

**UM ESTUDO SOBRE O MERCADO BRASILEIRO  
DE AÇÕES A PARTIR DE DADOS DO TWITTER**



HUGO SILVA SANTOS

**UM ESTUDO SOBRE O MERCADO BRASILEIRO  
DE AÇÕES A PARTIR DE DADOS DO TWITTER**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

ORIENTADOR: ALBERTO H. F. LAENDER  
COORIENTADOR: ADRIANO C. M. PEREIRA

Belo Horizonte

Março de 2016

**Ficha catalográfica elaborada pela Biblioteca do ICEx - UFMG**

Santos, Hugo Silva.

S237e Um estudo sobre o mercado brasileiro de ações a partir de dados do twitter. / Hugo Silva Santos. – Belo Horizonte, 2016.

xxii, 62 f.: il.; 29 cm.

Dissertação (mestrado) - Universidade Federal de Minas Gerais – Departamento de Ciência da Computação.

Orientador: Alberto Henrique Frade Laender.

Coorientador: Adriano César Machado Pereira.

1. Computação - Teses. 2. Redes sociais on-line. 3. Twitter. 4. Bolsa de valores I. Orientador. II. Coorientador. III. Título.

CDU 519.6\*04 (043)




UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

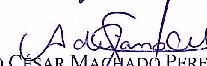
## FOLHA DE APROVAÇÃO

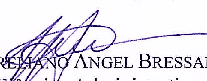
Um estudo sobre o mercado brasileiro de ações a partir de dados do twitter


**HUGO SILVA SANTOS**

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

  
PROF. ALBERTO HENRIQUE FRAIDE LAENDER - Orientador  
Departamento de Ciência da Computação - UFMG

  
PROF. ADRIANO CÉSAR MACHADO PEREIRA - COORIENTADOR  
Departamento de Ciência da Computação - UFMG

  
PROF. AURELIANO ANGEL BRESSAN  
Departamento de Ciências Administrativas - UFMG

  
PROF. FABRÍCIO BENEVENUTO DE SOUZA  
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 06 de abril de 2016.



*Dedico este trabalho a todos aqueles que, de alguma forma, colaboraram para a sua realização.*





# Agradecimentos

Gostaria de agradecer a todos que me apoiaram durante os anos do mestrado. Podem ter certeza que cada apoio, cada palavra e cada oração foi fundamental para me dar mais forças e energias para estudar durante as madrugadas e finais de semana. Como cristão, agradeço a Deus por me proporcionar disciplina que excedeu meu próprio entendimento. Agradeço à minha esposa, pela compreensão em ir dormir só, enquanto eu precisava honrar compromissos nos estudos. Agradeço aos meus pais que depositaram em mim confiança e palavras de reconhecimento. Meus sinceros agradecimentos ao Prof. Alberto Laender e ao Prof. Adriano Pereira por compartilharem comigo seus imensos conhecimentos. Meus agradecimentos aos irmãos da Igreja Batista da Fé que atuam como apoio familiar e de valores de sabedoria, tolerância e humildade. Aos colegas da Denox por, juntos, criar um ambiente descontraído e de excelência técnica e profissional que me permitiu realizar trocas de contexto entre o mestrado e o trabalho de forma agradável. Ao Instituto Nacional de Ciência e Tecnologia para a Web (InWeb) por compartilhar os dados fundamentais para o desenvolvimento deste trabalho.



*“How much better to get wisdom than gold, to get insight rather than silver!”*  
(Proverbs 16:16 - NIV)



# Resumo

Nesta dissertação, apresentamos um estudo do mercado brasileiro de ações baseado em uma ampla caracterização e análise de dados do Twitter. Em nossa análise, mostramos que os eventos e notícias sobre o mercado de ações são capazes de gerar picos de publicação pelos usuários do Twitter e que a frequência de postagem segue o início do pregão de negociação e se mantém por aproximadamente três horas após o encerramento do mercado de ações. Além disso, com base em uma pesquisa realizada com um nicho específico de usuários do Twitter, fomos capazes de estimar que 0,5% dos usuários têm algum conhecimento do mercado de ações brasileiro e são investidores, principalmente individuais, interessados na postagem e visualização de notícias, sendo que 45% deles usaram o Twitter como uma fonte para as decisões de investimento. Observamos também que o número total de ordens e o volume financeiro das ações são positivamente correlacionados para 66% das ações mencionadas no Twitter, enquanto que a oscilação diária e a máxima oscilação diária não apresentaram nenhuma correlação. Em particular, encontramos que 82% das URLs compartilhadas na rede são de notícias e esses endereços são altamente replicados de tal forma que somente 18,2% deles são únicos, além de identificarmos que seus tempos de exposição no Twitter variam entre 0,9 e 43,4 horas. Em resumo, este estudo fornece uma base de padrões comportamentais, tanto no Twitter como na bolsa de valores, para a elaboração de indicadores sociais de apoio à decisão.

**Palavras-chave:** Web, Mídias Sociais, Caracterização de Dados, Twitter, Mercado Brasileiro de Ações.



# Abstract

In this dissertation, we present a study of the Brazilian stock exchange market based on a large characterization and analysis of Twitter data. In our analysis, we show that events and news about the stock market are capable of generating peaks of postings by Twitter users and that the frequency of posts follows the starting of the exchange trading day and maintains for about three hours after the stock market closing hour. Moreover, based on a survey conducted with a specific niche of Twitter users, we have been able to estimate that 0.5% of those users have some knowledge of the Brazilian stock market and are mostly individual investors interested in posting and consuming news about this market, having 45% of them used Twitter as a source for investment decisions. Finally, we have observed that the total number of orders and the financial volume are positively correlated for 66% of the stocks mentioned on Twitter, whereas the oscillation and maximum oscillation dimensions present no correlation. In particular, we have found that 82% of the URLs shared in the network are news and that these links are highly replicated, being only 18.2% of them unique. In addition, we have measured their exposure time on Twitter and shown that it varies from 0.9 to 43.4 minutes. In general, this study provides a basis of behavioral patterns related to both Twitter and the stock exchange, to develop social indicators to support decision making.

**Keywords:** Web, Twitter, Social Media, Data Characterization, Brazilian Stock Market.





# Lista de Figuras

1.1	Padrão de <i>candlesticks</i> . . . . .	2
2.1	Metodologia utilizada . . . . .	12
2.2	Horários de negociação para cada tipo de mercado. . . . .	16
3.1	Distribuição das menções aos ativos do IBOVESPA . . . . .	22
3.2	Postagens do robô convidando os usuários a participar da pesquisa . . . . .	23
3.3	Perfis encontrados a partir dos resultados da pesquisa . . . . .	25
4.1	Número de <i>tweets</i> por dia ao longo dos 13 meses . . . . .	28
4.2	Postagens sumarizadas por hora . . . . .	29
4.3	Distribuição da participação dos usuários . . . . .	30
4.4	Distribuição da participação dos usuários em relação ao volume total de <i>tweets</i> por mês . . . . .	31
5.1	Coefficiente de Spearman calculado para cada ação. . . . .	34
6.1	Distribuição das URLs em função do seu número total de ocorrências na nossa coleção. . . . .	41
6.2	Restrição de acesso presente no <i>site</i> bastter.com. . . . .	42
6.3	Distribuição do tempo de exposição das notícias. . . . .	44
6.4	Notícia da empresa Petrobras publicada no <i>site</i> ADVFN em 16/07/2013. . . . .	45
6.5	Intervalo de confiança para o tempo de exposição dos <i>tweets</i> de notícia. . . . .	45



## Lista de Tabelas

2.1	Estatísticas gerais das nossas coleções de dados . . . . .	13
2.2	Questionário usado na pesquisa com os usuários do Twitter . . . . .	15
2.3	Escala de Cohen. . . . .	16
2.4	Escala positiva-negativa baseada na escala de Cohen. . . . .	17
3.1	Análise estática dos <i>tweets</i> . . . . .	21
3.2	Resultado da pesquisa com os usuários do Twitter . . . . .	24
4.1	Primeiros e últimos sete dias da coleção e seus respectivos totais de menções postadas . . . . .	28
5.1	Percentual de correlação para cada dimensão considerando o conjunto de todos os <i>tweets</i> que mencionaram as ações . . . . .	35
5.2	Percentual de correlação para cada dimensão considerando o conjunto de todos os <i>tweets</i> que mencionaram as 31 ações. . . . .	35
5.3	Conjunto das 22 ações que são positivamente correlacionadas com o volume financeiro e o número total de ordens e as cinco que não apresentaram correlação. . . . .	36
6.1	Catálogo de URLs Completas . . . . .	40
6.2	Amostra com as primeiras e as últimas quatro URLs da ordenação de repetições. (Coleção completa no Apêndice B) . . . . .	41
6.3	Classificação das 130 URLs selecionadas . . . . .	42
6.4	Amostra a identificação das fontes das primeiras e as últimas quatro URLs da ordenação de repetições . . . . .	43
6.5	Mapeamento de fontes com comportamento em que o próprio autor postava a mesma URL repetidas vezes na rede. . . . .	44
A.1	Códigos que referenciam ativos do IBOVESPA ou seu próprio índice. . . . .	55

B.1	130 URLs de maior ocorrência e suas classes . . . . .	59
-----	---	----

# Sumário

<b>Agradecimentos</b>	<b>ix</b>
<b>Resumo</b>	<b>xiii</b>
<b>Abstract</b>	<b>xv</b>
<b>Lista de Figuras</b>	<b>xvii</b>
<b>Lista de Tabelas</b>	<b>xix</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	3
1.2 Objetivos . . . . .	4
1.3 Contribuições . . . . .	4
1.4 Trabalhos Relacionados . . . . .	5
1.5 Organização . . . . .	8
<b>2 Metodologia</b>	<b>11</b>
2.1 Visão Geral da Metodologia . . . . .	11
2.1.1 Coleta dos Dados e Descrição das Coleções . . . . .	12
2.1.2 Análise Estática . . . . .	14
2.1.3 Análise Temporal . . . . .	14
2.1.4 Análise de Correlação . . . . .	14
2.1.5 Análise de Conteúdo . . . . .	18
2.2 Questões de Pesquisa . . . . .	19
<b>3 Análise Estática</b>	<b>21</b>
3.1 Mapeamento das Postagens . . . . .	21
3.2 Pesquisa com os Usuários . . . . .	23
3.3 Caracterização dos Usuários . . . . .	24

3.4	Conclusões . . . . .	25
<b>4</b>	<b>Análise Temporal</b>	<b>27</b>
4.1	Padrões Temporais das Postagens . . . . .	27
4.2	Participação dos Usuários . . . . .	30
4.3	Conclusões . . . . .	31
<b>5</b>	<b>Análise de Correlação</b>	<b>33</b>
5.1	Correlação entre as Quatro Dimensões e as Postagens . . . . .	33
5.2	Ações Correlacionadas . . . . .	36
5.3	Conclusões . . . . .	37
<b>6</b>	<b>Análise de Conteúdo</b>	<b>39</b>
6.1	Classificação e Ordenação de Conteúdo . . . . .	39
6.2	Análise das Fontes . . . . .	43
6.3	Tempo de Exposição das Notícias . . . . .	44
6.4	Conclusões . . . . .	46
<b>7</b>	<b>Conclusões Finais e Trabalhos Futuros</b>	<b>47</b>
7.1	Conclusões Finais . . . . .	47
7.2	Trabalhos Futuros . . . . .	49
	<b>Referências Bibliográficas</b>	<b>51</b>
	<b>Apêndice A IBOVESPA e Ações Consideradas</b>	<b>55</b>
	<b>Apêndice B Lista das 130 URLs de Maior Ocorrência</b>	<b>59</b>

# Capítulo 1

## Introdução

Compreender o comportamento de ações em bolsa de valores é ainda hoje um desafio e tema de pesquisa em inúmeras áreas de conhecimento. Muitos estudos com o objetivo de mapear o comportamento de ativos já foram elaborados, mas com o surgimento de novas tecnologias computacionais e, sobretudo, das redes sociais *online*, como o Twitter<sup>1</sup> e o Facebook<sup>2</sup>, novas linhas de pesquisa se tornaram viáveis [Sprenger et al., 2014] tendo em vista a quantidade de dados disponíveis na Web que, de certa forma, refletem comportamentos e opiniões dos seus usuários [Maia et al., 2008].

O mercado de ações sempre atraiu muita atenção da academia. Algumas pesquisas tentam encontrar padrões de comportamento dos ativos de forma que seja possível caracterizar tendências, padrões preditivos ou aleatórios. Outros estudos mapeiam as oscilações dos preços como sendo ocasionadas pelo surgimento de novas informações [Fama, 1965], ou seja, notícias, ao invés de preços passados e presentes.

O Brasil, por sua vez, tem na Bolsa de Valores de São Paulo (BM&FBOVESPA ou simplesmente BOVESPA) sua principal bolsa de valores para negociação de ações. Fundada em 1890, a BOVESPA está entre as 20 maiores bolsas de valores do mundo, sendo a maior da América do Sul<sup>3</sup>. É também a mais importante instituição brasileira na intermediação de operações do mercado de ações e a única que opera com valores mobiliários, mercadorias e futuros de câmbio no Brasil. A BOVESPA atua como um direcionador para os mercados de capitais brasileiros. Existem cerca de 381 empresas listadas na BOVESPA e seu indicador de referência é o índice BOVESPA<sup>4</sup>

---

<sup>1</sup><http://twitter.com>

<sup>2</sup><http://facebook.com>

<sup>3</sup><http://www.insidermonkey.com/blog/trading-places-the-20-largest-stock-exchanges-in-the-world-335310/>

<sup>4</sup>Os ativos do IBOVESPA são responsáveis por 80% do movimento financeiro da negociação de ações da BOVESPA

(IBOVESPA).

As redes sociais *online*, ao se tornarem extremamente populares, também se transformaram em ambientes virtuais de disseminação em larga escala de novas informações. De acordo com um relatório da Nielsen<sup>5</sup> as mídias sociais superaram os *e-mails* como atividade *online* mais popular, sendo que dois terços da população global *online* participa de redes sociais e *blogs*.

O Twitter, rede social *online* e serviço de *microblogging*, emergiu como um meio de propagação de acontecimentos recentes [Kwak et al., 2010]. Nesse serviço é possível que usuários sigam uns aos outros sem, necessariamente, haver reciprocidade como ocorre em outras redes como *Facebook* e *LinkedIn*. Além disso, o limite de postagem de 140 caracteres é um direcionador para os usuários realizarem postagens de forma rápida e sucinta. Por exemplo, em 24 de Junho de 2014, uma informação publicada com exclusividade no Twitter demonstrou a capacidade do serviço de influenciar as negociações na BOVESPA. Neste caso, um único *tweet* postado pelo governador do Estado do Paraná reverteu a tendência de alta das ações da COPEL, a companhia de energia do referido estado<sup>6</sup>.

Nesta dissertação, apresentamos um estudo sobre o mercado brasileiro de ações a partir da caracterização e análise em larga escala de quase três milhões de *tweets* e mais de 600 mil padrões *candlesticks* pertencentes a 73 ações oficiais da BOVESPA (Apêndice A). Esses padrões de *candlestick* descrevem as oscilações de todas as ações que compõem o IBOVESPA a cada 15 minutos (veja a Figura 1.1).

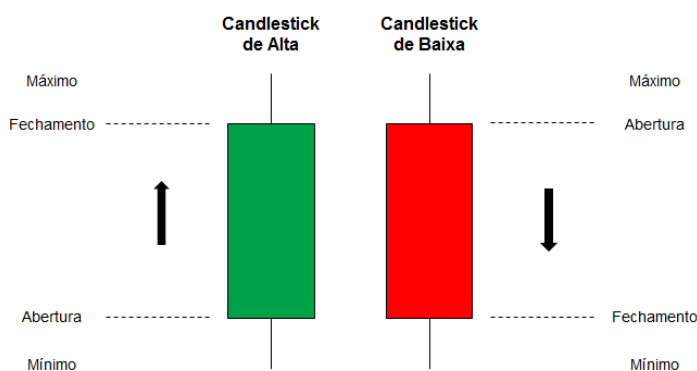


Figura 1.1: Padrão de *candlesticks*

<sup>5</sup>Nielsen *Online Report. Social networks & blogs now 4th most popular online activities* - <http://goo.gl/iG5L9R>

<sup>6</sup>Infomoney: Governador do Paraná veta pelo Twitter reajuste da Copel; ações viram para queda - <http://goo.gl/T4YTEt>



## 1.1 Motivação

Apesar das inúmeras pesquisas existentes sobre o mercado de ações que utilizam, por exemplo, técnicas de mineração de dados e redes neurais [Poh et al., 1998], existem poucos estudos orientados às redes sociais *online* que, além de utilizar o Twitter como fonte de dados voltados à BOVESPA, consideram a língua portuguesa como idioma base para extração e análise textual dos dados.

O estudo e a caracterização do comportamento de usuários brasileiros de redes sociais [Benevenuto et al., 2009] que compartilham informações no Twitter a respeito da BOVESPA permitem a identificação individual de investidores potenciais, profissionais ou influenciadores, que buscam na rede por informação relevante para o acompanhamento dos ativos ou mesmo para tomada de decisão.

Como mencionamos, pretende-se nesta dissertação compreender através de um estudo de caracterização quais são as informações postadas diariamente no Twitter a respeito do mercado de ações brasileiro, se os usuários envolvidos possuem algum padrão de comportamento e, sobretudo, se há alguma correlação dessas informações com o comportamento dos ativos na BOVESPA.

As motivações para este trabalho podem ser encontradas no crescente uso da Web como canal para investimento em ações através do aumento da tendência de utilização de *homebrokers* para tomada de decisão sobre investimentos, até a busca de informações relevantes que ofereçam vantagem competitiva para os investidores. Assim, as principais motivações para o presente trabalho são:

1. A tendência do crescimento do volume e da velocidade com que as informações são compartilhadas na Web, sobretudo, em redes sociais, que podem afetar as decisões dos investidores e, conseqüentemente, o mercado financeiro;
2. O Twitter tem se mostrado uma boa fonte para monitoramento de eventos em tempo real;
3. A necessidade de maiores estudos que considerem a BOVESPA e as análises em larga escala de *tweets* no idioma Português;
4. A compreensão de padrões de comportamento, tanto no Twitter como na bolsa de valores, que possam gerar uma base para a elaboração de indicadores sociais de apoio à decisão.

## 1.2 Objetivos

Tendo em vista a carência de trabalhos relacionados ao contexto brasileiro, esta dissertação tem como principal objetivo contribuir para um melhor entendimento de como se relacionam as informações publicadas pelos usuários do Twitter, bem como do comportamento das ações da BOVESPA.

Para isso foi feita uma caracterização detalhada de como ocorrem as diferentes postagens pelos usuários do Twitter a respeito do mercado de ações brasileiro. As análises consideram os aspectos das postagens dos *tweets*, assim como variações dos atributos das ações na BOVESPA. Primeiramente, os dados foram explorados de maneira agregada, levando em conta os atributos estáticos (número total de *tweets*, percentual de *tweets* que mencionam ações em seu texto, quantidade de *tweets* que contêm URLs), além da identificação do perfil dos usuários que postam essas mensagens. Além disso, considerou-se o aspecto temporal, ou seja, como ocorrem as variações das postagens ao longo do tempo e se essas variações constituem padrões comportamentais. Outro aspecto importante da dissertação é a compreensão se correlacionam as oscilações das ações na bolsa de valores e o seu respectivo número de menções nos *posts* no Twitter. Por fim, o último aspecto é a análise do conteúdo publicado a partir da descompactação das URLs e do estudo da natureza dos *links* publicados nas postagens do Twitter e de suas fontes.

Considerando todos esses aspectos, o principal objetivo desta dissertação é apresentar uma amplo estudo do mercado brasileiro de ações a partir da caracterização e análise em larga escala dos *tweets* que mencionam explicitamente as ações e dos padrões *candlesticks* oficiais para as 73 ações do índice IBOVESPA. A partir disso, compreender padrões de comportamento e perfil dos usuários do Twitter, quais tipos de conteúdo eles postam e estão interessados, além de estudar como o padrão de postagens se relaciona com as variações dos ativos no BOVESPA.

## 1.3 Contribuições

Os resultados desta dissertação permitem entender melhor o perfil dos usuários do Twitter que postam informações sobre o mercado de ações brasileiro. Além disso, as questões de pesquisa propostas tornam possível estruturar um paralelo entre as informações relevantes do Twitter e o comportamento dos ativos do IBOVESPA de forma a compreender como esses dois universos se relacionam e, de forma mais específica, através de quais tipos de informação e fontes. Assim, as principais contribuições desta dissertação são:

- Uma *análise estática* aprofundada com base em dados do Twitter, complementadas por uma pesquisa realizada em um nicho específico de seus usuários. Os estudos recentes na literatura caracterizam a informação presente nos *tweets* e executam algum tipo de análise de sentimento, mas não avaliam se os usuários que realizam as postagens são de fato investidores. Neste estudo, entrevistamos mais de 3.300 usuários do Twitter a fim de conhecer os seus interesses específicos, e descobrimos que 45% deles usam a rede social como uma fonte de informação para tomar decisões de investimento.
- Uma *análise temporal* detalhada que ilustra os padrões que emergem dos dados e dos comportamentos dos usuários à medida que analisamos a variação das suas postagens no Twitter em função do tempo.
- Um estudo de *correlação* transversal que traça um paralelo entre as menções dos ativos nas postagens no Twitter e como as frequências dessas menções se relacionam com as variações dos ativos do IBOVESPA em função do seu volume financeiro, número total de ordens, oscilação diária e oscilação diária máxima .
- Uma *análise do conteúdo* das URLs postadas nos *tweets*, incluindo a identificação de suas fontes e o tempo de exposição dessas URLs na rede social. Nessa análise, por exemplo, classificamos os tipos de postagem como análise financeira, opinião e notícia, com o objetivo de compreender qual tipo de informação é mais compartilhado na rede e, conseqüentemente, se possui algum potencial para influenciar o comportamento dos ativos na bolsa de valores.

Parte dos resultados descritos nesta dissertação foram apresentados na *16th International Conference on Electronic Commerce and Web Technologies (EC-Web 2015)*, realizada durante a *26th DEXA Conferences and Workshops* [Santos et al., 2015].

## 1.4 Trabalhos Relacionados

O estudo apresentado nesta dissertação refere-se a um amplo espectro de pesquisa envolvendo redes sociais *online* [Maia et al., 2008; Mislove et al., 2007], mercados de ações [Zhang & Skiena, 2010; Ruiz et al., 2012; Gilbert & Karahalios, 2010] e, mais especificamente, o Twitter [Kwak et al., 2010; Cha et al., 2012; Sprenger et al., 2014]. Este estudo tem como principal diferencial ser o primeiro a mostrar quais ativos do IBOVESPA pode ter o seu volume financeiro e número total de ordens monitorados através do Twitter.

Nos últimos anos, o surgimento das redes sociais proporcionou uma grande disponibilidade de dados na Web e, com isso, vários estudos surgiram para compreender comportamentos, identificar perfis e mapear redes de influência, além de categorizar as informações postadas. Por exemplo, Maia et al. [2008], a partir de dados extraídos do Youtube<sup>7</sup>, elaboraram uma metodologia baseada em classes de comportamento para caracterizar e identificar usuários em redes sociais. Mislove et al. [2007], por sua vez, desenvolveram um dos primeiros estudos em larga escala sobre a estrutura das redes sociais a partir de uma coleção de mais de 11 milhões de usuários e 328 milhões de *links* extraídos do Flickr<sup>8</sup>, YouTube, LiveJournal<sup>9</sup> e Orkut<sup>10</sup>. Como resultado desse estudo, eles identificaram que as redes sociais tendem a possuir um núcleo densamente ligado de nodos de alto grau e que esse núcleo liga pequenos grupos de *clusters* formados por nodos de baixo grau que, por sua vez, estão em zonas periféricas das redes.

No contexto dos trabalhos que relacionam o mercado de ações e a análise de informações em mídias, Zhang & Skiena [2010] utilizaram técnicas de processamento de linguagem natural em larga escala sobre textos provenientes de *blogs* e notícias para estudar como notícias relacionadas às empresas antecipam ou refletem os volumes de negociação de ações dessas mesmas instituições e seus respectivos retornos financeiros. Com base nos resultados, eles desenvolveram uma estratégia de negociação baseada em análise de sentimentos. A principal relação desse trabalho com esta dissertação está no fato de utilizar, entre outros, dados de 2009 do Twitter e, a partir deles, mostrar que as postagens em *blogs* e a análise de sentimento de notícias representam um método aplicável ao contexto do mercado financeiro.

Particularmente, o Twitter tem sido uma importante fonte de informação em tempo real para vários estudos que objetivam analisar padrões de comportamento e tendências do mercado. Sharma et al. [2012] desenvolveram um serviço para inferir e caracterizar atributos do Twitter de usuários usando dados coletados a partir de suas listas de interesse (um usuário pode criar uma lista de outros usuários que compartilham os mesmos interesses). Cheong & Lee [2009], por outro lado, abordaram o Twitter como uma plataforma com potencial de fornecer inteligência coletiva através da obtenção de opiniões e fatos para tomada de decisão. Eles desenvolveram uma abordagem para analisar padrões de tendências baseada na inteligência coletiva caracterizada pela base de mensagens e características demográficas dos seus usuários.

Vale ressaltar que o Twitter também tem sido usado como ferramenta de moni-

---

<sup>7</sup><http://www.youtube.com>

<sup>8</sup><http://www.flickr.com/>

<sup>9</sup><http://www.livejournal.com>

<sup>10</sup><http://www.orkut.com>

toramento de eventos correlacionados com o mundo real. Como Gomide et al. [2011] analisaram como a epidemia de Dengue é refletida no Twitter e de que forma essas informações podem ser usadas pela fiscalização de agentes públicos de combate à doença. Nesse trabalho eles verificaram a alta correlação entre o número de casos relatados pelas estatísticas oficiais e o número de *tweets* postados durante o mesmo período de tempo. Earle et al. [2012], por sua vez, apresentaram um procedimento de detecção de terremoto baseado em dados do Twitter. Eles se fundamentaram no fato das pessoas postarem *tweets* segundos após perceberem o tremor.

Bollen et al. [2011] realizaram um dos trabalhos científicos de base computacional de maior repercussão no mercado financeiro norte-americano. Em particular, eles pesquisaram se as medições de estados de humor coletivos derivados da análise em larga escala de dados do Twitter estariam correlacionadas com as variações do valor do *índice Dow Jones (Dow Jones Industrial Average - DJIA)* ao longo do tempo. Nesse estudo, foi utilizada uma coleção de *tweets* postados entre 28 de fevereiro e 19 dezembro de 2008 (9,8 milhões de *tweets* postados por aproximadamente 2,7 milhões usuários). O conteúdo dos textos dos *tweets* foi então submetido à avaliação de duas ferramentas de análise de sentimento: a *OpinionFinder*, que é capaz de classificar o humor presente em um *tweet* como uma expressão de opinião positiva ou negativa, e o *Google-Profile of Mood States (GPOMS)*, que mensura o humor em termos de seis dimensões: Calma, Vigilância, Confiança, Vitalidade, Gentileza e Felicidade. Como resultado eles encontraram uma precisão de 87,6% na previsão das variações de fechamento diário do DJIA. Vale ressaltar que o estudo realizado nesta dissertação difere fundamentalmente da pesquisa realizada por eles por não buscar padrões de previsibilidade ou casualidade entre o Twitter e a BOVESPA.

Sprenger [2011] estudou o Twitter como uma plataforma para a troca de informações sobre o mercado de ações. Em seu trabalho, ele encontrou associações entre o volume de *tweets* e o volume de transações, bem como entre os sentimentos expressos pelos *tweets* e os retornos das ações, concluindo assim que os usuários que forneciam conselhos de investimentos acima da média eram mais seguidos e os seus *tweets* recebiam mais *retweets*. A partir disso, propôs o desenvolvimento do serviço de microblog TweetTrader.net<sup>11</sup>. Nesse serviço ele aplicou métodos linguísticos computacionais para analisar cerca de 250 mil *tweets* relacionados com o mercado de ações e, a partir disso, extrair o sentimento dos investidores *online* a respeito de todas as empresas de capital aberto do índice do mercado financeiro norte-americano S&P 500<sup>12</sup>.

Em particular, Nofer [2015a,b] tem investigado a capacidade preditiva das mí-

---

<sup>11</sup><http://tweettrader.net>

<sup>12</sup>O *Standard & Poor's 500* ou, somente, S&P 500

dias sociais em relação ao retorno das ações no mercado financeiro a partir de quatro artigos. Em seu estudo o autor avalia às anomalias de mercado em plataformas de mídia social e lida com a questão de saber se os mercados eletrônicos são eficientes. A anomalia de mercado<sup>13</sup> (ou ineficiência de mercado) é uma distorção de preços que ocorre tipicamente no mercado financeiro e parece contradizer a Hipótese do Mercado Eficiente [Fama, 1965]. A anomalia de mercado geralmente se refere aos fatores estruturais, tais como uma concorrência desleal, à falta de transparência do mercado, às ações de entidades reguladoras, ou mesmo, a comportamentos enviesados de agentes econômicos.

Nofer & Hinz [2014], por sua vez, avaliam o valor preditivo do conteúdo gerado pelos usuários em relação aos retornos das ações. Os dados foram coletados diretamente de aplicações de mídias sociais, que oferecem possibilidades interessantes para pesquisadores e profissionais da área de previsão de preço de ações.

Por fim, Nofer et al. [2014] investigam a influência da privacidade e da segurança sobre o comportamento do consumidor ou investidor. Privacidade e segurança foram identificados como condições prévias importantes para que os usuários da Internet se dispusessem a compartilhar conteúdo em plataformas de mídia social [Fogel & Nehmad, 2009].

Finalmente, o principal diferencial do estudo apresentado nesta dissertação é uma visão exclusiva e abrangente do mercado de ações brasileiro, que mostra quais ações da BOVESPA podem ter o comportamento monitorado através do Twitter, e uma caracterização detalhada de seus usuários que são de fato investidores, uma questão que não tem sido explorada na literatura, pois os estudos atuais têm como viés que os usuários do Twitter que postam informações sobre o mercado financeiro são, em termos gerais, investidores desse mercado.

O estudo realizado nesta dissertação envolve quatro análises, estática, temporal, correlacional, e de conteúdo, todas elas tendo como objetivo estudar os *tweets* postados que mencionam os ativos e as variações dos respectivos ativos na BOVESPA. Vale ressaltar que esta dissertação não tem como objetivo propor um modelo preditivo, mas sim apresentar uma visão exclusiva do mercado brasileiro de ações tendo como base o comportamento e informações provenientes dos usuários do Twitter no Brasil.

## 1.5 Organização

O restante desta dissertação está organizado como se segue.

---

<sup>13</sup><http://www.investorhome.com/anomaly.htm>

O Capítulo 2 descreve a metodologia de pesquisa adotada que abrange a coleta de dados e as análises estática, temporal, correlacional e de conteúdo que compõem o estudo objeto desta dissertação. Além disso, esse capítulo também apresenta as principais perguntas de pesquisa propostas.

O Capítulo 3 discute os resultados obtidos a partir da análise dos atributos estáticos da coleção. Esse capítulo também apresenta os resultados da pesquisa realizada com os usuários do Twitter. Nela buscou-se compreender se os usuários do Twitter que postam informações sobre o mercado financeiro brasileiro são, de fato, investidores.

O Capítulo 4 apresenta os resultados que descrevem os aspectos temporais da coleção, ou seja, analisa os dados considerando a evolução dos seus atributos ao longo do tempo e a dinâmica dos registros da coleção.

O Capítulo 5 apresenta as análises de correlação entre o comportamento das ações do IBOVESPA e suas respectivas menções nas postagens dos usuários no Twitter.

O Capítulo 6 apresenta uma análise do conteúdo das URLs postadas nos *tweets*, identifica os endereços mais compartilhados, classifica os *tweets* de acordo com a URL compartilhada, identifica suas fontes (usuários) e, finalmente, determina o tempo de exposição dos *tweets* que foram classificados como notícias.

Por fim, o Capítulo 7 apresenta as conclusões e as perspectivas de trabalhos futuros.





## Capítulo 2

# Metodologia

Conforme mencionamos anteriormente, ainda hoje compreender comportamentos dos ativos em bolsa de valores é um desafio. Percebemos que esse problema tem motivado inúmeras pesquisas em diferentes partes do mundo e, sobretudo, que algumas delas têm buscado nas redes sociais evidências de comportamentos coletivos que tenham relação com os ativos de bolsas de valores.

Em particular, o Twitter, que é a rede social objeto do estudo apresentado nesta dissertação, provê uma grande quantidade de dados na Web que, de certa forma, reflete opiniões e notícias recentes compartilhadas por seus usuários. Diante disso, para que os estudos sejam válidos e os padrões identificados possam ter aplicação em outros contextos, é necessário atentar para alguns aspectos dos dados utilizados:

- É importante que a coleção contenha dados do Twitter e da BOVESPA suficientes para evitar que fenômenos excepcionais ocasionados por comportamentos específicos sejam assumidos como padrões comuns.
- Os dados devem compreender um intervalo de tempo suficiente para observar padrões recorrentes que ocorram ao longo do tempo como máximas de postagens e variações ao longo dos dias da semana.
- As análises devem considerar os diferentes tipos de conteúdo das postagens, para evitar que aspectos específicos de cada tipo possam influenciar conclusões gerais.

### 2.1 Visão Geral da Metodologia

Nesta seção apresentamos a metodologia utilizada para realizar o estudo sobre o mercado de ações a partir de dados do Twitter. Para isso, A Figura 2.1 apresenta cada

uma das etapas desde a coleta dos dados a partir do Twitter e da BOVESPA até a etapa final de análise de conteúdo.

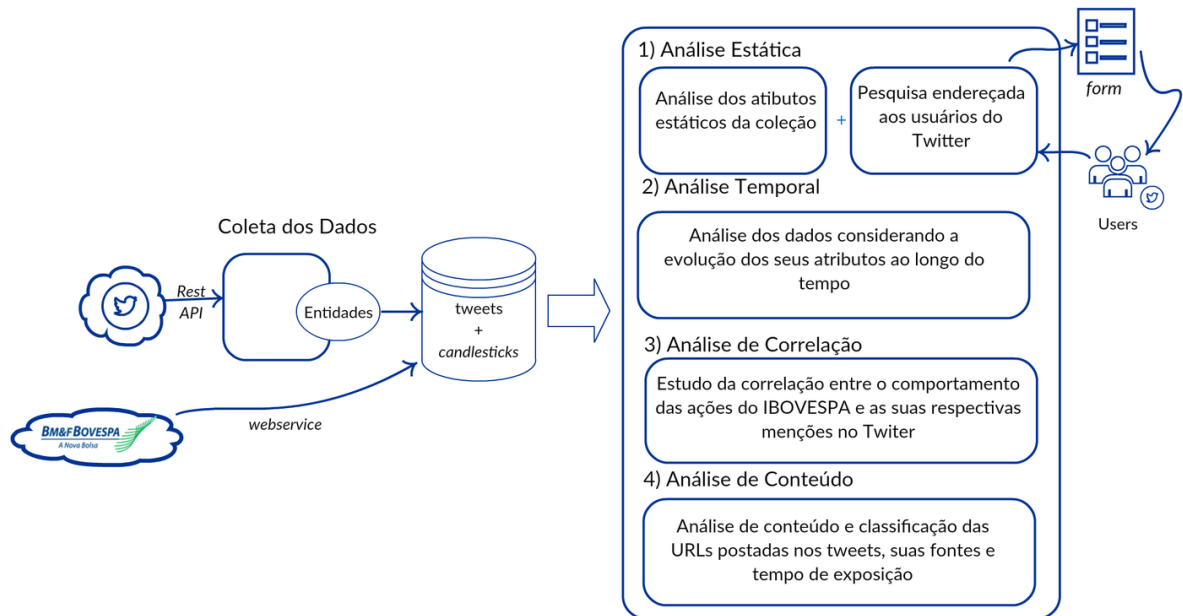


Figura 2.1: Metodologia utilizada

### 2.1.1 Coleta dos Dados e Descrição das Coleções

Para elaborar o estudo do mercado de ações a partir de dados do Twitter, coletamos *tweets* postados por usuários cujos textos continham menções às empresas do índice BOVESPA, ao segmento ou setor da economia, ou ao código dos ativos. Os dados deste estudo foram coletados em parceria com o InWeb<sup>1</sup> (Instituto Nacional de Ciência e Tecnologia para Web), que tem como objetivo desenvolver modelos, algoritmos e novas tecnologias que permitam aumentar a integração da Web com a sociedade. Para realizar nossa análise, foram coletados dados do Twitter usando a sua API REST<sup>2</sup>. Foram considerados os *tweets* publicados entre julho de 2013 e julho de 2014. Tais dados podem ser acessados atualmente através do projeto “Observatório do Investimento”<sup>3</sup>.

Além disso, antes de ser incluído na coleção de experimentação, cada *tweet* coletado teve o seu texto submetido ao reconhecimento de entidades [Parameswaran et al.,

<sup>1</sup><http://www.inweb.org.br>

<sup>2</sup><https://dev.twitter.com/docs/api>

<sup>3</sup><https://app.observatoriodoinvestimento.com/>

2010]. As entidades em questão foram nome da empresa, código da ação, setor e segmento, todas elas de acordo com o IBOVESPA. Como exemplo, os quatro *tweets* a seguir fazem parte da coleção:

1. “@clubedopairico *PETR4* não foi a R\$ 13,65 em 21/11/2008?”;
2. “*OGX* (#*OGXP3*)<sup>4</sup> abre em baixa de 29,11%, a R\$ 0,56”;
3. “Acho que quinta vou ir ver meus amigos do Itaú/Bradesco/Santander/Daycoval”;
4. “União garante crédito para capital de giro da Eletrobras <http://t.co/EcoPqkPBpc>”.

Para efeitos de nossa análise, definimos a coleção ideal para o estudo apresentado nesta dissertação como sendo formada exclusivamente pelos *tweets* que mencionam um ou mais códigos de identificação de ações. Por exemplo, na lista anterior os *tweets* 1 e 2 fazem parte da nossa coleção, enquanto os *tweets* 3 e 4 não foram considerados nos experimentos. Essa seleção se fez necessária para garantir que as postagens analisadas fossem consideradas, de fato, relacionadas ao mercado financeiro, uma vez que para determinar se os *tweets* 3 e 4 estariam relacionados seria necessário aplicar métodos específicos de análise de texto que fogem do objetivo desta dissertação.

Para análise dos dados da BOVESPA, foram utilizados dados coletados por meio de seu *webservice* oficial<sup>5</sup> por um dos nossos parceiros de pesquisa. Nessa coleta recuperamos *candlesticks* gerados a cada 15 minutos para cada ação do IBOVESPA, restringindo este conjunto de dados ao mesmo período de referência do conjunto de dados do Twitter, ou seja, de julho de 2013 a julho de 2014.

Coleção de Dados do Twitter	2.973.410 <i>tweets</i>
Primeiro <i>tweet</i> coletado	01/07/2013 00:22:03
Último <i>tweet</i> coletado	31/07/2014 23:10:24
Coleção de Dados da BOVESPA	619,125 <i>candlesticks</i>
Primeiro <i>candlestick</i> coletado	01/07/2013 10:00:00
Último <i>candlestick</i> coletado	31/07/2014 17:45:00

Tabela 2.1: Estatísticas gerais das nossas coleções de dados

A Tabela 2.1 descreve as estatísticas básicas das coleções de dados do Twitter e da BOVESPA usadas nos experimentos. Em ordem cronológica, o primeiro *tweet* foi postado em uma segunda-feira dia 01/07/2013 à 0 hora e 22 minutos, sendo que

<sup>4</sup>Durante o período de análise a empresa OGX ainda fazia parte do índice. (<http://goo.gl/vj6n>)

<sup>5</sup><http://goo.gl/1SWujW>

o último, em uma segunda-feira dia 31/07/2014 às 23 horas e 10 minutos. Foram coletados 2.973.410 *tweets* postados e 619.125 *candlesticks*.

### 2.1.2 Análise Estática

Ainda de acordo com a Figura 2.1, após a criação das coleções, seguimos para o efetivo estudo do mercado de ações. Nessa primeira etapa construímos algoritmos para extrair os atributos estáticos dos *tweets* em nossa coleção. Entendemos como atributos estáticos o texto, as URLs, as menções, *hashtags*, indicação de *retweet* e *reply*, entre outros. Em particular, nessa etapa diferenciamos o nosso trabalho por incluir, também, uma pesquisa direta envolvendo os usuários do Twitter com o objetivo de compreender o seu perfil predominante. A Tabela 2.2 apresenta o questionário utilizado na pesquisa com os usuários do Twitter presentes em nossa coleção.

### 2.1.3 Análise Temporal

A segunda etapa da construção da nossa visão é a análise temporal. Conforme mencionado anteriormente, nessa etapa estudamos a evolução dos atributos dos *tweets* ao longo do tempo. Nesse momento construímos algoritmos para iterar sobre a nossa coleção realizando a agregação dos dados por períodos de horas, semanas e meses, em busca de padrões de comportamento que se manifestam em diferentes unidades de tempo. Além disso, comparamos esses padrões com aspectos temporais da BOVESPA, como o horário do pregão de negociação do mercado à vista e os dias úteis de sua operação<sup>6</sup> (Figura 2.2).

### 2.1.4 Análise de Correlação

Nessa etapa aprofundamos a nossa análise realizando um estudo de correlação entre as postagens no Twitter e o comportamento dos ativos na BOVESPA. Para isso, usamos como ferramenta as séries temporais que descrevem os *tweets* ao longo do período de análise da nossa coleção e os *candlesticks*, também restritos ao mesmo período. Para mensurar a correlação, partimos do coeficiente de Spearman [Zar, 1972] e da escala de Cohen [Cohen et al., 2013].

---

<sup>6</sup><http://www.bmfbovespa.com.br/pt-br/regulacao/horarios-de-negociacao/acoes.aspx?idioma=pt-br>

- 1) Como você se enquadra em relação ao mercado de ações brasileiro?
- a) Eu sou investidor
  - b) Eu não sou investidor, mas pretendo investir no futuro
  - c) Eu não sou investidor e não pretendo investir no futuro
- 2) Em qual das opções a seguir você melhor se enquadra?
- a) Investidor individual
  - b) Agente Independente de Investimento
  - c) Profissional de Mídia Especializada
  - d) Analista Financeiro
  - e) Profissional de Relacionamento com Investidor
  - f) Nenhuma das anteriores
- 3) No contexto do mercado financeiro, você posta *tweets* com maior frequência sobre?
- a) Sua opinião pessoal
  - b) Análise técnica
  - c) Notícias
  - d) Outros assuntos
  - e) Não posto
- 4) De forma geral, qual o tipo de postagem de outros usuários sobre o mercado financeiro você considera mais relevante?
- a) Opiniões pessoais
  - b) Notícias
  - c) Análise financeira
  - d) Outros assuntos
  - e) Eu não uso Twitter
- 5) Os *tweets* que você posta sobre o mercado financeiro são geralmente relacionados à sua carteira de ações?
- a) Sim
  - b) Não
  - c) Eu não posto *tweets* sobre o mercado de ações
- 6) Você usa ou já usou informação proveniente do Twitter para decidir sobre algum investimento?
- a) Sim
  - b) Não

Tabela 2.2: Questionário usado na pesquisa com os usuários do Twitter

#### 2.1.4.1 Coeficiente de Spearman

O coeficiente de correlação de Spearman entre duas variáveis  $X$  e  $Y$  considera uma escala contínua de  $-1$  a  $1$ , em que  $-1$  e  $1$  são valores de limite em que as duas variáveis podem ser interpretadas como perfeitamente correlacionadas. Assim, se  $X$  ou  $Y$  assumem  $-1$  elas são negativamente correlacionadas, ou seja, caso o valor de  $X$  cresça, há uma tendência de que o valor de  $Y$  diminua e vice-versa. Por outro lado, caso  $X$  e  $Y$  assumam  $1$  como valor para o coeficiente de correlação, diz-se que elas são positivamente correlacionadas e, com isso, caso o valor de  $X$  cresça, há uma tendência de que o valor de  $Y$  também cresça.

A escala de Cohen, por sua vez, é utilizada para avaliar o coeficiente de correlação

The screenshot shows the BM&FBOVESPA website interface. At the top, there is a navigation menu with options: A BM&FBOVESPA, MERCADOS, EDUCACIONAL, SERVIÇOS, REGULAÇÃO (highlighted), PARTICIPANTES, and COMO INVESTIR. Below the navigation is a search bar and a breadcrumb trail: Home / Regulação / Horários de Negociação / Ações. The main heading is 'Horários de Negociação no Mercado de Ações (Mercado de Bolsa)'. Below this, there is a table with columns for 'Mercado2', 'Cancelamento de Ofertas', 'Pré-Abertura', 'Negociação', 'Call de Fechamento', and 'After-Market' (subdivided into 'Cancelamento de Ofertas' and 'Negociação'). Each of these columns has 'Início' and 'Fim' sub-columns. The table lists trading hours for various market segments like 'Mercado a vista', 'Fracionário', 'Mercado a termo', etc.

Mercado2	Cancelamento de Ofertas		Pré-Abertura		Negociação		Call de Fechamento		After-Market			
	Início	Fim	Início	Fim	Início	Fim	Início	Fim	Cancelamento de Ofertas		Negociação	
Mercado a vista	09:30	09:45	09:45	10:00	10:00	17:55	17:55	18:00	-	-	-	-
Fracionário	09:30	09:45	09:45	10:00	10:00	17:55	17:55	18:00	-	-	-	-
Mercado a termo	-	-	-	-	10:00	18:20	-	-	-	-	-	-
Mercado de opções	09:30	09:45	09:45	10:00	10:00	17:55	17:55	18:15	-	-	-	-
BOVESPA Mais - Todos os ativos	09:30	09:45	09:45	10:00	10:00	17:55	17:55	18:00	-	-	-	-
ETFs	09:30	09:45	09:45	10:00	10:00	17:55	17:55	18:15	-	-	-	-
Mercado de balcão organizado- Todos os ativos	09:30	09:45	09:45	10:00	10:00	17:55	17:55	18:00	-	-	-	-
Opções sobre índice de ações	09:30	09:45	09:45	10:00	10:00	17:50	17:50	18:15	-	-	-	-

Figura 2.2: Horários de negociação para cada tipo de mercado.

de forma a determinar a força do relacionamento, ou o tamanho do efeito, onde o valor do coeficiente estando entre 0,10 e 0,29 representa uma pequena associação, entre 0,30 e 0,49 representa uma associação média e acima de 0,50 representa uma associação maior ou mesmo um relacionamento. A Tabela 2.3 descreve em detalhe a escala de Cohen.

Com base nesses dois conceitos, definimos uma escala de correlação ilustrada na Tabela 2.4 com o objetivo de identificar a correlação como positiva, negativa ou inexistente, sem, necessariamente, considerar maiores variações na magnitude da correlação.

Coeficiente Correlação	Descrição
0.0-0.1	trivial, muito baixo, insignificante
0.1-0.3	pequeno, baixo, menor
0.3-0.5	moderado, médio
0.5-0.7	elevado, alto, maio
0.7-0.9	muito elevado, muito alto, gigante
0.9-1	praticamente ou quase: perfeito, distinto, infinito

Tabela 2.3: Escala de Cohen.

Coefficiente de Serman	Descrição
$-1 \leq \rho \leq -0.1$	negativamente correlacionado
$-0.1 \leq \rho \leq 0.1$	não correlacionado
$0.1 \leq \rho \leq 1$	positivamente correlacionado

Tabela 2.4: Escala positiva-negativa baseada na escala de Cohen.

#### 2.1.4.2 Dimensões dos Ativos

Para a nossa análise de correlação selecionamos quatro características dos ativos que servem tanto de referência para os investidores em ações, como para as próprias empresas e para o desempenho da bolsa. Essas quatro características foram denominadas nesta dissertação de dimensões de variação (ou “dimensões”) e são as seguintes:

1. **Numero Total de Ordens:** quando um investidor deseja realizar a compra ou a venda de uma ação negociada no mercado, ele o faz através do envio de uma ordem ao intermediário contratado. De acordo com o Portal do Investidor<sup>7</sup>, *site* educativo da Comissão de Valores Monetários - CVM<sup>8</sup>, ordem é o ato pelo qual o cliente determina que um intermediário negocie ou registre uma operação com valor mobiliário, em seu nome e nas condições que especificar. Nesse caso, o número total de ordens indica quantas ordens de compra ou venda um determinado ativo recebeu ao longo do dia.
2. **Volume Financeiro:** o volume é originado após a concretização de um negócio em um ativo e pode ser classificado em três tipos: volume de ações, volume financeiro e volume de negócios. Em nosso estudo utilizamos somente o volume financeiro. Para entender melhor o cálculo desse volume, vamos considerar como exemplo a compra de 1.000 ações da Petrobras (PETR4) no valor de R\$25,00. Na execução total desta única ordem obtivemos um total financeiro movimentado de R\$25.000. Esse valor corresponde ao volume financeiro.
3. **Oscilação Diária:** para exemplificar a oscilação diária de um ativo, vamos retomar a ilustração do *candlestick* da Figura 1.1. Nela é possível ver que todo ativo possui um valor de abertura e um valor de fechamento, ambos relativos ao início e fim do pregão de negociação da bolsa, respectivamente. A oscilação diária é a diferença entre o valor de fechamento e o valor de abertura. Por exemplo, caso a ação da VALE (VALE5) inicie o pregão com o valor de R\$30,00 e termine

<sup>7</sup><http://www.portaldoinvestidor.gov.br>

<sup>8</sup><http://www.cvm.gov.br>

com o valor de R\$31,00, então a oscilação diária é de R\$1,00 (Nesse caso houve um aumento do valor da ação no dia).

4. **Oscilação Diária Máxima:** semelhante à dimensão anterior, a partir da Figura 1.1 podemos compreender a oscilação diária máxima como sendo a diferença entre o maior e o menor valor alcançado pela ação durante o dia de negociação. Como no exemplo anterior, caso a ação da VALE (VALE5) inicie o pregão valendo R\$30,00 e termine com o valor de R\$31,00, mas durante o pregão tenha atingido a mínima de R\$29,80 e a máxima de R\$31,10, então o valor para a oscilação diária máxima é de  $R\$29,80 - R\$31,10 = R\$1,30$ . É importante ressaltar como essa dimensão difere do valor da dimensão anterior, pois, em contextos em que a bolsa apresente elevada volatilidade e comportamento especulativo, a oscilação diária máxima poderá refletir melhor esses comportamentos do que a própria oscilação diária.

Em particular, para as dimensões 3 e 4 trabalhamos com a variação em módulo percentual, tendo em vista que ao compararmos as séries temporais para o cálculo do coeficiente de Spearman, não faz sentido uma quantidade de menções negativas aos ativos durante o dia no Twitter.

### 2.1.5 Análise de Conteúdo

Nesta última etapa de análise que faz parte da visão do mercado de ações, entendemos que compreender os padrões presentes nas postagens dos usuários no Twitter é importante, pois nos permite mapear quais tipos de informação esses usuários compartilham com maior frequência. Para isso, nos concentramos no estudo das URLs compartilhadas, na classificação delas, no mapeamento dos usuários que as publicam (fontes) e, por fim, no tempo de exposição delas na rede.

Em particular, nessa etapa, o primeiro passo foi compreender a estrutura dos *links* compactados. Basicamente sabemos que analisar diretamente os *links* compactados não é conclusivo, pois esses *links* possuem ambiguidade, uma vez que, para uma dada URL completa e única, podem haver diversos *links* compactados que apontam para aquele mesmo endereço. Em seguida, realizamos a classificação manual das URLs mais compartilhadas em quatro categorias: opinião, análise financeira, dados técnicos e notícia. Feito isso, estudamos o tempo de exposição das notícias na rede através do cálculo do intervalo de confiança da subcoleção formada somente pelos *tweets* de maior ocorrência classificados como notícias.



Na próxima seção, apresentamos nossas questões de pesquisa levantadas sobre as coleções de dados do Twitter e da BOVESPA.

## 2.2 Questões de Pesquisa

Tendo coletado e pré-processado os dados para o reconhecimento das entidades, nosso objetivo concentrou-se em caracterizar de forma abrangente quais tipos de informação são postados no Twitter diariamente sobre o mercado financeiro e, sobretudo, se as informações vinculadas possuem correlação com o comportamento dos ativos na BOVESPA e vice-versa.

A análise experimental dos dados coletados foi conduzida de forma a responder quatro questões de pesquisa:

- Q1:** Qual é o volume de dados compartilhados no Twitter sobre o mercado de ações brasileiro? Existe um nicho específico entre os usuários do Twitter? Se sim, são eles investidores reais?
- Q2:** Quais padrões temporais e comportamentais são refletidos no Twitter a partir da análise da variação dos dados ao longo do tempo?
- Q3:** Existe uma correlação entre a oscilação das ações e seu respectivo número de menções em *posts* do Twitter?
- Q4:** Quais tipos de conteúdo são compartilhados com maior frequência no Twitter e qual a sua fonte?

Cada capítulo a seguir aborda a pesquisa desenvolvida, orientada por uma das perguntas propostas, na ordem em que foram levantadas.

Para a primeira questão de pesquisa (Q1), a análise estática, abordada no Capítulo 3 apresenta apenas estatísticas gerais dos *tweets* e os resultados da pesquisa com os usuários, desconsiderando qualquer aspecto temporal.

A segunda pergunta da pesquisa (Q2), por sua vez, é abordada no Capítulo 4, que complementa a análise estática considerando os padrões comportamentais dos dados ao longo dos 13 meses.

Para a terceira questão (Q3), a análise de correlação é abordada no Capítulo 5, onde avaliamos a correlação entre as postagens no Twitter e os dados dos ativos pertencentes ao IBOVESPA.

Por fim, a quarta questão de pesquisa é abordada no Capítulo 6. Nele realizamos a análise de conteúdo sobre as URLs postadas nos textos dos *tweets* em busca da

identificação das suas fontes e endereços para páginas Web. Além disso, classificamos os *tweets* em função dessas URLs e, finalmente, calculamos o tempo de exposição dessas postagens na rede.

# Capítulo 3

## Análise Estática

Neste capítulo abordamos a questão de pesquisa Q1, analisando os atributos dos *tweets* postados que mencionam explicitamente algum código de ação da BOVESPA. Para isso, dividimos o capítulo em duas seções: na primeira realizamos o mapeamento da coleção de acordo com os atributos dos *tweets* e, na segunda, apresentamos os resultados da pesquisa conduzida com os usuários que realizaram as postagens pertencentes à coleção.

### 3.1 Mapeamento das Postagens

A Tabela 3.1 apresenta as principais características da coleção de dados. Como podemos ver, apenas 3,1% dos *tweets* podem ser classificados como relacionadas com o mercado de ações, ou seja, eles mencionam explicitamente um código de um determinado ativo pertencente ao IBOVESPA ou o próprio índice (ver Apêndice A). Além disso, esses *tweets* foram postados por um nicho de 0,5% dos usuários considerados, o que significa que este grupo específico de usuários está realmente interessado no mercado de ações, tendo em vista o conhecimento especializado envolvido nos textos das postagens.

<i>#tweets</i> que mencionam ações	93,179 (3,1% de 2,973,410)
<i>#usuários</i> interessados em ações	4,816 (0,5% de 1,194,771)
<i>#retweets</i>	4,467 (5,0%)
<i>#replies</i>	2,657 (2,8%)
<i>#tweets</i> com URL	74,293 (80,3%)

Tabela 3.1: Análise estática dos *tweets*.

De acordo com as métricas adotadas por Benevenuto et al. [2010], o grande número de *links* encontrados nesses *tweets* sugere um padrão comportamental caracteri-

zado por postagens de usuários que estão principalmente interessados na divulgação de novas informações (notícias) ou mesmo um comportamento de *SPAM*. Diante disso, aprofundamos nossos estudos na análise dessas URLs que será exibido posteriormente no Capítulo 6.

Em particular, analisamos para todo o período dos dados quais foram os códigos mais mencionados em nossa coleção, entre códigos dos ativos e do próprio IBOVESPA. O resultado está na Figura 3.1. Nela é possível perceber os cinco códigos mais mencionados. Em primeiro o código IBOV11 (IBOV), que representa o próprio índice, em seguida os códigos da Petrobras (PETR4), OGX (OGXP3), novamente Petrobras (PETR3) e Bradesco (BBDC4).



Figura 3.1: Distribuição das menções aos ativos do IBOVESPA

Vale ressaltar que o código do IBOVESPA é o mais referenciado pelo fato dos *tweets* mencionarem frequentemente o seu código junto com outros ativos, conforme *tweets* de exemplo a seguir. A Petrobras, por sua vez, se mostrou a empresa de maior popularidade no Twitter durante o período de análise, já que as suas duas ações apareceram entre as cinco mais mencionadas, conforme os *tweets* de exemplo abaixo:

- “Depois de várias quedas as bandas de bolinger estão se fechando nos gráficos diários do #Ibov #Vale5 #Petr4”;
- “Novos estudos graficos do indice bovespa (#IBOV) e outras acoes disponiveis em <http://t.co/sIwG4LrsRu> #IBOV #OGXP3 #EIKE OGX suspende trabalhos em três campos: <http://t.co/sgPdswjzRS> #AL”;

Uma vez que identificamos que 4.816 usuários postaram mensagens no Twitter mencionando explicitamente ações do IBOVESPA, endereçamos a questão de pesquisa seguinte tendo em vista a ausência de estudos na literatura sobre isso. Já que identificamos um nicho específico de usuários do Twitter interessados no mercado de ações, é possível afirmar que esses usuários são investidores reais? Abordamos essa pergunta na seção seguinte.

## 3.2 Pesquisa com os Usuários

Nesta seção caracterizamos o perfil dos usuários do Twitter que postam informações sobre o mercado financeiro. Assim, estendemos nossa análise estática a fim de investigar se os usuários que postam mensagens relacionadas ao mercado financeiro são, de fato, investidores e quais são os seus principais interesses.

Para isso, foram selecionados 3.302 usuários daqueles 4.816 interessados em ações que postaram pelo menos uma vez no período de 1º de Janeiro à 31 de julho de 2014. A partir disso, implementamos um robô para postar *tweets* convidando os referidos usuários a participar da pesquisa através de um *link* específico e responder a um breve formulário elaborado no Google Forms<sup>1</sup>.

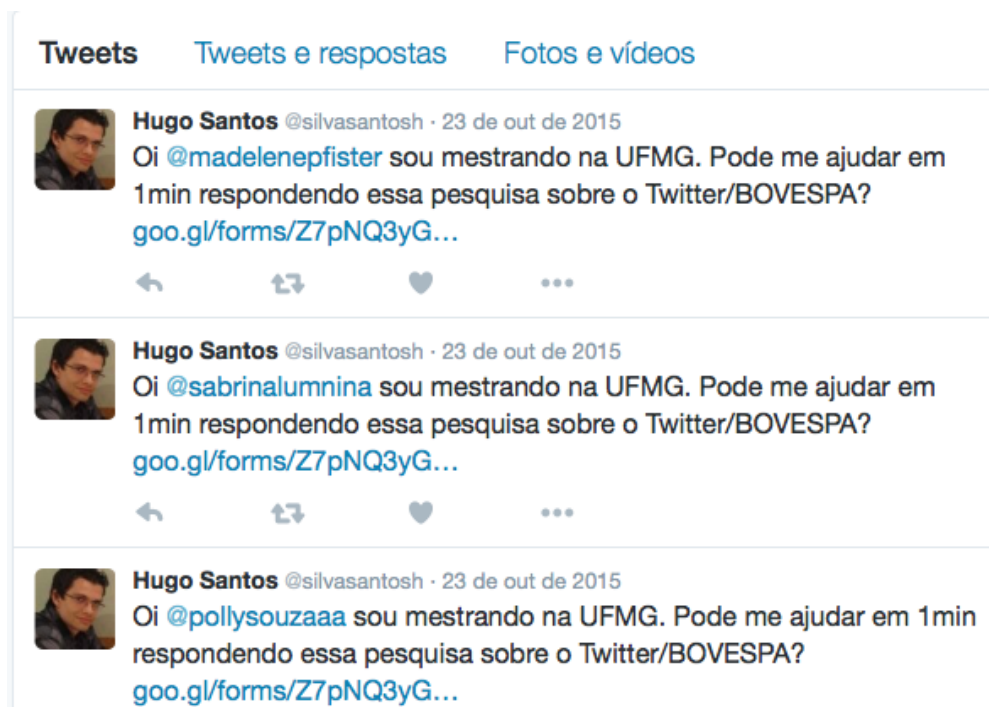


Figura 3.2: Postagens do robô convidando os usuários a participar da pesquisa

Após a disseminação dos convites pelo robô (Figura 3.1) foi possível obter 330 respostas de usuários distintos. As perguntas de 1 a 6 e a compilação das respostas dos usuários são apresentadas na Tabela 3.2.

<sup>1</sup><http://goo.gl/forms/67y6tYQmDN>

1) Como você se enquadra em relação ao mercado de ações brasileiro?		2) Em qual das opções a seguir você melhor se enquadra?	
a) Eu sou investidor	63,0%	a) Investidor individual	71,5%
b) Eu não sou investidor, mas pretendo investir no futuro	30,9%	b) Agente Independente de Investimento	7,0%
c) Eu não sou investidor e não pretendo investir no futuro	6,1%	c) Profissional de Mídia Especializada	3,0%
		d) Analista Financeiro	4,2%
		e) Profissional de Relacionamento com Investidor	0,9%
		f) Nenhuma das anteriores	13,3%
3) No contexto do mercado financeiro, você posta <i>tweets</i> com maior frequência sobre?		4) De forma geral, qual o tipo de postagem de outros usuários sobre o mercado financeiro você considera mais relevante?	
a) Sua opinião pessoal	30,0%	a) Opiniões pessoais	16,4%
b) Análise técnica	9,1%	b) Notícias	32,1%
c) Notícias	38,2%	c) Análise financeira	44,2%
d) Outros assuntos	4,2%	d) Outros assuntos	5,2%
e) Não posto	18,5%	e) Eu não uso Twitter	2,1%
5) Os <i>tweets</i> que você posta sobre o mercado financeiro são geralmente relacionados à sua carteira de ações?		6) Você usa ou já usou informação proveniente do Twitter para decidir sobre algum investimento?	
a) Sim	24,2%	a) Sim:	42,4%
b) Não	51,8%	b) Não:	57,6%
c) Eu não posto <i>tweets</i> sobre mercado de ações	23,9%		

Tabela 3.2: Resultado da pesquisa com os usuários do Twitter

### 3.3 Caracterização dos Usuários

A partir da pesquisa realizada foi possível identificar que 63,0% dos usuários consideram-se investidores (Pergunta 1) e que a maioria deles considera-se investidor individual (Pergunta 2). Podemos perceber também que esses usuários postam mais frequentemente sobre notícias e suas opiniões pessoais (Pergunta 3), ao mesmo tempo em que estão interessados principalmente em notícias e análises financeiras compartilhadas na rede (Pergunta 4).

Além disso, quando analisamos os resultados para as perguntas que abordam o processo decisório de investimento desses usuários, identificamos que 42,4% deles

afirmam usar, ou já ter usado, informação proveniente do Twitter para decidir sobre algum investimento. Por fim, descartamos a possibilidade desse grupo postar sobre ativos que pertencem à sua carteira, pois 75,7% dele afirmaram não postar sobre sua carteira de ações ou mesmo sobre o mercado financeiro. A Figura 3.2 apresenta uma perspectiva dos perfis mapeados a partir dos resultados da pesquisa realizada nessa etapa.



Figura 3.3: Perfis encontrados a partir dos resultados da pesquisa

Considerando a Hipótese do Mercado Eficiente (EMH) [Fama, 1965], que sugere que os preços das ações são mais influenciados por novas informações do que preços passados e presentes, é possível dizer que há um padrão comportamental em torno dos usuários do Twitter interessados em usar a rede social como canal de informação o que, conseqüentemente, torna esse serviço capaz de influenciar o mercado financeiro e vice-versa.

### 3.4 Conclusões

Considerando a primeira questão de pesquisa levantada (Seção 2.2), sobre os padrões que emergem a partir da análise estática dos atributos ao longo dos 13 meses de análise dos *tweets*, a análise apresentada neste capítulo revela uma interessante perspectiva do mercado de ações brasileiro.

Em particular, apesar de pequena (3,1%), a quantidade de *tweets* postados que fazem menção explícita a ativos do mercado financeiro não pode ser vista como inexpressiva. Em seguida, identificamos que, de fato, há um grupo especializado de usuários do Twitter que o utilizam interessados em acompanhar as notícias e análises financeiras relacionadas aos ativos da BOVESPA. Além disso, verificamos que o próprio índice

IBOVESPA e as ações da Petrobras apresentam os códigos mais frequentemente mencionados na rede.

Também foi levantado um interessante paralelo com pesquisas anteriores no momento em que perguntamos diretamente para os usuários do Twitter como eles se percebem diante do mercado financeiro, ponto até então não abordado diretamente pela literatura. Nessa etapa foi possível identificar os interesses desses usuários e caracterizá-los principalmente como investidores individuais em sua maioria, interessados em acompanhar notícias e análises financeiras, além de não postarem informações sobre sua carteira de ações. Além disso, aproximadamente metade deles admite se basear ou já ter se baseado em informação do Twitter para tomar alguma decisão de investimento.

Por fim, concluímos este capítulo traçando um paralelo dos resultados encontrados e a Hipótese do Mercado Eficiente [Fama, 1965]. Nesse paralelo identificamos a possibilidade do Twitter ser influenciado e influenciar o comportamento dos ativos na BOVESPA. Essa possibilidade é justificada pelo interesse dos usuários da rede social em compartilhar entre si notícias e análises financeiras que julgarem relevantes para o mercado.



# Capítulo 4

## Análise Temporal

No capítulo anterior, realizamos uma análise estática do conjunto de dados do Twitter com base apenas nos atributos dos *tweets*. Neste capítulo, abordamos a questão de pesquisa Q2 por meio da análise deste mesmo conjunto de dados ao longo do período de 13 meses. Mais especificamente, estudamos como as mensagens diárias variam ao longo dos dias da semana e os dias de negociação da BOVESPA, o número de postagens por hora do dia e, finalmente, a segmentação dos usuários responsáveis pelas postagens.

### 4.1 Padrões Temporais das Postagens

Uma vez que os usuários da nossa coleção se consideram, em sua maioria, investidores, é esperado que existam padrões temporais relacionados com a sua frequência de postagens como, por exemplo, a que horas ocorrem os maiores volumes de postagem, se determinadas notícias tidas como relevantes pelo mercado financeiro são capazes de gerar picos de publicação no Twitter, entre outros.

A frequência das mensagens pode variar de acordo com rotinas diárias dos usuários, dias de negociação da BOVESPA e dias úteis da semana. Na Tabela 4.1 apresentamos os primeiros e últimos sete dias da nossa coleção, e o total de menções aos ativos postado nesses mesmos dias. É possível observar pelos valores a queda do volume de *tweets* durante os finais de semana, por exemplo, o sábado e o domingo apresentam totais de menções abaixo da média dos dias úteis. Vale observar que no dia 31 de julho de 2014 temos um total significativo de postagens. Analisaremos esse resultado logo mais.

Na Figura 4.1 ilustramos uma compilação de todas as semanas do nosso intervalo. Nela observamos claramente para todo o período de análise como o padrão de postagens segue os dias de pregão da BOVESPA, tendo um número regular de mensagens durante

<i>Dia da Semana</i>	<i>Total de Menções</i>
segunda-feira, 1 de julho de 2013	262
terça-feira, 2 de julho de 2013	271
quarta-feira, 3 de julho de 2013	355
quinta-feira, 4 de julho de 2013	348
sexta-feira, 5 de julho de 2013	458
sábado, 6 de julho de 2013	24
domingo, 7 de julho de 2013	9
...	...
sexta-feira, 25 de julho de 2014	139
sábado, 26 de julho de 2014	23
domingo, 27 de julho de 2014	40
segunda-feira, 28 de julho de 2014	172
terça-feira, 29 de julho de 2014	132
quarta-feira, 30 de julho de 2014	390
<i>quinta-feira, 31 de julho de 2014</i>	<i>1.063</i>

Tabela 4.1: Primeiros e últimos sete dias da coleção e seus respectivos totais de menções postadas

os dias úteis da semana e um volume quase irrelevante durante finais de semana, conforme mostrado anteriormente na Tabela 4.1.

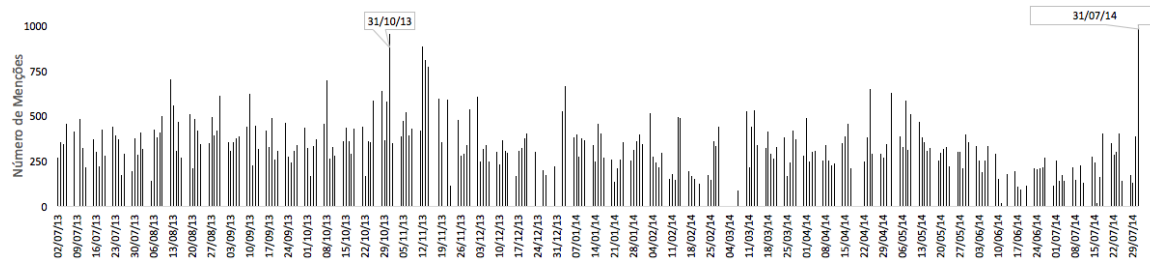


Figura 4.1: Número de *tweets* por dia ao longo dos 13 meses

Observando a Figura 4.1, também é possível identificar claramente picos de postagens de *tweets*. Em uma pesquisa por notícias e em canais de relacionamento com investidores, verificamos que esses picos representam eventos significativos no mercado de ações. Por exemplo, o dia 31 julho de 2014 pode ser explicado por uma sequência de resultados positivos publicados por empresas brasileiras no segundo trimestre daquele ano. Esses resultados incluem a alta do lucro do Bradesco, um lucro de 283% da Vale (VALE5), inversão de prejuízo em lucro pela Embraer (EMBR3), além do aumento de 16% do lucro da Ambev (ABEV3) devido à Copa do Mundo realizada no Brasil. Os *tweets* listados abaixo representam uma pequena amostra do que foi postado nesse dia:

- “Embraer reverte prejuízo e tem lucro de R\$ 327 milhões no 2º tri: A Embraer (EMBR3) teve lucro de R\$ 327,2 ... <http://t.co/IN2lcqWITD>”;
- “Com efeito Copa, lucro da Ambev sobe 16% no trimestre, a R\$ 2,2 bilhões: A Ambev (ABEV3), maior empresa de ... <http://t.co/Hdxcr00zda>”;
- “Vale registra lucro líquido de R\$ 3,187 bilhões no 2º trimestre: Resultado foi 283% maior na... <http://t.co/5f7fu6nz0I> #infomoney #vale5”;
- “Lucro do Bradesco sobe 28% com alta de crédito e queda da inadimplência: O Bradesco (BBDC4) registrou lucro... <http://t.co/dN6GKWDal5>”.

Outro dia atípico em que houve pico de postagens foi o dia 31 de outubro de 2013. Nesse dia verificamos que as postagens são em sua maioria relacionadas com a retirada das ações da empresa OGX da composição do IBOVESPA<sup>1</sup>.

Para entender melhor os padrões de postagens dos usuários, sobretudo, se a frequência de *tweets* acompanha os horários de negociação da bolsa de valores, também analisamos a distribuição de *tweets* em uma granularidade menor de tempo. Assim, a Figura 4.2 mostra a distribuição de *tweets* através das horas do dia para todo o período de análise.

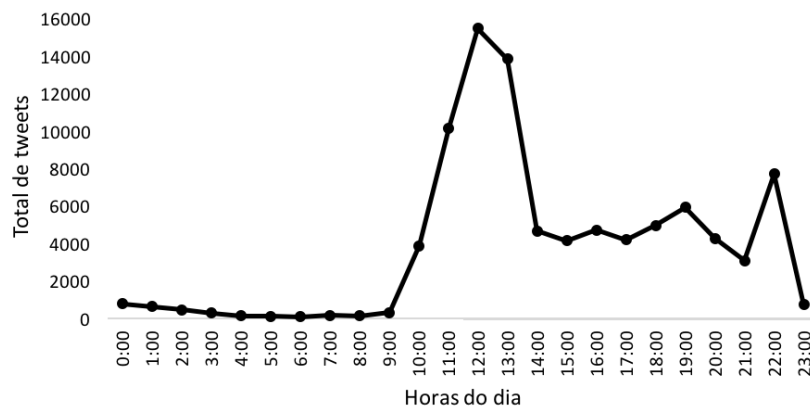


Figura 4.2: Postagens sumarizadas por hora

Sabe-se que o mercado inicia as suas negociações às 10 horas da manhã e termina às 17 horas. Assim, a partir da Figura 4.2 observa-se que o período em que os usuários estão mais engajados em postar no Twitter é de 11 horas às 13 horas. Além disso,

<sup>1</sup>Valor Econômico - <http://goo.gl/6UVav9>

observamos que a frequência de postagens no Twitter permanece após o fechamento do mercado, o que ocorre às 18 horas. Esse comportamento pode ser justificado por fatores como a abertura e fechamento de bolsas no exterior, notícias relacionadas publicadas em noticiário especializado e, também, informações estratégicas publicadas pelas empresas que ocorrem após o fechamento do mercado, de acordo com a regulamentação da CVM<sup>2</sup>.

## 4.2 Participação dos Usuários

Na seção anterior analisamos a frequência de postagens ao longo dos 13 meses relativos ao período de coleta, tendo como referência os valores diários e os horários das postagens. A partir disso, mapeamos como o volume de postagens acompanha os horários de negociação

Nesta seção vamos concentrar a nossa análise no entendimento de como ocorre e está distribuída a participação dos usuários no Twitter. Para isso, na Figura 4.3, traçamos a distribuição de mensagens para todo o nosso período de análise. Nela ilustramos o percentual dos *tweets* em função do percentual de participação dos usuários).

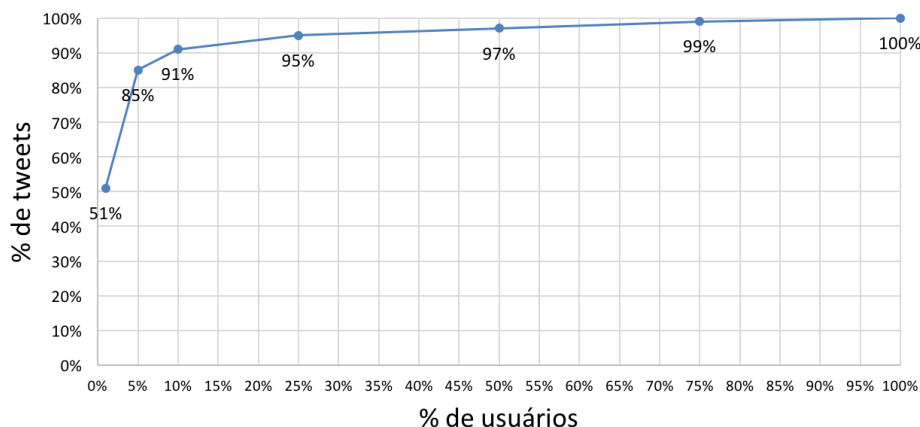


Figura 4.3: Distribuição da participação dos usuários

A partir da distribuição da Figura 4.3, pode-se observar que 10% dos usuários são responsáveis pela geração de mais de 90% dos *tweets* coletados. A partir disso, é possível identificar alguns grupos de usuários com alto nível de envolvimento na geração de conteúdo, enquanto os demais apenas atuam como consumidores. Acreditamos que esses usuários geradores de conteúdo podem agir como *influenciadores* na rede social por postar informações aos outros 90% dos usuários.

<sup>2</sup>Comissão de Valores Mobiliários - <http://www.cvm.gov.br>

Em particular, nesta etapa também buscamos compreender se, necessariamente, uma maior participação de usuários distintos representa um maior volume de postagens na rede. Por exemplo, para os meses em que houve um maior número de postagens houve, também, um maior engajamento de diferentes usuários? Os resultados estão ilustrados na Figura 4.4:

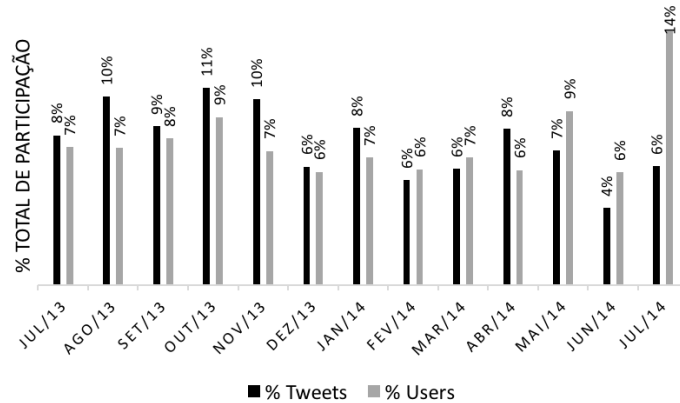


Figura 4.4: Distribuição da participação dos usuários em relação ao volume total de *tweets* por mês

Diante dos resultados encontrados, não é possível determinar que exista uma correlação positiva entre o aumento do volume de postagem e o respectivo aumento do volume de participação de diferentes usuários. Por exemplo, de forma mais significativa, o último mês da nossa análise, isto é, julho de 2014, percebemos que ele representou 6% dos *tweets* de toda a nossa coleção, mas obteve 14% de participação de usuários. Por outro lado, em outubro de 2013, há o maior número de *tweets* postados, mas somente 9% dos usuários participaram, valor consideravelmente menor do que o último mês da coleção.

### 4.3 Conclusões

Neste capítulo, abordamos a questão de pesquisa Q2 (Seção 2.2) por meio da análise da nossa coleção ao longo do período de 13 meses. A análise apresentada neste capítulo fornece indícios para a construção do estudo do mercado brasileiro de ações tema desta dissertação. Mais especificamente, estudamos a variação do volume de mensagens ao longo dos dias da semana e dos dias de negociação da BOVESPA, o número de postagens por hora do dia e, finalmente, a participação dos usuários nas postagens.

Inicialmente mapeamos como a frequência das mensagens pode variar de acordo com as rotinas diárias dos usuários. Observamos a queda significativa do número de *tweets* durante os finais de semana, ou seja, dias em que não há pregão de negociação na bolsa. Também identificamos como os eventos do mercado financeiro, tido como relevantes pelos usuários do Twitter, são capazes de gerar picos de postagens. Esse resultado demonstra uma ligação com os resultados encontrados no capítulo anterior, em que mostramos que a rede social é um importante canal para propagação de informações de investimento pelos usuários que se consideram investidores individuais.

Em particular, neste capítulo também identificamos que a frequência de postagens ao longo das horas do dia iniciam-se, em sua maioria, no mesmo horário de abertura do pregão de negociação da BOVESPA, mas que de forma contraintuitiva, se mantêm por cerca de duas a três horas após o fechamento das negociações. Este comportamento pode ser explicado pelo acompanhamento dos noticiários, abertura das bolsas externas e divulgação de resultados operacionais das empresas.

Por fim, fechamos o capítulo analisando como ocorrem as participações dos usuários em relação às postagens no Twitter. Verificamos que 10% dos usuários podem atuar como influenciadores na rede, pois publicam mais de 90% do conteúdo. Também identificamos na Figura 4.4 que não há uma relação direta entre o aumento do número de *tweets* e o aumento da participação dos usuários.

# Capítulo 5

## Análise de Correlação

Nos capítulos anteriores nos concentramos na análise apenas sobre os dados do Twitter. Neste capítulo, abordamos a questão de pesquisa Q3 (Seção 2.2) através do estudo de correlação entre o comportamento diário das ações da BOVESPA e da frequência de menções a estas ações no Twitter. Nossa análise é baseada no coeficiente de correlação de Spearman [Zar, 1972] e na escala de Cohen [Cohen et al., 2013], conforme apresentados a Seção 2.1.4 do Capítulo 2.

### 5.1 Correlação entre as Quatro Dimensões e as Postagens

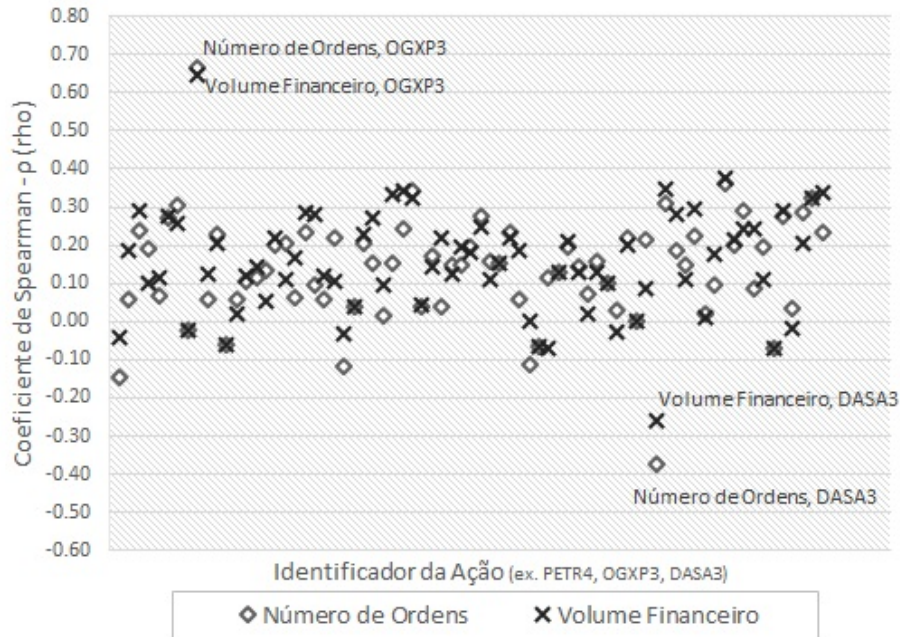
A partir da seleção das quatro dimensões de referência dos ativos do BOVESPA (Seção 2.1.4.2), construímos suas respectivas séries temporais a partir da nossa coleção de dados de *tweets* e *candlesticks*. Feito isso, para cada par da série “menção-*tweet* versus dimensão-ativo” foi calculado o coeficiente de correlação de Spearman. Os resultados para as 73 ações consideradas nesta dissertação estão apresentadas na Figura 5.1.

Como podemos ver na Figura 5.1(a), para algumas ações, tais como OGXP3 e DASA3, há uma correlação significativa entre o número de menções no Twitter e as dimensões *número total de ordens* e volume financeiro. Essas duas ações atingiram valores extremos em nossos experimentos para o coeficiente de Spearman, sendo 0,66 para OGXP3 e -0,36 para DASA3, ou seja, eles mostram, respectivamente, uma alta correlação positiva e uma correlação negativa mediana para o *número total de ordens* e o *volume financeiro*, respectivamente. Por outro lado, a Figura 5.1(b) mostra valores extremos de aproximadamente 0,40 para OGXP3 e -0,25 para LREN3, indicando que o número de menções dessas duas ações no Twitter têm, respectivamente, uma moderada

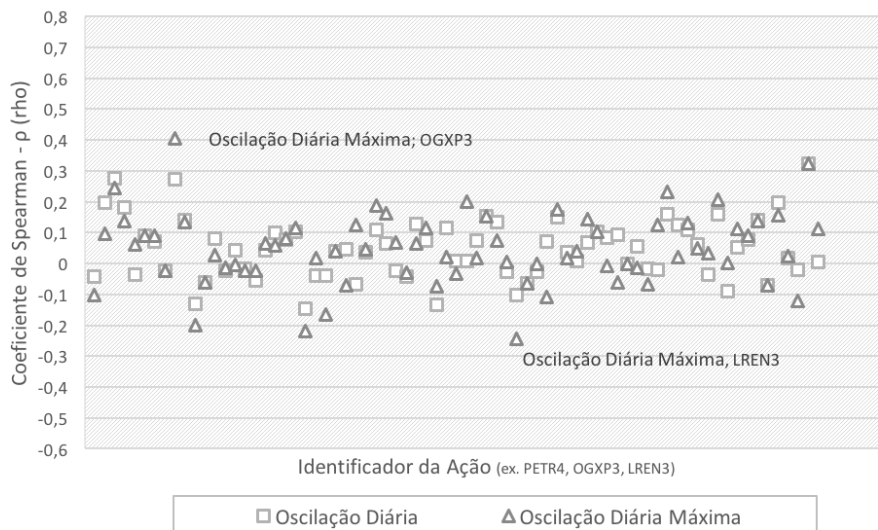
correlação positiva e uma baixa correlação negativa com a dimensão oscilação diária máxima.

Figura 5.1: Coeficiente de Spearman calculado para cada ação.

(a) Coeficiente para as dimensões de número total de ordens e o volume financeiro.



(b) Coeficiente para as dimensões de oscilação diária e oscilação diária máxima.





Apesar dos valores extremos mencionados no parágrafo anterior, os resultados das Figuras 5.1(a) e (b) mostram que a maioria das ações apresentam valores do coeficiente de Spearman de “não correlação”, de acordo com a nossa escala (Tabela 2.4) e, sobretudo, que esses valores são divergentes dependendo da ação e dimensão consideradas. Assim, a fim de aprofundar a nossa análise, calculamos para cada dimensão seu percentual total de correlação em relação ao total das ações em questão que foram mencionadas no Twitter, considerando novamente a nossa escala positiva-negativa. Para isso, consideramos primeiro 100% dos dias analisados e, em seguida, selecionamos apenas aquelas ações que apresentaram pelo menos uma menção a cada dois dias. Os resultados são apresentados nas Tabelas 5.1 e 5.2.

Dimensão (100%)	Correlação das 73 ações		
	Negativa	Sem Correlação	Positiva
Número Total de Ordens	5%	29%	66%
Volume Financeiro	1%	25%	74%
Oscilação Diária	5%	66%	29%
Oscilação Diária Máxima	10%	58%	33%

Tabela 5.1: Percentual de correlação para cada dimensão considerando o conjunto de todos os *tweets* que mencionaram as ações

Dimensão (50%)	Correlação das 31 ações selecionadas		
	Negativa	Sem Correlação	Positiva
Número Total de Ordens	3%	26%	71%
Volume Financeiro	0%	16%	84%
Oscilação Diária	3%	58%	39%
Oscilação Diária Máxima	6%	48%	45%

Tabela 5.2: Percentual de correlação para cada dimensão considerando o conjunto de todos os *tweets* que mencionaram as 31 ações.

Os resultados apresentados nas Tabelas 5.1 e 5.2 mostram que há uma correlação positiva entre o número total de ordens e o volume financeiro com respeito às menções dos ativos no Twitter, isto é, à medida que o volume de negociação financeira aumenta, há a tendência de que o número de menções no Twitter aumente e vice-versa. Por outro lado, ambas as tabelas mostram que não é possível determinar uma correlação entre o número de menções no Twitter e as dimensões de oscilação diárias e oscilação diária máxima.

## 5.2 Ações Correlacionadas

A partir dos resultados da seção anterior, encontramos que, para as dimensões de volume financeiro e número total de ordens, os usuários do Twitter podem obter a partir da rede social alguns indicadores sobre o comportamento das ações na BOVESPA. Vale ressaltar que essas duas dimensões são importantes para o mercado financeiro, pois elas podem ser utilizadas pelos investidores como referência de liquidez financeira dos ativos.

Positiva	EMBR3, OGXP3, BBDC4, VIVT4, BVMF3, BBDC3, LLXL3, CRUZ3, LIGT3, BTOW3, GFSA3, USIM5, ITUB4, SBSP3, OIBR4, ALLL3, BBAS3, VALE5, PETR3, PETR4, CSNA3, PDGR3
Sem Correlação	MMXM3, BISA3, MRVE3, NATU3, BRML3

Tabela 5.3: Conjunto das 22 ações que são positivamente correlacionadas com o volume financeiro e o número total de ordens e as cinco que não apresentaram correlação.

A Tabela 5.3 mostra o conjunto de ações que apresentaram sua frequência de menções no Twitter correlacionadas com o volume financeiro e número total de ordens na BOVESPA, ao mesmo tempo. Essa tabela inclui as 22 ações com correlação positiva que foram consideradas relevantes pelo nosso grupo de usuários do Twitter. Ela inclui, por exemplo, ações de empresas de petróleo e aço, como Petrobras (PETR4), Usiminas (USIM5) e Vale (VALE5), bem como grandes bancos como Bradesco (BBDC4), Banco do Brasil (BBAS3) e Itaú Unibanco (ITUB4). Também identificamos que cinco ações (MMXM3, BISA3, MRVE3, NATU3 e BRML3), em geral, não possuem um padrão de mensagens correlacionados com qualquer uma das dimensões consideradas. Ou seja, não há nenhuma evidência para inferir qualquer padrão de postagem baseado em menções a esses ativos no Twitter que fosse útil como um indicador para investidores interessados nessas ações. Finalmente, as últimas quatro ações apresentaram resultados inconclusivos, pois não são nem negativas e nem positivamente correlacionadas com as duas dimensões ao mesmo tempo.

Finalmente, a partir dos resultados desta seção, é possível ressaltar que eles demonstram também que as ações mais populares pertencentes ao IBOVESPA são também as mais mencionadas no Twitter e, portanto, tendem a apresentar uma correlação com o seu comportamento na BOVESPA.

## 5.3 Conclusões

Neste capítulo, abordamos a questão de pesquisa Q3 (Seção 2.2) por meio da análise de correlação entre o comportamento diário das ações da BOVESPA e a frequência de menções a essas ações no Twitter. A análise apresentada neste capítulo fornece resultados inéditos no cenário brasileiro.

Em particular, na Seção 5.1 identificamos que as dimensões volume financeiro e número total de ordens dos ativos apresentam correlação positiva com a frequência de menções no Twitter para alguns ativos. Verificamos também que não foi possível inferir qualquer correlação entre as dimensões oscilação diária e oscilação diária máxima e as menções às ações no Twitter. Ao selecionarmos as ações que apresentaram pelo menos uma menção a cada dois dias, evidenciamos que as dimensões volume financeiro e número de ordem são as que possuem maior potencial de correlação com as menções dos ativos. Com relação especificamente ao volume financeiro, nossos resultados estão alinhados com os encontrados por Sprenger [2011], mas, no nosso caso, aplicados à BOVESPA.

Na Seção 5.2, por fim, mapeamos as 22 ações que podem ter nas suas respectivas frequências de menção no Twitter indicadores para monitorar seus comportamentos na BOVESPA, ou seja, demonstramos que partir da análise dos dados do serviço de *microblogging* é possível estruturar indicadores que ofereçam ao investidor maiores informações para o seu processo decisório, principalmente, quando a análise financeira realizada pelo investidor considerar como relevante o volume financeiro e o número total de ordens do ativo.



# Capítulo 6

## Análise de Conteúdo

Conforme a Seção 3.1, identificamos que cerca de 80% dos *tweets* postados continham URLs em seu texto. Diante deste dado significativo, neste capítulo específico apresentamos um estudo aprofundado das características desses *links*, visando responder à questão de pesquisa Q4 (Seção 2.2).

Os três *tweets* a seguir são uma amostra da nossa coleção e expressam o conjunto de *tweets* que serão objeto de estudo deste capítulo. É possível perceber que todos eles possuem menções a ativos da BOVESPA e incluem uma URL compactada:

- “*RT @ativacorretora: Nossa análise para o pregão de hoje o papel USIM5 Ficou com dúvidas? Fale com a gente! <http://t.co/Kdn5XTD7Zm>”;*
- “*Opinião - As ações mais indicadas pelos analistas para investir nesta semana: As ações da Vale (VALE5) foram a... <http://t.co/LvcAfoQlpe>”;*
- “*Vídeo muito engraçado sobre a dura realidade do iniciante na bolsa: <http://t.co/OAPPIXDdcB> - “Perdendo na Bolsa - #OGXP3 e #CCXC3”;*

Nas próximas seções realizaremos uma análise do conteúdo das URLs postadas nos *tweets* coletados. Nessa análise identificaremos os endereços (*links*) completos e domínios mais compartilhados, classificaremos os *tweets* de acordo com a URL compartilhada, identificaremos suas fontes (usuários) e, finalmente, determinaremos o tempo de exposição das postagens classificadas como notícias.

### 6.1 Classificação e Ordenação de Conteúdo

Conforme mencionado na Seção 2.1.5, para iniciar o nosso processo de análise das URLs o primeiro passo foi compreender a estrutura dos *links* compactados. Diante disso,

realizamos a extração das URLs da nossa coleção e as submetemos à API do serviço *UNSHORTEN.IT*<sup>1</sup>. Feito isso, para cada URL submetida o serviço retornou o endereço completo correspondente. Após processar todos os 74.858 *tweets*, estruturamos o nosso catálogo de URLs completas para início da análise de conteúdo. As estatísticas iniciais desse catálogo estão na Tabela 6.1 a seguir.

<i>Descrição</i>	<i>Total</i>	<i>%</i>
URLs Completas	71.316	95,2%
URLs Inválidas	3.542	4,8%
URLs Únicas	13.032	18,2%

Tabela 6.1: Catálogo de URLs Completas

Conforme descrito na Tabela 6.1, foi possível descompactar 95,2% das URLs da nossa coleção inicial, estabelecendo um total de 71.316 URLs que serão os *links* objeto no estudo deste capítulo. Vale ressaltar, ainda de acordo com a mesma tabela, que uma vez que obtivemos as URLs completas foi possível realizar o processo de desambiguação, o qual revelou que 18,2% delas são únicas. Por fim, durante o processo de descompactação dos *links* obtivemos algum tipo de erro ao tentar extrair a URL completa para cerca de 4,8%, como *http 404* que indica que o endereço não está mais acessível.

A partir dessa coleção completa de endereços, elaboramos uma ordenação pelo número total de vezes em que cada uma delas ocorreu na rede durante o nosso período de avaliação. A partir disso, verificamos a distribuição, conforme Figura 6.1, e selecionamos manualmente uma amostra dos 130 primeiros endereços que mais se repetem. Para isso, selecionamos as URLs que apresentaram um total de, no mínimo, 43 repetições (ponto indicado na Figura 6.2). Feito isso, realizamos uma inspeção manual de cada um delas com o objetivo de classificar o seu conteúdo entre: Dados Técnicos, Notícia, Análise Financeira ou Opinião. Uma amostra com as quatro primeiras e quatro últimas das 130 URLs selecionadas aparece na Tabela 6.2 (a relação completa está no Apêndice B).

Na Tabela 6.2 podemos perceber que as notícias 4 e 127 que estão entre as mais compartilhadas na rede ocorreram no dia 31/07/2014, data em que, conforme ilustramos na Seção 4.1, verificamos um dos maiores volumes de publicações sobre o mercado de ações no Twitter. Particularmente, as duas URLs dizem respeito a notícias relevantes do mercado financeiro que tratam de bons resultados operacionais das empresas mencionadas nos *tweets*.

<sup>1</sup><http://unshorten.it>

Índice	URL	Ocorrência	Classificação
1	<a href="http://pardotrader.blogspot.com.br/search/label/fechamento">http://pardotrader.blogspot.com.br/search/label/fechamento</a>	232	Dados Técnicos
2	<a href="http://www.bastter.com/mercado/Cadastro.aspx?alerta=login">http://www.bastter.com/mercado/Cadastro.aspx?alerta=login</a>	201	Análise Financeira
3	<a href="http://economia.uol.com.br/noticias/reuters/2014/07/30/bradesco-santander-e-btg-entram-em-novo-socorro-a-eletricas-diz-fonte.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/reuters/2014/07/30/bradesco-santander-e-btg-entram-em-novo-socorro-a-eletricas-diz-fonte.htm?utm_source=twitterfeed</a>	181	Notícia
4	<a href="http://economia.uol.com.br/noticias/redacao/2014/07/31/lucro-do-bradesco-sobe-no-trimestre-com-alta-na-oferta-de-credito.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/redacao/2014/07/31/lucro-do-bradesco-sobe-no-trimestre-com-alta-na-oferta-de-credito.htm?utm_source=twitterfeed</a>	152	Notícia
...	...	...	...
127	<a href="http://economia.uol.com.br/noticias/redacao/2014/07/31/embraer-reverte-prejuizo-e-tem-lucro-de-r-327-milhoes-no-2-trimestre.htm?utm_medium=twitter">http://economia.uol.com.br/noticias/redacao/2014/07/31/embraer-reverte-prejuizo-e-tem-lucro-de-r-327-milhoes-no-2-trimestre.htm?utm_medium=twitter</a>	43	Notícia
128	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58571200">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58571200</a>	43	Notícia
129	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58331542">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58331542</a>	42	Notícia
130	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58329984">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58329984</a>	42	Notícia

Tabela 6.2: Amostra com as primeiras e as últimas quatro URLs da ordenação de repetições. (Coleção completa no Apêndice B)

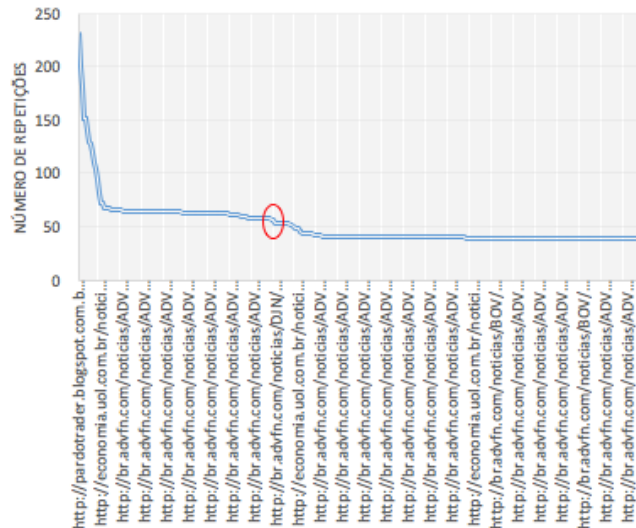


Figura 6.1: Distribuição das URLs em função do seu número total de ocorrências na nossa coleção.

A Tabela 6.3, por sua vez, descreve o resultado da classificação das 130 URLs mais repetidas na nossa coleção. Na tabela é possível perceber a quantidade expressiva de *links* que foram classificados como notícia (81,5%), validando assim os resultados encontrados nos capítulos anteriores e, principalmente, na pesquisa com os usuários em que verificamos que eles estão mais interessados em notícias, ou seja, de fato esse conteúdo está disponível na rede.

Ainda sobre a Tabela 6.3, um resultado interessante dessa classificação é que, apesar das URLs referentes a páginas com conteúdo relativo a Dados Técnicos e Análise Financeira possuírem uma baixa ocorrência, as duas URLs mais comparti-

<i>Classificação da URL</i>	<i>Total</i>	<i>%</i>
Dados Técnicos	2	1,5%
Notícia	106	81,5%
Análise Financeira	3	2,3%
Opinião	1	0,8%
Não acessível	18	13,9%

Tabela 6.3: Classificação das 130 URLs selecionadas

Figura 6.2: Restrição de acesso presente no *site* bastter.com.

lhadas (Tabela 6.2) pertencem a essas duas classes. Ao acessarmos manualmente essas URLs verificamos que um dos motivos para elas terem alta frequência é que elas não correspondem a informação volátil. Por exemplo, no caso da primeira URL o *site* <http://pardotrader.blogspot.com.br> é popular por divulgar análises e dados técnicos e, essa URL, diariamente informa o fechamento da BOVESPA. A segunda, por sua vez, é consequência de um comportamento de segurança implementado pelo *site* <http://www.bastter.com>, popular pela divulgação de análises fundamentalistas. Nesse caso, o usuário somente pode acessar determinado endereço do *site* se estiver cadastrado (Figura 6.2). Vale ressaltar que a elevada frequência dessas duas URLs na rede representa uma alta popularidade dos respectivos *sites* de conteúdo no Twitter. Por fim, por estarmos realizando uma análise com dados coletados há dois anos, nos deparamos com um considerável índice de URLs (13,9%) que não estão mais disponíveis.

Na próxima seção faremos uma análise das URLs com o objetivo de entender quais são as fontes (usuários) responsáveis pelas postagens delas na rede.



## 6.2 Análise das Fontes

Na seção anterior realizamos a classificação de conteúdo das URLs a partir do endereço descompactado e da ordenação delas de acordo com a sua popularidade. Nesta seção, analisaremos a nossa coleção identificando as fontes dos *tweets* (usuários que realizam a primeira postagem da URL na rede) e como eles se propagam. A partir disso, identificamos alguns comportamentos específicos em que o próprio autor postava a mesma URL repetidas vezes na rede.

A partir do conjunto das 130 URLs que mais se repetem em nossa coleção, identificamos quais são as fontes que as publicaram na rede. Na Tabela 6.4 apresentamos uma amostra com as quatro primeiras e quatro últimas URLs desse conjunto. Verificamos que para as duas URLs mais publicadas as fontes são exatamente os usuários proprietários do *site* externo à rede e, também, do perfil no Twitter. Entretanto, a partir da nossa amostra, inferimos que para os *links* que dizem respeito às notícias, há uma diversidade maior de fontes, ou seja, qualquer usuário que tenha acesso antecipado à informação irá compartilhá-la na rede.

<i>Índice</i>	<i>URL</i>	<i>Ocorrência</i>	<i>Fonte (1º Usuário)</i>
1	<a href="http://pardotrader.blogspot.com.br/search/label/fechamento">http://pardotrader.blogspot.com.br/search/label/fechamento</a>	232	pardoTRADER
2	<a href="http://www.bastter.com/mercado/Cadastro.aspx?alerta=login">http://www.bastter.com/mercado/Cadastro.aspx?alerta=login</a>	201	Bastter
3	<a href="http://economia.uol.com.br/noticias/reuters/2014/07/30/bradesco-santander-e-btg-entram-em-novo-socorro-a-eletricas-diz-fonte.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/reuters/2014/07/30/bradesco-santander-e-btg-entram-em-novo-socorro-a-eletricas-diz-fonte.htm?utm_source=twitterfeed</a>	181	forasteiro_mix
4	<a href="http://economia.uol.com.br/noticias/redacao/2014/07/31/lucro-do-bradesco-sobe-no-trimestre-com-alta-na-oferta-de-credito.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/redacao/2014/07/31/lucro-do-bradesco-sobe-no-trimestre-com-alta-na-oferta-de-credito.htm?utm_source=twitterfeed</a>	152	operiodico
...	...	...	...
127	<a href="http://economia.uol.com.br/noticias/redacao/2014/07/31/embraer-reverte-prejuizo-e-tem-lucro-de-r-327-milhoes-no-2-trimestre.htm?utm_medium=twitter">http://economia.uol.com.br/noticias/redacao/2014/07/31/embraer-reverte-prejuizo-e-tem-lucro-de-r-327-milhoes-no-2-trimestre.htm?utm_medium=twitter</a>	43	lspetana10
128	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58571200">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58571200</a>	43	mrcasemiro
129	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58331542">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58331542</a>	42	dragao1313
130	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58329984">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58329984</a>	42	dragao1313

Tabela 6.4: Amostra a identificação das fontes das primeiras e as últimas quatro URLs da ordenação de repetições

Em seguida, aprofundamos nossa análise na busca por comportamento específico em que o próprio autor postava a mesma URL periodicamente. Para isso, buscamos entender a frequência com que cada fonte publicou um mesmo endereço, dentre as 130 mais publicadas da nossa amostra. O resultado está na Tabela 6.5, onde é possível ver de forma mais clara como determinado usuário se esforça para promover o seu próprio *site* na rede. Por exemplo, a URL de maior ocorrência na rede apresenta mais da metade (68%) de suas postagens realizadas pelo mesmo usuário.

Na seção seguinte faremos um estudo final das URLs em que analisaremos o tempo de exposição das URLs classificadas como notícia.

<i>URL</i>	<i>Ocorrência</i>	<i>Fonte</i>	<i>Posts</i>
http://pardotrader.blogspot.com.br/search/label/fechamento	232	pardoTRADER	159
http://www.bastter.com/mercado/Cadastro.aspx?alerta=login	201	Bastter	85
http://lex10at.blogspot.com.br	119	pralex10	73

Tabela 6.5: Mapeamento de fontes com comportamento em que o próprio autor postava a mesma URL repetidas vezes na rede.

### 6.3 Tempo de Exposição das Notícias

Nesta seção analisaremos o tempo de exposição na rede das URLs que foram classificadas como notícias. A principal motivação é que sendo as notícias influenciadoras do comportamento dos ativos na BOVESPA, compreender o tempo de exposição delas na rede social pode representar o tempo pelo qual os ativos na bolsa poderiam sofrer alguma influência.

Para extrair o tempo de exposição calculamos a diferença entre a última e a primeira ocorrência da URL em nossa coleção. A Figura 6.3 mostra a distribuição dos tempos de duração das URLs que foram classificadas como notícias e seus respectivos tempos de exposição. Vale ressaltar que nessa distribuição retiramos o endereço *http://exame.abril.com.br/*, pois entendemos ser ele como não representativo de uma notícia específica, tendo em vista que apesar de ser classificado como notícia, o endereçamento é para um portal de informações, não sendo uma noticia específica. Assim, o tempo de exposição estava consideravelmente maior.

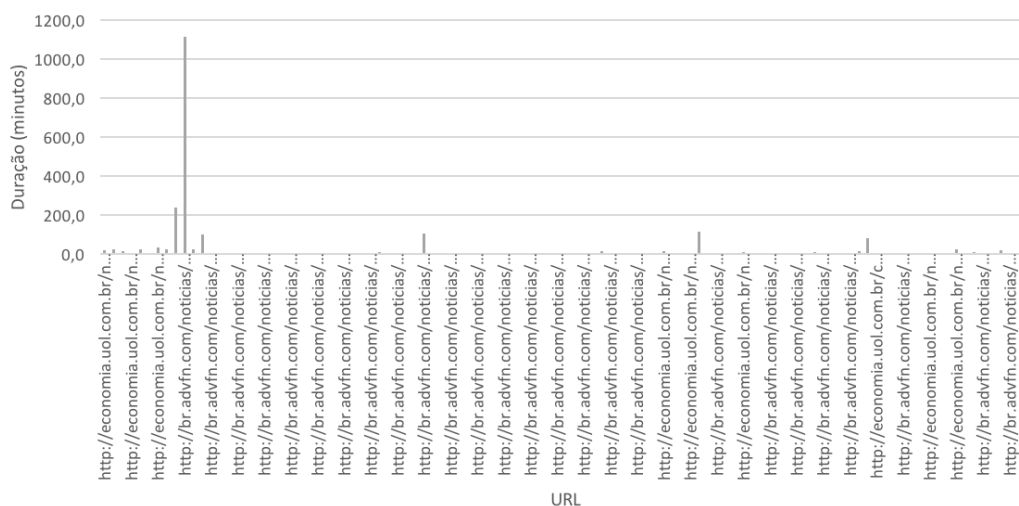


Figura 6.3: Distribuição do tempo de exposição das notícias.

A partir da Figura 6.3, é possível identificar um valor con-

sideravelmente acima da média. Nesse caso, trata-se do endereço <http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58434739> que diz respeito à notícia de investimento da empresa Petrobras. Essa URL teve um tempo de exposição medido na rede de 1.111,7 horas, ou seja, aproximadamente 46 dias. De certa forma esse resultado sugere que por 46 dias essa notícia teve potencial para influenciar o comportamento das ações da empresa na BOVESPA, pois, além de disponível, ela foi compartilhada no Twitter durante esse período.



Figura 6.4: Notícia da empresa Petrobras publicada no *site* ADVFN em 16/07/2013.

Por fim, finalizamos esta seção calculando o intervalo de confiança para o tempo de exposição dos endereços das notícias. O resultado está na Figura 6.5, sendo que o resultado para 95% de confiança foi de no mínimo 0,9 e no máximo 43,4 minutos. Ou seja, para uma dado *tweet* publicado que contenha uma URL que mencione uma notícia do mercado financeiro, existe a probabilidade de 95% dele possuir um tempo de exposição entre 0,9 e 43,4 minutos.

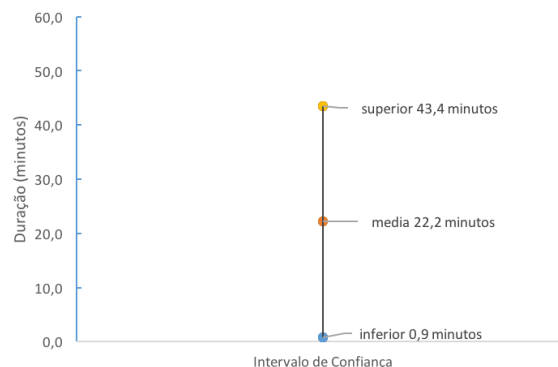


Figura 6.5: Intervalo de confiança para o tempo de exposição dos *tweets* de notícia.

## 6.4 Conclusões

Neste capítulo finalizamos o nosso estudo sobre o mercado brasileiro de ações a partir de dados do Twitter. Nele abordamos a questão de pesquisa Q4 (Seção 2.2) por meio de uma análise de conteúdo. Nessa análise estudamos os *tweets* através da caracterização das suas fontes. Além disso, ao calcular o tempo de exposição na rede dos *tweets* classificados como notícia estamos sugerindo um período pelo qual o ativo mencionado poderia ter seu comportamento na BOVESPA influenciado pela informação compartilhada na rede social.

Na Seção 6.1 encontramos que apenas 18,2% das URLs da nossa coleção são únicas, ou seja, a maior parte dos endereços são compartilhados de forma recorrente na rede. Ainda nessa seção verificamos que 82% delas são notícias (Tabela 6.3). Em seguida, identificamos que as fontes das duas URLs de maior ocorrência em nossa coleção são os usuários proprietários dos *sites* externos promovendo-as na .

Em particular, identificamos dentre as 130 URLs selecionadas que os *links* classificados como notícias apresentam uma diversidade maior de fontes, ou seja, não cabe necessariamente ao perfil proprietário do domínio do *link* postá-lo na rede. Em seguida, realizamos uma análise evidenciando o tempo exposição das URLs na rede que foram classificadas como notícia. A partir disso, identificamos o tempo de exposição está entre 0,9 e 43,4 horas, conforme intervalo de confiança da Figura 6.5. Novamente, esses limites encontrados podem ser utilizados como indicador para tomada de decisão, tendo em vista que representam uma margem de 95% de confiança de que uma nova notícia postada na rede poderá influenciar o ativo dentro do período de tempo descrito por esses valores.

## Capítulo 7

# Conclusões Finais e Trabalhos Futuros

Neste capítulo concluímos o nosso estudo apresentando os principais resultados encontrados de, também, os trabalhos futuros que representam os direcionamentos para novas pesquisas.

### 7.1 Conclusões Finais

Nesta dissertação apresentamos um estudo sobre o mercado de ações brasileiro a partir da análise em larga escala de dados do Twitter coletados de julho de 2013 à julho de 2014 em parceria com o *InWeb*, Instituto Nacional de Ciência e Tecnologia para Web. As análises consideraram os aspectos das postagens dos *tweets* assim como variações dos atributos das ações no BOVESPA. Primeiramente, os dados foram analisados considerando somente os atributos estáticos. Em seguida o aspecto considerado foi o temporal, ou seja, como ocorreram as variações das postagens ao longo dos 13 meses. Outro aspecto importante desta dissertação foi o estudo de correlação entre as oscilações das ações na bolsa de valores e o seu respectivo número de menções nas postagens do Twitter. Por fim, estudamos os conteúdos das URLs publicadas por meio da descompactação desses endereços presentes nos textos dos *tweets*.

A partir da análise estática, descobrimos que apenas 3,1% dos *tweets* postados durante o período considerado foram classificados como relacionados com o mercado de ações, ou seja, esses *tweets* mencionaram explicitamente pelo menos um código de ações da BOVESPA em seu texto. Além disso, também encontramos um grupo especializado correspondente a 0,5 % dos usuários que está realmente interessado no mercado de ações. A partir desse resultado, diferenciamos este estudo de trabalhos anteriores ao

entrevistar diretamente 3.302 usuários do Twitter e, em seguida, conseguimos mapear interesses e atributos desses usuários como, por exemplo, que eles em sua maioria são investidores individuais interessados em acompanhar notícias e análises financeiras.

A análise temporal, por sua vez, mostrou padrões de comportamento do usuário a partir da análise dos dados ao longo de todo o período da coleção. Um resultado interessante para essa etapa foi que nossa análise mostrou que os eventos e notícias sobre o mercado de ações são capazes de gerar picos de publicações no Twitter. Além disso, descobrimos que a frequência de publicações no Twitter está concentrada principalmente nos horários do pregão de negociação. Entretanto, influenciada pela abertura de bolsas externas e resultados operacionais publicados, a frequência de postagens no Twitter se mantém mesmo após o fechamento do pregão de negociação da BOVESPA. Ao final do capítulo encontramos que 10% dos usuários podem atuar como influenciadores na rede, pois são responsáveis por mais de 90% do conteúdo publicado.

Em particular, na análise de correlação foi traçado um paralelo entre o comportamento diário das ações na BOVESPA e a frequência de menção aos ativos no Twitter. Inicialmente, foi calculado o coeficiente de correlação de Spearman para as ações do IBOVESPA e as séries temporais correspondentes às menções dos ativos no Twitter. Nessa análise, observamos que o número total de ordens e o volume financeiro são positivamente correlacionados com 66% das ações mencionadas no Twitter, enquanto que a oscilação diária e a oscilação diária máxima não demonstram claramente um padrão de correlação. Como resultado final, mapeamos as 22 ações que podem ter nas suas respectivas frequências de menções no Twitter indicadores para monitorar seus comportamentos na BOVESPA. Ou seja, demonstramos que partir da análise dos dados do serviço de *microblogging* é possível estruturar indicadores que ofereçam ao investidor maiores informações para o seu processo de decisão, principalmente quando a análise financeira realizada considerar como relevante o volume financeiro e o número total de ordens do ativo.

Por fim, culminamos a construção do estudo que é o objetivo dessa dissertação através da análise de conteúdo. Essa análise produziu resultados interessantes que enquadram nosso estudo como o primeiro a classificar as informações compartilhadas no Twitter, além de compreender suas fontes e o tempo de exposição. Nessa etapa conseguimos encontrar 95,2% dos endereços completos e verificamos que apenas 18,2% deles são únicos, ou seja, a maior parte das URLs são compartilhadas de forma recorrente na rede. Ainda nessa análise, selecionamos os 130 endereços que mais foram compartilhados pelos usuários e verificamos que 82% deles são notícias (Tabela 6.3). Finalizamos o capítulo aprofundando nossa análise ao identificarmos o tempo de exposição entre 0,9 e 43,4 horas, conforme intervalo de confiança da Figura 6.5. Esse resultado sugere

que uma nova notícia postada na rede poderá influenciar o ativo mencionado dentro do período de tempo descrito por esses valores limites.

Em geral, as análises desenvolvidas foram relevantes por evidenciarem padrões comportamentais dos usuários no Twitter sobre os quais ainda pouco se sabia. O trabalho cumpriu com o objetivo principal de fornecer informações relevantes sobre um cenário carente de pesquisas, pois existem poucos estudos computacionais orientados às redes sociais *online* que além de utilizar o Twitter como fonte, são aplicáveis à Bolsa de Valores de São Paulo (BOVESPA) e consideram a língua portuguesa como idioma base para extração e análise textual dos dados. O estudo e a caracterização do comportamento de usuários realizada é base para identificar quais tipos de informação são tidas como relevantes para o mercado de ações brasileiro. Apesar dos procedimentos de caracterização terem sido executados sobre um conjunto de dados específico, os padrões encontrados possuem aplicabilidade geral, assim como os procedimentos e a metodologia. Os padrões e resultados encontrados possuem aplicações práticas diversas, tais como a elaboração de indicadores de suporte à decisão de investimento e mecanismos automatizados de operação em bolsa que considerem dados mensurados a partir do coletivo expresso nas redes sociais, em particular, o Twitter no Brasil.

## 7.2 Trabalhos Futuros

Nesta dissertação, foi apresentado um estudo sobre o mercado brasileiro de ações a partir de dados do Twitter composta por análises estáticas, temporais, correlacionais e de conteúdo. Apesar da abrangência de nossas análises, existem muitos outros aspectos promissores a serem investigados no contexto de intercessão entre o Twitter e outras redes sociais e a BOVESPA.

Assim, como o trabalho futuro, pretendemos explorar um modelo de grafo para explicar o comportamento do nosso conjunto de usuários do Twitter com base em conceitos como centralidade, *clustering* e comunidades. Além disso, pretendemos expandir este trabalho através da realização de uma análise de sentimento em nosso conjunto de dados do Twitter, a fim de investigar a forma como a opinião e humor dos usuários se correlaciona com as oscilações de ações no mercado brasileiro.

Por fim, além de explorar outras possibilidades de análise, pretendemos investir tempo no desenvolvimento uma ferramenta que seja composta por indicadores de investimento que considerem os resultados obtidos nesta dissertação. Por exemplo, nessa ferramenta podemos ter um *dashboard* de notícias indicando o período de publicação e o seu tempo de exposição esperada na rede, além de um indicador de compartilhamento

de *tweets* de análises financeiras para sugerir ao investidor se alguma decisão financeira poderá ser tomada considerando o conteúdo publicado.



# Referências Bibliográficas

- Benevenuto, F.; Magno, G.; Rodrigues, T. & Almeida, V. (2010). Detecting Spammers on Twitter. In *Proceedings of the 7th Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference*, volume 6, p. 12.
- Benevenuto, F.; Rodrigues, T.; Cha, M. & Almeida, V. (2009). Characterizing User Behavior in Online Social Networks. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference*, pp. 49 – 62. ACM.
- Bollen, J.; Mao, H. & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1 – 8.
- Cha, M.; Benevenuto, F.; Haddadi, H. & Gummadi, P. K. (2012). The World of Connections and Information Flow in Twitter. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 42(4):991–998.
- Cheong, M. & Lee, V. (2009). Integrating Web-Based Intelligence Retrieval and Decision-Making from the Twitter Trends Knowledge Base. In *Proceedings of the 2nd ACM Workshop on Social Web Search and Mining*, pp. 1 – 8.
- Cohen, J.; Cohen, P.; West, S. G. & Aiken, L. S. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Routledge.
- Earle, P.; Bowden, D. & Guy, M. (2012). Twitter earthquake detection: earthquake monitoring in a social world. *Annals of Geophysics*, 54(6).
- Fama, E. F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business*, 38(1):34 – 105.
- Fogel, J. & Nehmad, E. (2009). Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in human behavior*, 25(1):153 – 160.

- Gilbert, E. & Karahalios, K. (2010). Widespread Worry and the Stock Market. In *Proceedings of the International Conference on Weblogs and Social Media*, pp. 59 – 65.
- Gomide, J.; Veloso, A.; Meira, Jr., W.; Almeida, V.; Benevenuto, F.; Ferraz, F. & Teixeira, M. (2011). Dengue surveillance based on a computational model of spatio-temporal locality of twitter. In *Proceedings of the 3rd International Web Science Conference*, WebSci '11, pp. 3:1 – 3:8, New York, NY, USA. ACM.
- Kwak, H.; Lee, C.; Park, H. & Moon, S. (2010). What is Twitter, a Social Network or a News Media? In *Proceedings of the 19th International Conference on The World Wide Web*, pp. 591 – 600.
- Maia, M.; Almeida, J. & Almeida, V. (2008). Identifying User Behavior in Online Social Networks. In *Proceedings of the 1st Workshop on Social Network Systems*, pp. 1 – 6.
- Mislove, A.; Marcon, M.; Gummadi, K. P.; Druschel, P. & Bhattacharjee, B. (2007). Measurement and Analysis of Online Social Networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, pp. 29 – 42.
- Nofer, D.-K. M.; Hinz, O.; Muntermann, J. & Roßnagel, H. (2014). The economic impact of privacy violations and security breaches. *Business & Information Systems Engineering*, 6(6):339 – 348.
- Nofer, M. (2015a). Market anomalies on two-sided auction platforms. In *The Value of Social Media for Predicting Stock Returns*, pp. 11 – 26. Springer.
- Nofer, M. (2015b). *The Value of Social Media for Predicting Stock Returns - Preconditions, Instruments and Performance Analysis*. Springer.
- Nofer, M. & Hinz, O. (2014). Are crowds on the internet wiser than experts? The case of a stock prediction community. *Journal of Business Economics*, 84(3):303 – 338.
- Nofer, M. & Hinz, O. (2015). Using Twitter to Predict the Stock Market. *Business & Information Systems Engineering*, 57(4):229 – 242.
- Parameswaran, A.; Garcia-Molina, H. & Rajaraman, A. (2010). Towards the Web of Concepts: Extracting Concepts From Large Datasets. *Proceedings of the VLDB Endowment*, 3(1-2):566 – 577.

- Poh, H.-L.; Yao, J. & Jašic, T. (1998). Neural networks for the analysis and forecasting of advertising and promotion impact. *International Journal of Intelligent Systems in Accounting, Finance & Management*, 7(4):253 – 268.
- Ruiz, E. J.; Hristidis, V.; Castillo, C.; Gionis, A. & Jaimes, A. (2012). Correlating Financial Time Series with Micro-blogging Activity. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, pp. 513 – 522.
- Santos, H. S.; Laender, A. H. & Pereira, A. C. (2015). A Twitter View of the Brazilian Stock Exchange Market. In *E-Commerce and Web Technologies*, pp. 112 – 123. Springer.
- Sharma, N. K.; Ghosh, S.; Benevenuto, F.; Ganguly, N. & Gummadi, K. (2012). Inferring Who-is-Who in the Twitter Social Network. *ACM SIGCOMM Computer Communication Review*, 42(4):533 – 538.
- Sprenger, T. O. (2011). Tweettrader.net: Leveraging Crowd Wisdom in a Stock Microblogging Forum. In *Proceedings of the International Conference on Weblogs and Social*.
- Sprenger, T. O.; Tumasjan, A.; Sandner, P. G. & Welpe, I. M. (2014). Tweets and trades: the information content of stock microblogs. *European Financial Management*, 20(5):926 – 957.
- Zar, J. H. (1972). Significance testing of the spearman rank correlation coefficient. *Journal of the American Statistical Association*, 67(339):578 – 580.
- Zhang, W. & Skiena, S. (2010). Trading strategies to exploit blog and news sentiment. In *Proceedings of the Fourth International Conference on Weblogs and Social Media*, pp. 375 – 378.



# Apêndice A

## IBOVESPA e Ações Consideradas

O IBOVESPA é o resultado de uma carteira teórica de ativos, elaborada de acordo com os critérios estabelecidos em sua metodologia, sendo que seu objetivo é ser o indicador do desempenho médio das cotações dos ativos mais negociados e de maior representatividade do mercado de ações brasileiro<sup>1</sup>. As 73 ações que fizeram parte do índice durante a elaboração desta dissertação estão descritas na Tabela A.1, assim como os códigos BOVA11 e IBOV11 que fazem referência ao próprio índice:

Tabela A.1: Códigos que referenciam ativos do IBOVESPA ou seu próprio índice.

<b>Código</b>	<b>Ação</b>
ALLL3	América Latina Logística
AMBV4 ou ABEV3	Ambev S/A
BBAS3	Banco do Brasil
BBDC4	Bradesco
BISA3	Brookfield
BRAP4	Bradespar
BRFS3	BRF SA
BRKM5	Braskem
BRML3	BR Malls Par
BTOW3	B2W Digital
BVMF3	BMFBOVESPA
CCRO3	CCR SA
CESP6	Cesp

<sup>1</sup><http://www.bmfbovespa.com.br/indices/ResumoIndice.aspx?Indice=IBOVESPA&Idioma=pt-br>

<b>Código</b>	<b>Ação</b>
CIEL3	Cielo
CMIG4	Cemig
CPFE3	CPFL Energia
CPLE6	Copel
CRUZ3	Souza Cruz
CSAN3	Cosan
CSNA3	SID Nacional
CTIP3	Cetip
CYRE3	Cyrela Realt
DASA3	Diagnosticos da America SA
DTEX3	Duratex
ELET3	Eletrobras
ELET6	Eletrobras
ELPL4	AES Eletropaulo
EMBR3	Embraer
ENBR3	Energias do Brasil
FIBR3	Fibria
GFSA3	Gafisa
GGBR4	Gerdau
GOAU4	Gerdau Met
GOLL4	Gol
HGTX3	CIA Hering
HYPE3	Hypermarcas
ITSA4	Itausa
ITUB4	Itaunibanco
JBSS3	JBS
KLBN4	Klabin S/A
LAME4	Lojas Americanas
LIGT3	Light S.A.

---

<b>Código</b>	<b>Ação</b>
LLXL3	LLX
LREN3	Lojas Renner
MMXM3	MMX
MRFG3	Marfrig
MRVE3	MRV
NATU3	Natura
OGXP3	OGX
OIBR3	Oi
OIBR4	Oi
PCAR4	P. Açúcar-CBD
PDGR3	PDG Realty
PETR3	Petrobras
PETR4	Petrobras
RENT3	Localiza
RSID3	Rossi Residencial
SANB11	Santander
SBSP3	Sabesp
SUZB5	Suzano
TIMP3	Tim
TRPL4	Transmissão Paulista
UGPA3	Ultrapar
USIM3	Usiminas
USIM5	Usiminas
VAGR3	Ecodiesel
VALE3	Vale
VALE5	Vale
VIVT4	Vivo
BBDC3	Bradesco
BRPR3	Br Properties
KROT3	Kroton
AEDU3	Anhanguera
IBOV11 ou BOVA11	Índice IBOVESPA





## Apêndice B

# Lista das 130 URLs de Maior Ocorrência

Tabela B.1: 130 URLs de maior ocorrência e suas classes

Índice	URL	Ocorrência	Classe
1	<a href="http://pardotrader.blogspot.com.br/search/label/fechamento">http://pardotrader.blogspot.com.br/search/label/fechamento</a>	232	Dados Técnicos
2	<a href="http://www.bastter.com/mercado/Cadastro.aspx?alerta=login">http://www.bastter.com/mercado/Cadastro.aspx?alerta=login</a>	201	Análise Financeira
3	<a href="http://economia.uol.com.br/noticias/reuters/2014/07/30/bradesco-santander-e-btg-entram-em-novo-socorro-a-eletricas-diz-fonte.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/reuters/2014/07/30/bradesco-santander-e-btg-entram-em-novo-socorro-a-eletricas-diz-fonte.htm?utm_source=twitterfeed</a>	181	Notícia
4	<a href="http://economia.uol.com.br/noticias/redacao/2014/07/31/lucro-do-bradesco-sobe-no-trimestre-com-alta-na-oferta-de-credito.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/redacao/2014/07/31/lucro-do-bradesco-sobe-no-trimestre-com-alta-na-oferta-de-credito.htm?utm_source=twitterfeed</a>	152	Notícia
5	<a href="http://invista-em-acoes.blogspot.com.br/">http://invista-em-acoes.blogspot.com.br/</a>	151	Opinião
6	<a href="http://economia.uol.com.br/noticias/redacao/2014/07/31/embraer-reverte-prejuizo-e-tem-lucro-de-r-327-milhoes-no-2-trimestre.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/redacao/2014/07/31/embraer-reverte-prejuizo-e-tem-lucro-de-r-327-milhoes-no-2-trimestre.htm?utm_source=twitterfeed</a>	130	Notícia
7	<a href="http://economia.uol.com.br/noticias/redacao/2014/07/31/lucro-do-santander-cresce-5-credito-sobe-pouco-e-inadimplencia-tem-alta.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/redacao/2014/07/31/lucro-do-santander-cresce-5-credito-sobe-pouco-e-inadimplencia-tem-alta.htm?utm_source=twitterfeed</a>	128	Notícia
8	<a href="http://lex10at.blogspot.com.br">http://lex10at.blogspot.com.br</a>	119	Análise Financeira
9	<a href="http://economia.uol.com.br/noticias/redacao/2014/07/31/com-efeito-copa-lucro-da-ambev-sobe-16-no-trimestre-a-r-22-bilhoes.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/redacao/2014/07/31/com-efeito-copa-lucro-da-ambev-sobe-16-no-trimestre-a-r-22-bilhoes.htm?utm_source=twitterfeed</a>	110	Notícia
10	<a href="http://economia.uol.com.br/noticias/redacao/2014/04/24/usiminas-sai-de-prejuizo-e-lucra-r-222-milhoes-no-1-trimestre.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/redacao/2014/04/24/usiminas-sai-de-prejuizo-e-lucra-r-222-milhoes-no-1-trimestre.htm?utm_source=twitterfeed</a>	104	Notícia
11	<a href="http://pregao-online.bmfbovespa.com.br/Principal.aspx?idioma=pt-BR">http://pregao-online.bmfbovespa.com.br/Principal.aspx?idioma=pt-BR</a>	95	Dados Técnicos
12	<a href="http://economia.uol.com.br/noticias/redacao/2014/07/24/gpa-avalia-trocar-algumas-unidades-do-extra-por-pao-de-acucar.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/redacao/2014/07/24/gpa-avalia-trocar-algumas-unidades-do-extra-por-pao-de-acucar.htm?utm_source=twitterfeed</a>	81	Notícia
13	<a href="http://economia.uol.com.br/noticias/redacao/2013/08/21/usiminas-cobra-empresa-de-eike-batista-por-atraso-em-obra-de-porto.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/redacao/2013/08/21/usiminas-cobra-empresa-de-eike-batista-por-atraso-em-obra-de-porto.htm?utm_source=twitterfeed</a>	73	Notícia
14	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58687782">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58687782</a>	72	Notícia
15	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58434739">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58434739</a>	68	Notícia
16	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58536051">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58536051</a>	67	Notícia
17	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59562104">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59562104</a>	67	Notícia
18	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58361343">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58361343</a>	66	Notícia
19	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58369939">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58369939</a>	66	Notícia
20	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58416833">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58416833</a>	66	Notícia

Índice	URL	Ocorrência	Classe
21	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58704722">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58704722</a>	66	Notícia
22	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58716694">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58716694</a>	66	Notícia
23	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58776255">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58776255</a>	66	Notícia
24	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58968776">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58968776</a>	66	Notícia
25	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59765462&amp;adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59765462&amp;adw=1126416</a>	65	Não abriu
26	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59712862&amp;adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59712862&amp;adw=1126416</a>	65	Não abriu
27	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59755152&amp;adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59755152&amp;adw=1126416</a>	65	Não abriu
28	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59746788&amp;adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59746788&amp;adw=1126416</a>	65	Não abriu
29	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59755325&amp;adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59755325&amp;adw=1126416</a>	65	Não abriu
30	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59807298&amp;adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59807298&amp;adw=1126416</a>	65	Não abriu
31	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59803352&amp;adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59803352&amp;adw=1126416</a>	65	Não abriu
32	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59730518&amp;adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59730518&amp;adw=1126416</a>	65	Não abriu
33	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59529709">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59529709</a>	65	Notícia
34	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59530261">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59530261</a>	65	Notícia
35	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59512666">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59512666</a>	65	Notícia
36	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59597844">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59597844</a>	65	Notícia
37	<a href="http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62907121?adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62907121?adw=1126416</a>	65	Notícia
38	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59226381">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59226381</a>	65	Notícia
39	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59661424">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59661424</a>	65	Notícia
40	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59490757">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59490757</a>	65	Notícia
41	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58847561">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58847561</a>	65	Notícia
42	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58279579">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58279579</a>	64	Notícia
43	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59268025">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59268025</a>	64	Notícia
44	<a href="http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62865181?adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62865181?adw=1126416</a>	64	Notícia
45	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/60124638&amp;adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/60124638&amp;adw=1126416</a>	64	Não abriu
46	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/60209519&amp;adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/60209519&amp;adw=1126416</a>	64	Não abriu
47	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/60069592&amp;adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/60069592&amp;adw=1126416</a>	64	Não abriu
48	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59879743&amp;adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59879743&amp;adw=1126416</a>	64	Não abriu
49	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/60118210&amp;adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/60118210&amp;adw=1126416</a>	64	Não abriu
50	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/60165931&amp;adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/60165931&amp;adw=1126416</a>	64	Não abriu
51	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58867209">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58867209</a>	64	Notícia
52	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58757306">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58757306</a>	64	Notícia
53	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59280135">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59280135</a>	64	Notícia
54	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59414906">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59414906</a>	64	Notícia
55	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59392192">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59392192</a>	64	Notícia
56	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59241957">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59241957</a>	64	Notícia
57	<a href="http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62143722?adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62143722?adw=1126416</a>	63	Notícia
58	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/60009455&amp;adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/60009455&amp;adw=1126416</a>	63	Não abriu
59	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59987609&amp;adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59987609&amp;adw=1126416</a>	63	Não abriu
60	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59863048&amp;adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/59863048&amp;adw=1126416</a>	63	Não abriu
61	<a href="http://daltonvieira.com/">http://daltonvieira.com/</a>	63	Análise Financeira
62	<a href="http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62199507?adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62199507?adw=1126416</a>	63	Notícia
63	<a href="http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62759938?adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62759938?adw=1126416</a>	63	Notícia
64	<a href="http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62821663?adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62821663?adw=1126416</a>	63	Notícia
65	<a href="http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62659631?adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62659631?adw=1126416</a>	63	Notícia
66	<a href="http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62179981?adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62179981?adw=1126416</a>	63	Notícia
67	<a href="http://br.advfn.com/noticias/ADVNEWS/2014/artigo/61766003?adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2014/artigo/61766003?adw=1126416</a>	63	Notícia
68	<a href="http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62693445?adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62693445?adw=1126416</a>	63	Notícia
69	<a href="http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62162707?adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62162707?adw=1126416</a>	63	Notícia
70	<a href="http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62725369?adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2014/artigo/62725369?adw=1126416</a>	63	Notícia
71	<a href="http://br.advfn.com/noticias/ADVNEWS/2014/artigo/61672820?adw=1126416">http://br.advfn.com/noticias/ADVNEWS/2014/artigo/61672820?adw=1126416</a>	63	Notícia

Índice	URL	Ocorrência	Classe
72	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/61834484?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/61834484?adw=1126416</a>	63	Notícia
73	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/61842147?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/61842147?adw=1126416</a>	63	Notícia
74	<a href="http://br.advn.com/noticias/ADVNEWS/2013/artigo/60000269&amp;adw=1126416">http://br.advn.com/noticias/ADVNEWS/2013/artigo/60000269&amp;adw=1126416</a>	63	Notícia
75	<a href="http://br.advn.com/noticias/ADVNEWS/2013/artigo/59069169">http://br.advn.com/noticias/ADVNEWS/2013/artigo/59069169</a>	63	Notícia
76	<a href="http://br.advn.com/noticias/DJN/2014/artigo/62079709?adw=1126416">http://br.advn.com/noticias/DJN/2014/artigo/62079709?adw=1126416</a>	62	Notícia
77	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/61866438?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/61866438?adw=1126416</a>	62	Notícia
78	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/62409931?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/62409931?adw=1126416</a>	62	Notícia
79	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/62375325?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/62375325?adw=1126416</a>	62	Notícia
80	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/61710162?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/61710162?adw=1126416</a>	62	Notícia
81	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/61541824?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/61541824?adw=1126416</a>	62	Notícia
82	<a href="http://economia.uol.com.br/noticias/redacao/2013/11/12/lucro-do-bb.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/redacao/2013/11/12/lucro-do-bb.htm?utm_source=twitterfeed</a>	62	Notícia
83	<a href="http://economia.uol.com.br/noticias/redacao/2013/09/16/llx-de-eike-assina-acordo-definitivo-para-investimento-de-r-13-bilhao.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/redacao/2013/09/16/llx-de-eike-assina-acordo-definitivo-para-investimento-de-r-13-bilhao.htm?utm_source=twitterfeed</a>	61	Notícia
84	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/61491103?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/61491103?adw=1126416</a>	61	Notícia
85	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/61264809?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/61264809?adw=1126416</a>	61	Notícia
86	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/60950187?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/60950187?adw=1126416</a>	61	Notícia
87	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/61419282?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/61419282?adw=1126416</a>	61	Notícia
88	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/61507627?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/61507627?adw=1126416</a>	61	Notícia
89	<a href="http://br.advn.com/noticias/ADVNEWS/2013/artigo/60279524?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2013/artigo/60279524?adw=1126416</a>	60	Notícia
90	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/60638891?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/60638891?adw=1126416</a>	59	Notícia
91	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/60929842?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/60929842?adw=1126416</a>	59	Notícia
92	<a href="http://economia.uol.com.br/noticias/redacao/2013/09/23/grupo-pao-de-acucar-escolhe-jean-charles-naouri-para-presidente-do-conselho.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/redacao/2013/09/23/grupo-pao-de-acucar-escolhe-jean-charles-naouri-para-presidente-do-conselho.htm?utm_source=twitterfeed</a>	59	Notícia
93	<a href="http://economia.uol.com.br/noticias/reuters/2014/07/30/bradesco-santander-e-btg-entram-em-novo-socorro-a-eletricas-diz-fonte.htm?utm_medium=twitter">http://economia.uol.com.br/noticias/reuters/2014/07/30/bradesco-santander-e-btg-entram-em-novo-socorro-a-eletricas-diz-fonte.htm?utm_medium=twitter</a>	58	Notícia
94	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/60607457?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/60607457?adw=1126416</a>	58	Notícia
95	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/60588925?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/60588925?adw=1126416</a>	58	Notícia
96	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/60670861?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/60670861?adw=1126416</a>	58	Notícia
97	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/60687450?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/60687450?adw=1126416</a>	58	Notícia
98	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/60619487?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/60619487?adw=1126416</a>	58	Notícia
99	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/60531907?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/60531907?adw=1126416</a>	58	Notícia
100	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/60531908?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/60531908?adw=1126416</a>	58	Notícia
101	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/60529650?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/60529650?adw=1126416</a>	58	Notícia
102	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/60703601?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/60703601?adw=1126416</a>	58	Notícia
103	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/60767735?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/60767735?adw=1126416</a>	58	Notícia
104	<a href="http://br.advn.com/noticias/ADVNEWS/2013/artigo/60378503?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2013/artigo/60378503?adw=1126416</a>	58	Notícia
105	<a href="http://br.advn.com/noticias/ADVNEWS/2013/artigo/60325403?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2013/artigo/60325403?adw=1126416</a>	58	Notícia
106	<a href="http://br.advn.com/noticias/ADVNEWS/2013/artigo/58666542">http://br.advn.com/noticias/ADVNEWS/2013/artigo/58666542</a>	56	Notícia
107	<a href="http://br.advn.com/noticias/ADVNEWS/2013/artigo/60443934?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2013/artigo/60443934?adw=1126416</a>	56	Notícia
108	<a href="http://br.advn.com/noticias/ADVNEWS/2013/artigo/59539531">http://br.advn.com/noticias/ADVNEWS/2013/artigo/59539531</a>	54	Notícia
109	<a href="http://economia.uol.com.br/noticias/redacao/2013/11/11/ogx-procura-dinheiro-novo-ate-janeiro-diz-jornal.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/redacao/2013/11/11/ogx-procura-dinheiro-novo-ate-janeiro-diz-jornal.htm?utm_source=twitterfeed</a>	54	Notícia
110	<a href="http://economia.uol.com.br/cotacoes/noticias/redacao/2013/09/09/troca-de-acoes-entre-abilio-e-casino-ocorrera-em-duas-partes.htm?utm_source=twitterfeed">http://economia.uol.com.br/cotacoes/noticias/redacao/2013/09/09/troca-de-acoes-entre-abilio-e-casino-ocorrera-em-duas-partes.htm?utm_source=twitterfeed</a>	54	Notícia
111	<a href="http://br.advn.com/noticias/DJN/2014/artigo/62014815?adw=1126416">http://br.advn.com/noticias/DJN/2014/artigo/62014815?adw=1126416</a>	54	Notícia
112	<a href="http://exame.abril.com.br/">http://exame.abril.com.br/</a>	54	Notícia
113	<a href="http://br.advn.com/noticias/DJN/2013/artigo/58731848">http://br.advn.com/noticias/DJN/2013/artigo/58731848</a>	53	Notícia
114	<a href="http://br.advn.com/noticias/ADVNEWS/2013/artigo/59531664">http://br.advn.com/noticias/ADVNEWS/2013/artigo/59531664</a>	53	Notícia
115	<a href="http://br.advn.com/noticias/ADVNEWS/2013/artigo/58960299">http://br.advn.com/noticias/ADVNEWS/2013/artigo/58960299</a>	53	Notícia
116	<a href="http://br.advn.com/noticias/ADVNEWS/2014/artigo/62461817?adw=1126416">http://br.advn.com/noticias/ADVNEWS/2014/artigo/62461817?adw=1126416</a>	51	Notícia

Índice	URL	Ocorrência	Classe
117	<a href="http://economia.uol.com.br/noticias/redacao/2013/08/01/lucro-da-gerdau-cai-27-no-2-trimestre.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/redacao/2013/08/01/lucro-da-gerdau-cai-27-no-2-trimestre.htm?utm_source=twitterfeed</a>	51	Notícia
118	<a href="http://economia.uol.com.br/noticias/redacao/2013/08/06/petrobras-diz-ter-dinheiro-suficiente-para-atuar-em-leilao-de-pre-sal.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/redacao/2013/08/06/petrobras-diz-ter-dinheiro-suficiente-para-atuar-em-leilao-de-pre-sal.htm?utm_source=twitterfeed</a>	50	Notícia
119	<a href="http://economia.uol.com.br/cotacoes/noticias/redacao/2013/07/22/acao-do-bradesco-cai-apos-corte-nas-projecoes-de-crescimento-em-2013.htm?utm_source=twitterfeed">http://economia.uol.com.br/cotacoes/noticias/redacao/2013/07/22/acao-do-bradesco-cai-apos-corte-nas-projecoes-de-crescimento-em-2013.htm?utm_source=twitterfeed</a>	49	Notícia
120	<a href="http://economia.uol.com.br/noticias/redacao/2014/07/31/santander-confirma-parceria-de-r-460-milhoes-com-banco-bonsucesso.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/redacao/2014/07/31/santander-confirma-parceria-de-r-460-milhoes-com-banco-bonsucesso.htm?utm_source=twitterfeed</a>	49	Notícia
121	<a href="http://economia.uol.com.br/noticias/redacao/2014/07/31/lucro-do-bradesco-sobe-no-trimestre-com-alta-na-oferta-de-credito.htm?utm_medium=twitter">http://economia.uol.com.br/noticias/redacao/2014/07/31/lucro-do-bradesco-sobe-no-trimestre-com-alta-na-oferta-de-credito.htm?utm_medium=twitter</a>	48	Notícia
122	<a href="http://br.advfn.com/noticias/DJN/2013/artigo/60238942&amp;adw=1126416">http://br.advfn.com/noticias/DJN/2013/artigo/60238942&amp;adw=1126416</a>	45	Não abriu
123	<a href="http://economia.uol.com.br/cotacoes/noticias/redacao/2013/07/22/acao-do-bradesco-passa-a-subir-apos-avaliacao-de-balanco.htm?utm_source=twitterfeed">http://economia.uol.com.br/cotacoes/noticias/redacao/2013/07/22/acao-do-bradesco-passa-a-subir-apos-avaliacao-de-balanco.htm?utm_source=twitterfeed</a>	44	Notícia
124	<a href="http://economia.uol.com.br/noticias/redacao/2014/07/31/lucro-do-santander-cresce-5-credito-sobe-pouco-e-inadimplencia-tem-alta.htm?utm_medium=twitter">http://economia.uol.com.br/noticias/redacao/2014/07/31/lucro-do-santander-cresce-5-credito-sobe-pouco-e-inadimplencia-tem-alta.htm?utm_medium=twitter</a>	44	Notícia
125	<a href="http://economia.uol.com.br/noticias/redacao/2013/09/10/diretoria-da-ogx-confirma-que-pode-obrigar-eike-a-injetar-us-1-bi.htm?utm_source=twitterfeed">http://economia.uol.com.br/noticias/redacao/2013/09/10/diretoria-da-ogx-confirma-que-pode-obrigar-eike-a-injetar-us-1-bi.htm?utm_source=twitterfeed</a>	44	Notícia
126	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58584358">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58584358</a>	44	Notícia
127	<a href="http://economia.uol.com.br/noticias/redacao/2014/07/31/embraer-reverte-prejuizo-e-tem-lucro-de-r-327-milhoes-no-2-trimestre.htm?utm_medium=twitter">http://economia.uol.com.br/noticias/redacao/2014/07/31/embraer-reverte-prejuizo-e-tem-lucro-de-r-327-milhoes-no-2-trimestre.htm?utm_medium=twitter</a>	43	Notícia
128	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58571200">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58571200</a>	43	Notícia
129	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58329984">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58329984</a>	42	Notícia
130	<a href="http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58331542">http://br.advfn.com/noticias/ADVNEWS/2013/artigo/58331542</a>	42	Notícia