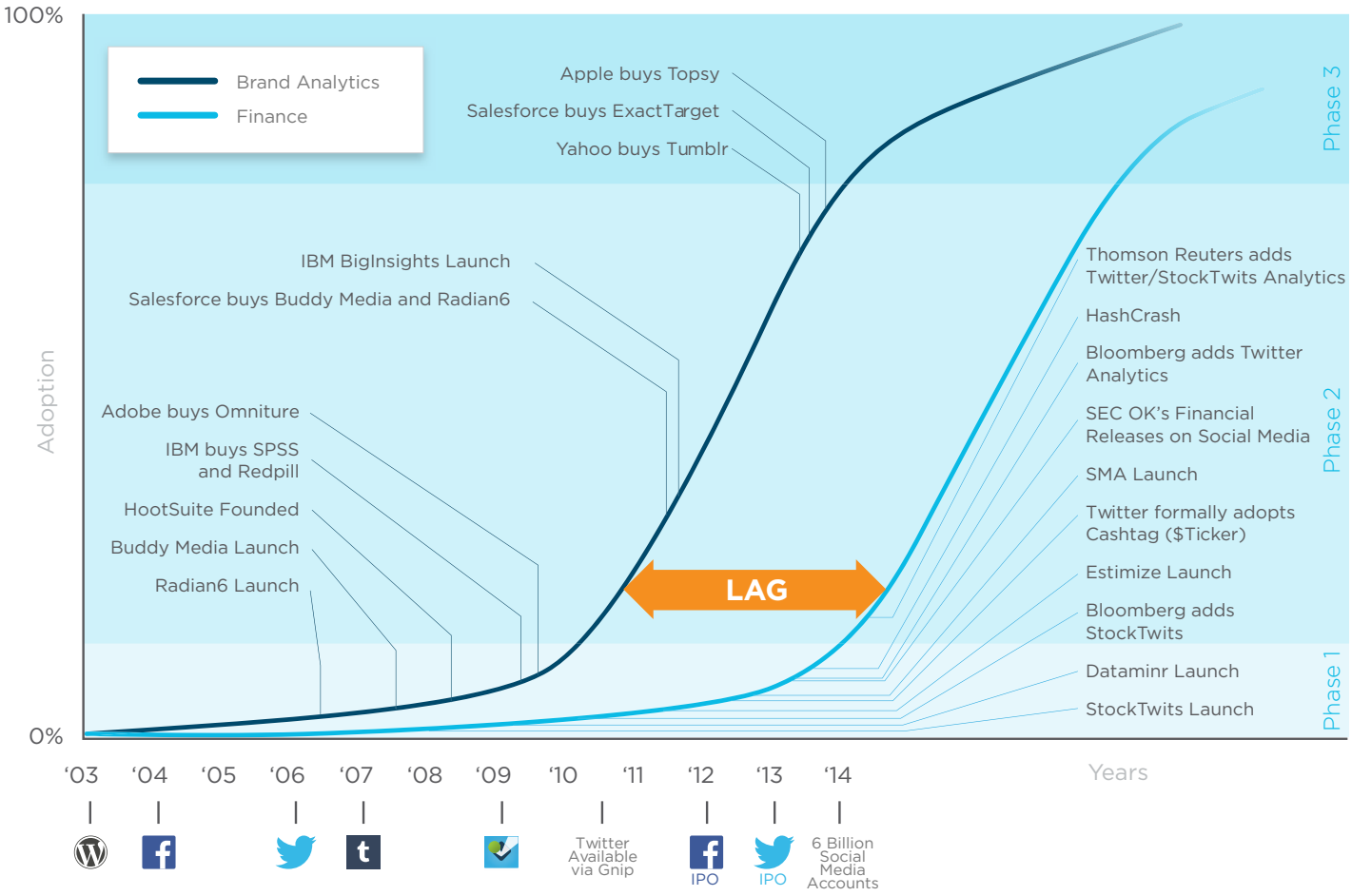
- The SEC confirmed that companies could use social media outlets to announce key information in compliance with Regulation Fair Disclosure;
- The Hash Crash in April drove 140 points off the Dow in two minutes (the AP Twitter account was hacked and tweeted about explosions in the White House);
- Iconic investor Carl Icahn tweeted "We currently have a large position in APPLE. We believe the company to be extremely undervalued. Spoke to Tim Cook today. More to come." causing the stock price to jump and add $12.5 billion to the market value[2].

In 2013 things changed. Notable news events and information continued to be released on Twitter, grabbing the interest of the financial community.

These events caught the eye of asset managers and financial professionals, and the last twelve months saw terminal industry behemoths Bloomberg and Thomson Reuters add or expand social media data monitoring and analytics tools to their professional workstations. These notable events inspired more hedge funds, prop traders and analytics providers to explore new approaches to help capture alpha from social data and the thousands of micro events that go unnoticed every day.

Academics often discuss an "S-Curve" of adoption and investment in new innovations. This "S-Curve" follows a long period of limited adoption, then a period of rapid adoption, concluding with a slowing of adoption as the market becomes mature and the incremental benefits diminish. The chart below compares the finance industry's use of social media data analytics to that of brand analytics.

SOCIAL MEDIA DATA ANALYTICS ADOPTION BY INDUSTRY (THE "S-CURVE")



Looking at the evolution of social media analytics in the brand space provides interesting context on what is next for social data analysis in finance. Brand analysis has been highly dependent on a "base layer" - the development of high quality content to build on. Even the best analytics can't make up for lack of, incomplete or skewed data. Looking at the timeline above we can see three distinct phases:

- Phase 1: Platform Creation/User Adoption
- Phase 2: Analytics Development
- Phase 3: Integration with Existing Workflows

Social media content applicable to finance has evolved in a similar pattern but on a delayed cycle. We see the same three phases emerging in finance but 3-plus years later than in brand based analytics. We're just reaching the exciting inflection point, early in Phase 2, where costs and barriers to use have decreased, allowing new entrants the ability to gain an information advantage. Now that the tools used to search and source meaningful social data have matured, even early startups and small hedge funds are able to start exploring the patterns and insights in the data.

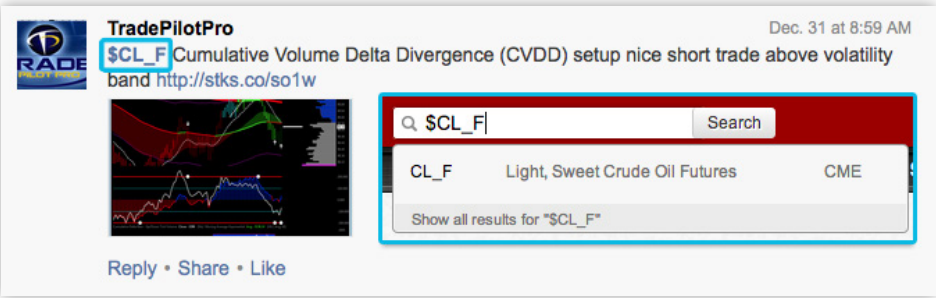
GROWTH OF FINANCIAL DISCUSSIONS

Consumer and brand discussions were the base layer that enabled marketers and brand managers to create valuable social media analytics. Much in the same way that the growth of discussion around capital markets, equities, macroeconomic indicators, FX and breaking news serve as the base layer for valuable social analytics for financial services. In financial use cases, platform creation, user adoption and relevant discussion have evolved significantly over the last few years in terms of structure, depth and breadth.

A key enabler of financial discussions on social media has been the adoption of the "Cashtag". Cashtagging is the convention of adding a "$TICKER(s)" tag to content to associate the discussion with tradable equities.

This convention was started by StockTwits in 2008 and officially adopted by Twitter in July 2012. Cashtagged discussions have grown massively over the past three years. In comparable periods from 2011 to 2014 Cashtagged conversations on Twitter around Russell 1000 securities increased more than 550% reaching several million messages per quarter.

StockTwits has evolved as the leading finance-specific social platform and has been an innovator in introducing new conventions and expanding financial discussion on other platforms including Twitter. Use of the Cashtag convention has expanded beyond equities to FX, futures, commodities and alternative investments, such as BitCoin, being tagged similarly.



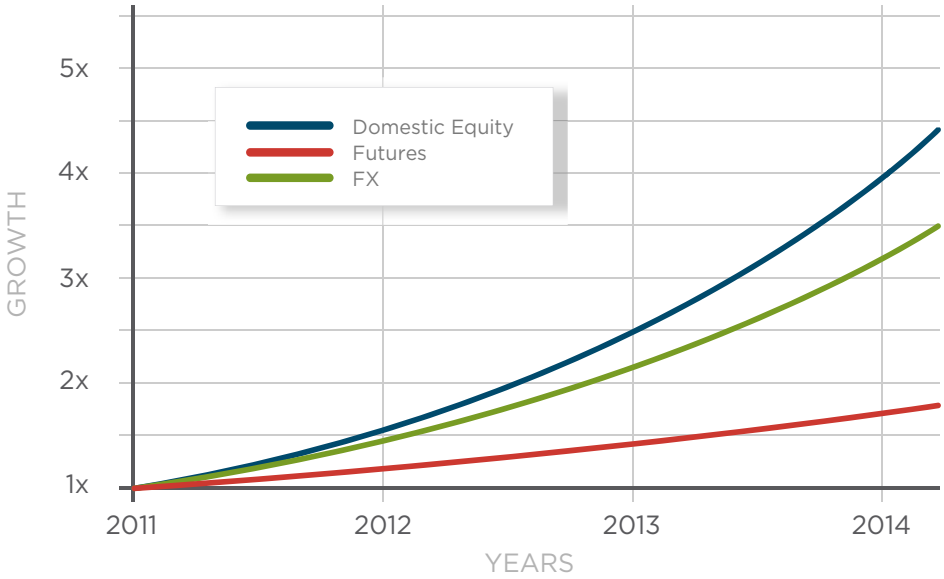
A key enabler of financial discussions on social media has been the adoption of the "Cashtag". Cashtagging is the convention of adding a "$TICKER(s)" tag to content to associate the discussion with tradable equities.

In comparable periods from 2011 to 2014 cashtagged conversations on Twitter around Russell 1000 securities increased more than 550% reaching several million messages per quarter.

Bitcoin also provides a useful example of where Cashtags sometimes only scratch the surface of conversations. Last month on Twitter there were only 1,000 messages tagged with $BCOIN but nearly 1.6 million messages containing the word Bitcoin[3]. With the trading of Bitcoin spreading to new exchanges in 2014, there is a huge opportunity for alpha for the firms that can best understand the volatility in Bitcoin markets.

Currency pair discussion is another interesting area of activity. As of 2013 there were already around 500,000 messages per month with at least one of the top 18 most actively traded currency pairs mentioned[4].

This chart shows normalized growth of Twitter discussions the different asset classes discussed. All three asset classes show increasingly exponential growth with equity discussion growing the most rapidly and FX discussion just beginning an exponential climb.



In addition to the short-form, micro-blogging content of Twitter and StockTwits, other structured social data has emerged. Estimize has been successful at creating a social platform around crowd-sourced and buy-side earnings estimates. More than 3500 analysts contribute to a transparent platform that breaks the traditional sell-side research model. A whitepaper published by Estimize contends that their consensus has proven more accurate than comparable sell side data sets more than 69% of the time[5]. A number of Estimize users also provide their StockTwits and Twitter handles along with their estimates allowing people to follow both their investment

Bitcoin provides a useful example of where Cashtags sometimes only scratch the surface of conversations. Last month on Twitter there were only 1,000 messages tagged with $BCOIN but nearly 1.6 million messages containing the word Bitcoin.

thoughts and numerical projections in tandem. SumZero has created a spiritually similar platform for research and trade idea sharing, but focused on long form research reports instead of quantitative data.

Furthermore, long form blog data and Foursquare "check-in" data is now being adopted, monitored and analyzed in the brand analytics space. It won't be long before financial market participants take note and start using these sources.

FINANCE-SPECIFIC TOOLS EMERGE

As financial social media sites and conversation has grown (the bottom tail of our S-Curve) analysts, traders, technologists and academics have begun to make sense of the data.

In our first whitepaper "Social Media in Markets: The New Frontier" we outlined three primary initial use cases:

1. Equity Sentiment Analysis
2. Breaking News Discovery
3. Macroeconomic Trend Analysis

An ecosystem of companies building analytics on top of social data for finance has begun to emerge for these use cases. The first class of companies we have seen in the marketplace focus on social media "monitoring" for finance. The companies and platforms embrace social media as news and have built displays to show filtered and possibly enhanced social media content to keep users informed. Companies include Eagle Alpha, Hedge Chatter, Market Prophit and Finmaven. Additionally, Bloomberg and Thomson Reuters now have filtered data from Twitter and StockTwits on their professional platforms. Non-financially focused social monitoring firms such as Brandwatch have also illustrated that investors and not just brands can use their platforms.

A second class of companies focuses on social media analytics for finance. These companies apply advanced analytics to create scores, signals and other derived data from Twitter or other social media. These companies include Social Market Analytics, Contix, Eagle Alpha, Market Prophit, Infinigon, TheySay, Knowsis, Dataminr, PsychSignal and mBlast. For example, Social Market Analytics takes a stream of equity related Tweets, filters out noise

A second class of companies focuses on social media analytics for finance. These companies apply advanced analytics to create scores, signals and other derived data from Twitter or other social media.

and spam and creates intraday sentiment scores. They've conducted research showing 55% return (1.92 sharpe ratio) over a two year test period[6]. Eagle Alpha has spent several years honing both machine-learning and human-screening processes to identify financial topic-based influencers to find and give perspective to key financial news breaking and being discussed on Twitter.

Use cases are now increasingly moving beyond equity markets. Credit, FX and commodity analytics platforms are coming to market. In the energy markets, IHS has begun to integrate social data and calculate a sentiment index. Further, new use cases have started to develop and are being explored both commercially and academically around risk management. Global consultancies McKinsey[7] and PwC[8] have both recommended the use of social media and textual analysis as a part of counterparty risk assessment. As this use case develops further, we expect to see new firms emerge to provide services in this specific area.

UNLOCKING VALUE IN SOCIAL DATA

For hedge funds and analytics providers alike, delivering meaningful value and turning broad streams of social data into actionable intelligence is a non-trivial task. Once consuming social data, the start to finish process of analytics has 5 general phases:

1. Sourcing Social Data
2. Filtering Noise
3. Specific Filtering
4. Sorting, Scoring and Aggregation
5. Link to Existing Data

Once consuming social data, the start to finish process of analytics has 5 general phases:

1. *Sourcing Social Data*
2. *Filtering Noise*
3. *Specific Filtering*
4. *Sorting, Scoring and Aggregation*
5. *Link to Existing Data*

1. SOURCING SOCIAL DATA

The first step is to collect the social data you need. The decision of where to get social data from is not necessarily an easy one. The reality is that each business has unique social data needs, yet there is no blueprint for how to determine your needs. Some considerations in finding the right data provider include:

- Can they provide all the relevant data you need?
- Can they provide the level of reliability you need?

- Can they provide historical data?
- Can they provide rapid-response exploratory search tools to help define your test data set?

2. FILTERING NOISE:

Filtering noise is truly a grey area. Over-filtering can remove valuable data for certain types of analysis while under-filtering can create biased noise, which can skew analytic results. Common tools for generic filtering include:

- **Data-Provider Filtering** – Sources such as Twitter may provide additional metadata about users which allow for base-level filtering. Twitter provides a layer of spam filtering even in "firehose" feeds. Data providers such as Gnip also allow for customers to create bespoke filters on top of the firehose.
- **3rd-Party User Scoring** - Third-party metrics such as Klout have proven useful in ranking users in terms of influence and importance. Additionally, Market Prophit and mBlast have created scores specific to the finance industry.
- **Repetition Filtering** – Text Similarity Metrics such as the Levenshtein distance or semantic similarity can be used to filter out Tweets that contain unoriginal content, but are not Retweets or reblogs.

Once analytics have been created, real value is derived by linking social data back to existing processes.

3. TOPIC-SPECIFIC FILTERING

Topic-specific filtering is where one starts to filter down activities or users specific to the end analysis. For example, in equity sentiment analysis topic-specific filtering would include filters limiting content only to activities that are reliably relatable to equity identifiers and contain or link to score-able content.

4. SORTING, SCORING AND AGGREGATION

After a cleaned and filtered data set is created, analytics can be built on top. The range of analytic lenses that can be put on top of social data is constantly growing, but they can generally be grouped into several broad categories.

- **User-Centric Analysis** – Identification of key users for either amplification analysis or influencer analysis.

- **Topic-Specific Scoring** – Sentiment, volume or other scoring that quantifies discussion around a focused topic, e.g. a company, brand, equity ticker, or FX pair.
- **Trend Tracking** – Counting or aggregation of counts within a topical framework that can identify trending topics or news.

5. LINK TO EXISTING DATA

Lastly, once analytics have been created, real value is derived by linking back to existing processes. In financial use cases, the most obvious examples are linking equity Cashtags back to ticker-driven data or linking sector or macroeconomic analysis back through benchmark indicators or macroeconomic data identifiers. Once a link is made to an existing identifier, and thus to price data and all existing datasets, the true value of social data is unlocked.

ACADEMICS WORK TO REFINE ANALYTIC TECHNIQUES

Initial work in academia focused on proving correlations between social sentiment and stock prices and social volume and trading volume. This work served as a base validation for value in tracking sentiment and volume movements. More recent studies have shifted to focus on optimization and application of social data in capital markets.

Recent studies explored various enhancements to the initial approach taken by Johan Bollen[9]. They looked at social media analytics modifying the analysis by examining different ways to process the data such as applying network graphs, combined with non-social news sources, quantifying user influence and applying topical groupings.

An interesting study from University of Illinois– Chicago and Hong Kong University[10] explored the application of a topic-based approach to predicting markets using social media sentiment. Machine learning techniques were applied to derive topic patterns from social data rather than specifying ticker-based linkage. After deriving topics, the study showed how sentiment on these topics could predict market movements. The results showed improved accuracy by employing a topic-based approach with the potential to uncover valuable links among securities.

Initial work in academia focused on proving correlations between social sentiment and stock prices and social volume and trading volume. More recent studies have shifted to focus on optimization and application of social data in capital markets.

In another advanced study, Ruiz Et al. (2012)[11] explored how layering social graphs on top of financial conversations could improve correlations. In the paper, they explored how factoring in interactions between users and identifying user clusters could help improve stock prediction based on the text of Tweets.

Finally, a third paper by Zhang et al (2012)[12] expanded analysis beyond equity and looked at Twitter buzz around commodities like gold and oil as well as general topics such as "economy". They explored a sentiment based on indexes of "hope", "fear" and "worry" and concluded that there is strong correlation between the sentiment and next-day price changes. A recent announcement by Twitter and Gnip to issue data grants to the academic community will surely speed up advances in academic research of social data.

WHAT'S NEXT?

Where are we now? When looking at the social data analysis within the financial sector relative to brand analytics - the financial industry is about three years behind. However, we are at the beginning of an inflection point and acceleration of the S-curve. In the past 12 months, we've seen both an escalation in the number of new firms embracing and innovating, as well as early adoption by some of the larger and somewhat more risk-averse players in the industry.

As financial discussion has become mainstream on social media, new tools have come to market. Hedge funds and HFT's were the leaders in adoption of social data for research and trading use cases, and we see this trend continuing in 2014. But now the broader financial community also has an eye on the pulse of discussion on social media. As the audience demand is broadening, incentives are emerging for new content to find its way to social media platforms. Two interesting areas that are starting to develop are investor relations and sell-side research. As new content sets start being posted to and discussed on social media, even more financial applications become feasible.

Challenges remain, though. Standards need to be developed for international growth and methods honed for mining non–Cashtag data. During 2014 there should be a cyclical acceleration as we move up the S-Curve with audience growth driving the creation of additional data (social conversations), and additional data driving the creation of additional analytics. As we've seen in brand analytics, the expansion of content has created opportunities for still new types of monitoring and analytics.

As new data and tools emerge to monitor global thought and discussion, financial professionals need to evaluate which ones fit with their investment theses. Those who are able to successfully implement will stand to reap advantage, information and returns.

ABOUT THE AUTHOR:

DAVID GREENFIELD

David is Director Financial Services at Gnip. He is responsible for building the finance business and product suite. Previously at FactSet for 10 years, David is a regular speaker on social data application in finance and markets. He is currently enrolled in the Data Science program at Columbia University.

ABOUT GNIP

Gnip is the world's largest and most trusted provider of social data. We deliver more than 100 billion social media activities per month. Our clients include business intelligence platforms, social media monitoring firms, hedge funds, prop traders and financial technology platforms.

REFERENCES

[1] Adobe- Which Social Networks matter in 2014
http://blogs.adobe.com/digitaleurope/2014/01/03/social-networks-care-2014/

[2] Carl Icahn's $12 billion apple Tweet
http://blogs.marketwatch.com/thetell/2013/08/13/carl-icahns-12-billion-apple-tweet/

[3]Gnip

[4]Top Currency Pairs
http://www.investopedia.com/ask/answers/06/maincurrencypairs.asp

[5]Crowdsourcing Forecasts: Competition for Sell-Side Analysts?
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2333671

[6]SMA Proof
https://www.socialmarketanalytics.com/proof

[7]Social-Media Analytics Offer New Insight on Risk
http://ww2.cfo.com/risk-management/2013/04/social-media-analytics-offer-new-insight-on-risk/2/

[8]Risk in review: Global risk in the transformation age
http://www.pwc.com/us/en/risk-assurance-services/publications/assets/pwc-risk-in-review-global-risk-transformation-age.pdf

[9]Johan Bollen et al Twitter mood predicts the stock Market
http://arxiv.org/pdf/1010.3003&

[10]Exploiting Topic based Twitter Sentiment for Stock Prediction
http://www.cs.uic.edu/~liub/publications/ACL-2013-Jianfeng-stock-short.pdf

[11]Correlating Financial Time Series with Micro-Blogging Activity
http://www.cs.ucr.edu/~vagelis/publications/wsdm2012-microblog-financial.pdf

[12]Predicting Asset Value Through Twitter Buzz
http://www.ickn.org/documents/Collin2011_Zhang_Fuehres_Gloor.pdf