

Uma Análise do Mercado de Ações Baseada na Correlação entre Ativos no StockTwits

Gabriela B. Alves, João Paulo S. R. Bastos,
Michele A. Brandão, Adriano C. M. Pereira

Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG – Brasil

{gabrielabrant, joaopaulosr, micheleabrandao, adrianoc}@dcc.ufmg.br

***Abstract.** StockTwits is a social microblog for the financial and investing community which is becoming very popular. In this paper, we investigate why companies are mentioned together in StockTwits. Furthermore, we analyze how this information can be used to help on decision making in the stock market. In particular, we propose a new graph-based model in which stocks are nodes and edges are formed when stocks are mentioned together in the same post. Then, we analyze the main features and isolated pairs in the network. The results show that stocks cited together are correlated with financial results on stock market.*

***Resumo.** O StockTwits é um microblog social cada vez mais popular e voltado para o público interessado em mercado financeiro. Neste trabalho, o porquê de empresas distintas serem citadas juntas no StockTwits é investigado. Além disso, analisamos como utilizar essa informação para auxiliar na tomada de decisão no mercado de ações. Especificamente, é proposta uma modelagem em grafo na qual os ativos são os nós e as arestas se formam quando os ativos são citados juntos em uma postagem. Em seguida, é feita uma análise das principais características e dos pares isolados da rede. Os resultados mostram que ativos citados juntos estão correlacionados com as variações no mercado financeiro.*

1. Introdução

As redes sociais são cada vez mais importantes e não param de surgir variações delas para as mais diversas finalidades (e.g., interação social, troca de fotos, compartilhamento de música, etc.). Entre as mais conhecidas e exploradas por pesquisas acadêmicas está o Twitter [Ciotec et al. 2014, Santos et al. 2015], que com seu formato dinâmico e postagens de no máximo 140 caracteres, atrai muitos usuários a compartilharem opiniões e informações. Na mesma linha de fluxo rápido de informações e postagens curtas, o StockTwits¹ é uma rede social criada para a disseminação de conteúdo sobre finanças. Diferente de outras redes sociais, o StockTwits possui como público-alvo pessoas interessadas no mercado financeiro, dando possibilidade aos usuários amadores de interagirem livremente com profissionais.

Até o momento, o mercado de ações associado a redes sociais ainda é pouco explorado pela área de Ciência da Computação [Sprenger et al. 2014]. Atualmente, é possível encontrar pesquisas que utilizam análise de sentimentos para prever movimentações no

¹StockTwits: <https://stocktwits.com/>

mercado de ações [Oh and Sheng 2011]. Outros se envolvem em esforços para encontrar os usuários mais experientes e que podem oferecer twits² de maior relevância. Esses twits podem ser utilizados como entrada para a modelagem de aprendizado de máquina objetivando prever movimentos no mercado de ações [Bar-Haim et al. 2011].

Diferente dos trabalhos citados, o objetivo principal aqui é analisar a correlação entre ativos, os quais identificam empresas que estão na bolsa de valores e são representados através de \$TICKERS. Por exemplo, \$GOOG e \$APPL representam os ativos da Google e da Apple, respectivamente. Inicialmente, a rede de ativos é modelada como um grafo em que os nós são os ativos e as arestas se formam quando os ativos são citados em um mesmo twit. Assim, são considerados apenas os twits que possuem dois ou mais ativos citados. Em seguida, são realizadas análises topológicas para entender as características da rede e análise dos pares isolados (componentes desconectados) para estudar a interação entre pares de ativos. Note que os pares isolados são considerados por não se repetirem com a mesma frequência que os pares pertencentes ao componente conectado.

O estudo da correlação entre ativos aqui apresentado é inédito, portanto, este trabalho mostra o valor dessa linha de análise para auxiliar à tomada de decisão no mercado financeiro. Além disso, é investigado como a quantidade de citações está relacionada com a variação no preço das ações dos pares mais citados. Também é brevemente analisado o impacto que notícias podem causar na variação dos resultados financeiros dos ativos. Especificamente, duas perguntas principais de pesquisa motivam este trabalho: **P1:** Quais as características de uma rede de ativos? e **P2:** Por que ativos são citados juntos?

O restante deste artigo está organizado da seguinte forma: a Seção 2 discute os trabalhos relacionados. A Seção 3 descreve os materiais e métodos utilizados para desenvolvimento da pesquisa. Em seguida, a Seção 4 discorre sobre os resultados encontrados e suas análises. Finalmente, a Seção 5 apresenta as principais conclusões.

2. Trabalhos Relacionados

Existem trabalhos que exploram conteúdo de redes sociais e revelam conclusões promissoras sobre a importância de analisar esses tipos de dados para prever o mercado de ações. Grande parte dos trabalhos realizam análise de sentimentos em sua metodologia. Por exemplo, Oliveira et al. (2013) utilizam indicadores de sentimento associados a um modelo de regressão para prever três variáveis do mercado: retorno, volatilidade e volume negociado. Tais autores chegam a conclusão de que prever o andamento do mercado é uma tarefa muito complexa e que são necessários modelos muito bem embasados. Estudos mais recentes, como o realizado por Li et al. (2017), consideram métodos mais específicos de análise de sentimentos, através de indicadores léxicos associados à redes neurais, e obtém resultados melhores na predição de sentimento das postagens (twits).

Ademais, Atkins et al. (2018) estudaram como as notícias impactam na volatilidade do mercado de ações. Nessa pesquisa, utilizaram métodos de aprendizado de máquina aplicado a notícias de jornais e concluíram que a predição do mercado de ações apresenta maior acurácia quando notícias midiáticas são consideradas nos modelos. Similarmente, neste trabalho, são utilizadas fontes de notícias que envolvem pares de ações selecionados, buscando entender a correlação entre eles.

²*Twits*: termo utilizado no StockTwits que funciona de forma similar aos *tweets* do Twitter.



Figura 1. Principais etapas para construir e investigar a rede de ativos.

Existem linhas de pesquisa que investigam a influência dos usuários em relação ao mercado financeiro. Por exemplo, Wang et al. (2014) utilizam dados do StockTwits e do SeekingAlpha³ para determinar quem são os usuários especialistas e se suas postagens têm utilidade para investidores individuais. Similarmente, Tu et al. (2016) investigam os especialistas no StockTwits por meio de técnicas de aprendizado de máquina e utilizam as postagens de usuários *experts* para gerar recomendações de portfólio.

Nesse contexto, a modelagem em grafos aqui proposta é pouco explorada, sendo que foi encontrada apenas uma pesquisa que modela os dados do StockTwits em grafo. Diferentemente de nossa abordagem, Cortez et al. (2016) modelam a interação entre usuários em grafo, considerando os *retweets*, compartilhamentos e *replies*. Assim, o objetivo é medir a influência dos usuários no StockTwits, partindo do pressuposto que usuários mais influentes são mais úteis para criar modelos de previsão no mercado financeiro.

Portanto, a principal contribuição deste trabalho é um estudo da correlação entre ativos no StockTwits através de uma modelagem em grafos em que os ativos são os nós. O objetivo é identificar as características (e.g., nível de agrupamento, conexão entre vizinhos) da rede e entender o motivo dos ativos serem citados juntos. Além disso, a análise dos pares considerando notícias fornece indícios do porquê deles serem citados juntos repetidas vezes. Assim, essas formas de análise têm potencial para contribuir para a tomada de decisão no mercado financeiro.

3. Materiais e Métodos

Esta seção apresenta os materiais e métodos utilizados para investigar a rede de ativos. Uma visão geral das etapas para realização deste trabalho é apresentada na Figura 1 com suas cinco principais etapas: (1) construção da rede de ativos na qual os nós são ações e a aresta se forma quando um par é citado no mesmo twit; (2) cálculo das métricas topológicas para cada mês com o objetivo de analisar a variação das características durante o ano; (3) remoção dos índices da rede; (4) seleção dos vinte pares mais citados de cada mês; e (5) análise de um par isolado para cada trimestre do ano para entender como a variação no número de citações se relaciona com os dados financeiros daquelas ações.

Descrição do conjunto de dados. Os dados utilizados foram gentilmente cedidos pela plataforma StockTwits, portanto, não houve fase de coleta ou tratamento dos dados. A base dos dados contém 22.434.953 twits postados de 1º de janeiro a 31 de dezembro de 2015, sendo que desses, 37,2% citam algum tipo de ativo. Visto que o principal objetivo é entender o motivo de ativos serem citados juntos, são considerados apenas twits com dois ou mais ativos. No total, foram processados um montante de 991.772 twits.

³Seeking Alpha: <https://seekingalpha.com/>



Figura 2. Modelo da rede de ativos

Tabela 1. Propriedades topológicas

Métrica	Definição	Interpretação
Coeficiente de Clusterização (CC)	<p>Sendo $T(u)$ o número total de triângulos aos quais u pertence e $deg(u)$ o grau (número de arestas conectadas) de u, então CC é:</p> $CC(u) = \frac{2T(u)}{deg(u)(deg(u)-1)}$	Representa a tendência de um nó formar uma comunidade, ou seja, a tendência dos anos em agruparem-se.
Neighborhood Overlap (NO)	<p>Sendo $\mathcal{N}(u)$ e $\mathcal{N}(v)$ o conjunto de nós u e v vizinhos, respectivamente, então $NO(u, v) = \frac{ \mathcal{N}(u) \cap \mathcal{N}(v) }{ \mathcal{N}(u) \cup \mathcal{N}(v) }$</p>	Mede a similaridade entre dois pares de nós vizinhos e pode gerar bons insumos a respeito de como os ativos interagem entre si.
Preferential Attachment (PA)	<p>Dado $\mathcal{N}(u)$ e \mathcal{N} como conjuntos de vizinhos de u e v, respectivamente, $PA(u, v) = \frac{ \mathcal{N}(u) }{ \mathcal{N}(v) }$</p>	Assume o potencial de receber novas arestas através do grau do nó, sendo que quanto maior o número de vizinhos, maior o valor do preferential attachment.
Coeficiente Adamic-Adar (AA)	<p>No contexto de rede social, as características são os vizinhos em comum e a métrica é formalizada da seguinte forma: $AA(u, v) = \frac{\sum_{z \in \mathcal{N}(u) \cap \mathcal{N}(v) }}{\log \mathcal{N}(z) }$</p>	Dado um conjunto de características, essa métrica foi originalmente criada para computar a similaridade entre duas páginas Web.

Construção da rede social de ativos. A rede foi modelada como um grafo ponderado $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, sendo \mathcal{V} o conjunto de nós (representando os ativos), e \mathcal{E} o conjunto de arestas não direcionadas (formadas quando um par de ativos é citado no mesmo twit). O peso das arestas é o total de vezes que um par de ativos é citado junto. A Figura 2 mostra um exemplo da formação da rede de ativos, baseado em dois twits aleatórios. Os pares de ativos ($\$TSLA, \$GOOG$) e ($\$FB, \$GOOG$) se conectam, pois foram citados juntos na mesma postagem. O peso é um valor fictício utilizado apenas para exemplificar.

Para entender o impacto da quantidade de vezes em que um par de ativos é citado junto, as análises são realizadas considerando os pares de ativos mensalmente. Dessa forma, a escolha mensal permite analisar dados mais específicos e melhor correlacionar com o mercado e suas variações.

Análise topológica. Para responder à pergunta de pesquisa **P1**, foram calculadas métricas topológicas para cada mês do ano de 2015. As métricas topológicas caracterizam a estrutura da rede, considerando como os nós e arestas interagem entre si. Em particular, o objetivo é comparar as variações destas características ao longo dos meses. Em nosso estudo, as métricas de análise de redes complexas utilizadas foram coeficiente de clusterização, *neighborhood overlap*, *preferential attachment* e Adamic Adar, calculadas na forma de média e detalhadas na Tabela 1.

Análise dos pares isolados. Além da análise topológica, é necessária uma análise mais aprofundada para entender como os pares de ativos citados juntos podem ser utilizados para auxiliar na tomada de decisão no mercado de ações. Assim, são feitas análises em pares de ativos com o objetivo de responder a pergunta **P2** deste trabalho.

Nessa etapa, consideramos apenas os vinte pares de ativos mais citados de cada mês (rede TOP20), pois o objetivo é fazer uma análise detalhada. Em média, o par mais citado possui aproximadamente 1098 citações e o vigésimo par mais citado 364 citações. A partir dos grafos TOP20 de cada mês, questiona-se o porquê de alguns pares aparecerem isolados do componente principal. Em geral, os componentes conectados principais são formados pelos mesmos ativos, como AAPL, FB, TWTR, GOOGL, TSLA, NFLX, etc. Portanto, os pares isolados são mais interessantes dado o seu comportamento atípico, em que aparecem uma vez ou poucas vezes entre os vinte pares mais citados nos meses. Os índices, *Exchange Traded Funds* (ETFs)⁴, não são considerados nas análises de pares isolados, visto que os índices representam diversos ativos e o escopo do trabalho é entender porque ativos são citados juntos.

Na rede TOP20 de cada mês, existem diversos pares isolados, portanto, é necessário uma abordagem para selecionar os pares a serem estudados, descrita a seguir. Inicialmente, o ano de 2015 foi dividido em trimestres, uma vez que três meses é o intervalo em que as empresas liberam seus balanços financeiros. Em seguida, analisamos os dados de cada trimestre e selecionamos o par isolado com maior número de citações. Após selecionado o par mais citado de cada trimestre, o gráfico de preço anual de cada par é comparado com a quantidade de citações mês a mês. Após a análise anual, os ativos são estudados em uma granularidade de tempo menor, ou seja, o mês em que o par mais citado do trimestre aparece é analisado em termos diários. No gráfico diário, são considerados apenas dias úteis, pois é quando o mercado de ações está aberto. Finalmente, a variação dos preços e números de citações é estudada utilizando notícias midiáticas com o objetivo de apresentar explicações para a correlação entre ativos.

4. Resultados

Esta seção apresenta resultados da análise topológica (Seção 4.1) para responder a pergunta de pesquisa P1 e da análise de pares isolados (Seção 4.2) para responder P2.

4.1. Análise Topológica

Para investigar as características da rede de ativos, as métricas coeficiente de clusterização, *neighborhood overlap*, *preferential attachment* e *Adamic Adar* são consideradas e apresentadas na Figura 3. Os valores das métricas são referentes à rede completa de cada mês, ou seja, consideram todos os ativos que apareceram naquele mês, incluindo índices. Note que essa análise não é feita para a rede TOP20, pois o relevante aqui é entender as características da rede como um todo e sua variação ao longo do ano.

A Figura 3(a) mostra que as características são homogêneas durante o ano. O coeficiente de clusterização se manteve estável ao longo de 2015, variando apenas entre 0,4 e 0,5. Isso significa que os pares de ativos tendem em média a formar triângulos [David and Jon 2010]. Em outras palavras, a citação de dois pares de ativos tende a ocasionar a citação conjunta de um terceiro par com ativos já envolvidos em citações anteriores. Para mercado financeiro, isso pode indicar, por exemplo, que ativos de mesmo setor tendem a se influenciar mutuamente. Os resultados para *neighborhood overlap* são constantes ao longo do ano, se mantendo em torno de 0,1. Em geral, valores próximos de

⁴Um índice de ação é uma carteira teórica (i.e., não existe de fato, é apenas calculada) criada com objetivo de que os investidores tenham um indicador de comparação de um conjunto de ações.

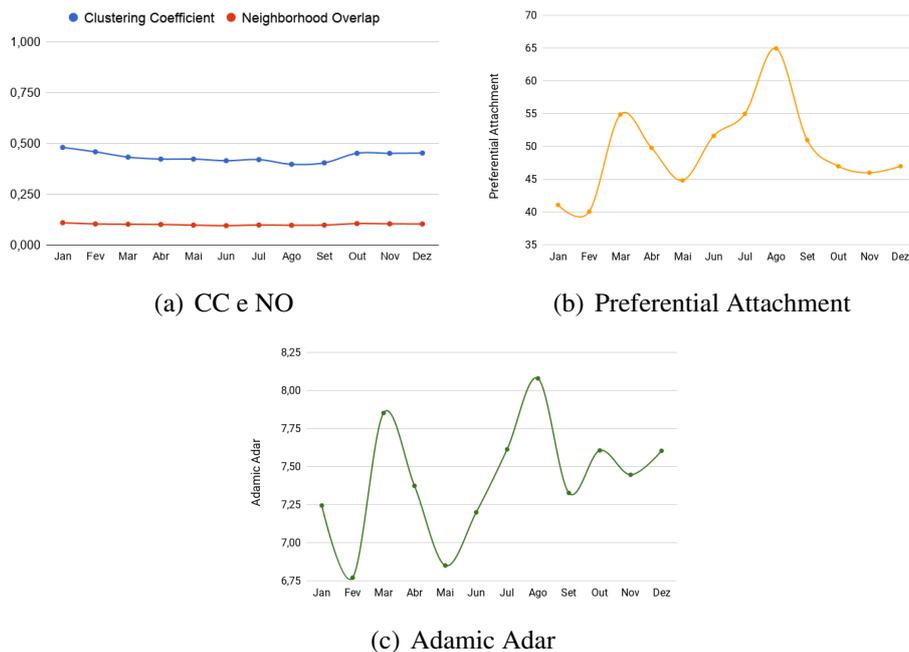


Figura 3. Valor médio das métricas topológicas para o grafo completo de citações de ativos de cada mês.

0 indicam que existem muitas pontes locais na rede e aqui representa que existem diversos ativos formando uma ponte entre componentes conectados [David and Jon 2010].

Por outro lado, a Figura 3(b) mostra picos de variação em alguns meses, como março e agosto que apresentam valores elevados comparados a seus adjacentes. *Preferential attachment* refere-se a ideia de que há uma tendência de novos nós se conectarem a nós que possuem alto grau, causando o efeito rico fica mais rico. Assim, considerando o contexto financeiro, um alto PA pode indicar que existem ativos que se destacam na rede em meses específicos. Em relação à Figura 3(c), existe uma variação grande entre os meses. A métrica *Adamic Adar* calcula a similaridade entre dois nós e a intuição por trás da métrica é que características raras são mais importantes do que as triviais. Dessa forma, pares com alto valor de AA tendem a ser novidade na rede e não são citados aleatoriamente ou influenciados por ativos sempre citados. A análise da Figura 3(c) mostra que nos meses de fevereiro e maio não houveram pares de ativos muito raros, enquanto que em março e agosto observa-se a maior presença de novos pares. Por exemplo, nesses quatro meses, os ativos AAPL e SPY são sempre citados em conjunto com outros pares. Assim, um ativo citado em conjunto com esses pode não indicar alguma mudança no mercado financeiro, enquanto que o surgimento de outra citação como BNI-UNP pode trazer mais informação relevante.

4.2. Análise dos Pares Isolados

Esta seção apresenta a análise dos pares isolados da rede dos pares TOP20 de ativos mais citados a cada trimestre conforme os grafos apresentados nas Figuras 4, 6, 8 e 10. Observa-se que em todos os meses existe um componente conectado predominante e outros isolados. Neste artigo, são realizadas análises sobre pares os isolados, como explicado na Seção 3.

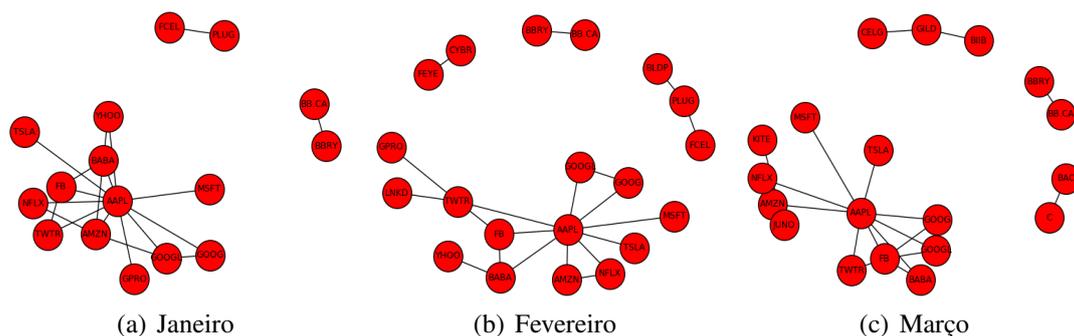


Figura 4. Primeiro trimestre - Rede de ativos TOP20



Figura 5. CYBR (laranja), FEYE (azul) - Variação de preço em contraste com número de citações

Primeiro trimestre. Começando a análise pelo primeiro trimestre de 2015, vemos na Figura 4 que existem diversos componentes nos meses de janeiro, fevereiro e março. No entanto, consideramos elegíveis apenas os componentes isolados e entre eles, analisaremos o mais citado. O par mais citado ocorre em fevereiro com 683 citações, sendo ele a aresta CYBR-FEYE. O fato desse par não aparecer nos outros meses nos grafos de TOP20 dá indícios de que houve algum acontecimento relevante naquele mês, levando essas duas empresas a serem citadas juntas muitas vezes.

Ambas empresas são do setor de Tecnologia da Informação, sendo que a CyberArk Software Ltd, representada pelo *ticker* CYBR, é uma empresa Israelense que oferece soluções em segurança da informação voltadas para proteção contra ataques cibernéticos. Em comparação, o *ticker* FEYE representa a FireEye Inc. que oferece soluções baseadas em inteligência artificial, também voltada para segurança cibernética.

A Figura 5(a) mostra o gráfico da variação do preço das duas ações justapostos no período de janeiro a dezembro de 2015. A linha laranja refere-se à ação CYBR e a linha azul à ação FEYE. A linha verde na parte inferior do gráfico mostra a quantidade de citações por mês desse par de ativos. É possível observar que o gráfico das duas ações em questão tem picos e comportamentos incomuns nos meses em que os ativos são mais citados. Além disso, analisando especificamente o mês de fevereiro de 2015, vemos na Figura 5(b) a variação diária no preço e número de citações do par CYBR-FEYE. Novamente, o número de citações é representada pela linha verde no canto inferior do gráfico. Vemos que no dia 12 de fevereiro de 2015 ocorreu a inversão das curvas das ações, colocando a FEYE acima da CYBR. Ao mesmo tempo, o número de citações cresceu até alcançar o

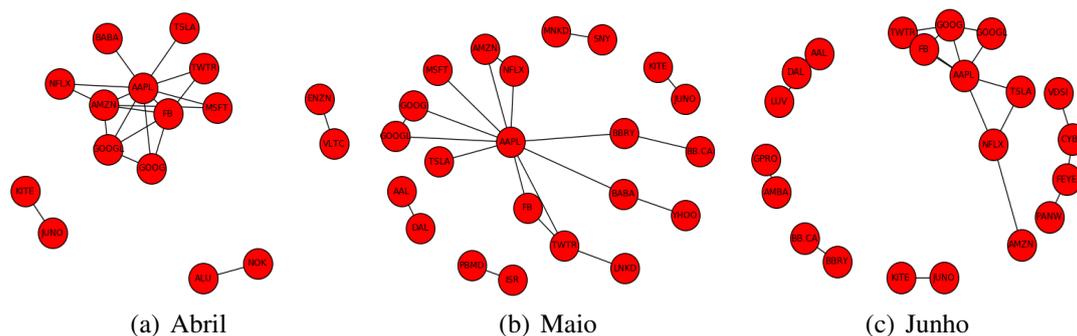


Figura 6. Segundo trimestre - Rede de ativos TOP20.

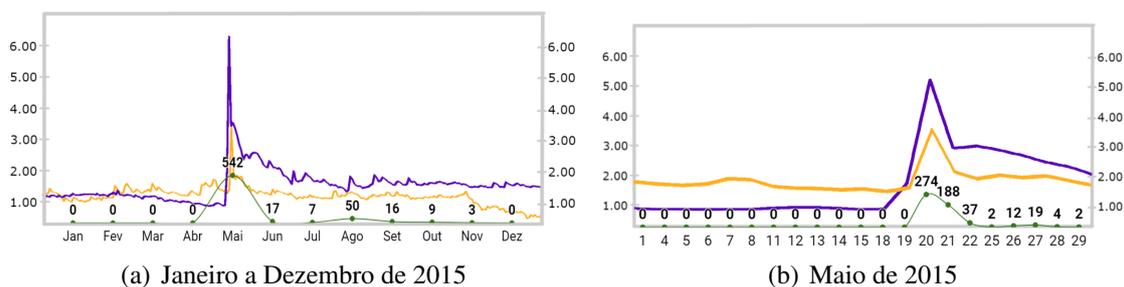


Figura 7. ISR (laranja), PBMD (azul) - Variação de preço em contraste com número de citações

pico no dia 13 de fevereiro. Essa subida abrupta no número de citações e a inversão de posições na linha de preços entre CYBR e FEYE se deu pela divulgação de resultados financeiros da CYBR no mês de fevereiro⁵. Em geral, os resultados superaram as expectativas de analistas de mercado e fizeram com que as ações da empresa tivessem seu valor elevado em aproximadamente 15%, sendo essa mesma subida nos preços apontada como uma possível bolha por analistas na semana seguinte.⁶

Segundo trimestre. No segundo trimestre, também vemos alguns componentes isolados na Figura 6 e o par ISR-PBMD é escolhido entre eles, uma vez que possui a maior quantidade de citações daqueles três meses (542 citações). O par em questão foi o segundo mais citado de maio, ficando atrás apenas de APPL-FB que possui 572 citações. Essa posição reforça a relevância dessa dupla no segundo trimestre.

A empresa IsoRay Inc. é do ramo de biotecnologia. Através de sua subsidiária IsoRay Medical Inc., a empresa desenvolve, manufatura e vende medicamentos a base de isótopos e aparelhos para o tratamento de câncer e outras doenças malignas. Por outro lado, a Prima BioMed Ltd, também voltada para biotecnologia, tem foco em pesquisa, desenvolvimento e comercialização de produtos licenciados no segmento imunoterapêutico.

Ao analisar o gráfico de preço anual das ações na Figura 7(a), o pico de preço da Prima BioMed chama atenção. Tal comportamento atípico está relacionado ao número de citações do par ISR-PBMD em maio, que foi o mais alto do ano inteiro. No mesmo mês, a

⁵The Street: <http://bit.ly/subida-preco-CYBR> - Acesso em 12/2017.

⁶SeekingAlpha: <http://bit.ly/possivel-bolha-CYBR> - Acesso em 12/2017.

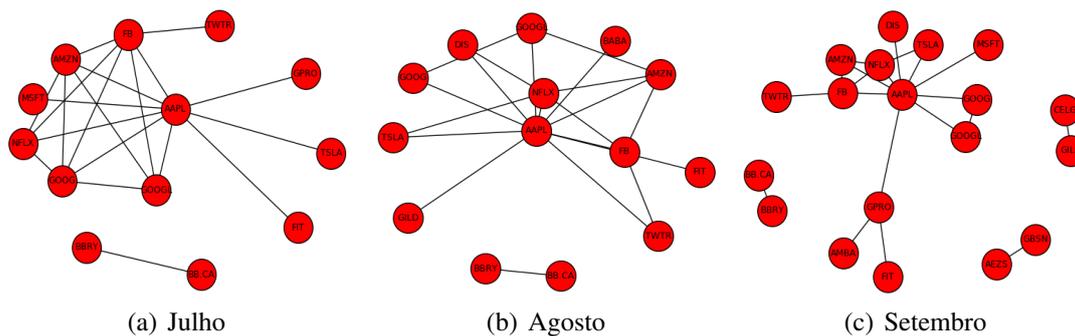


Figura 8. Terceiro trimestre - Rede de ativos TOP20.

IsoRay Inc. também tem um pico, porém menos significativo que a Prima BioMed. Além do gráfico de preços durante todo o ano, temos também o gráfico que mostra a variação de preço e do número de citações durante o mês de maio, exibidos na Figura 7(b). Em uma reportagem de 20 de maio de 2015, explica-se a subida de 270% no preço das ações da Prima Biomed após o anúncio de resultados promissores nos testes de uma nova droga para o tratamento de câncer de ovário⁷. Na Figura 7(b), vemos que esta é exatamente a data na qual o par de ações tem maior número de citações no mês de maio. Além disso, é interessante ver que o número de citações era baixo e constante nos dias anteriores à essa notícia, tendo um aumento abrupto com a novidade. Por fim, considerando-se que ambas empresas são da mesma área, a descoberta de novos tratamentos em uma delas, gera expectativa de novidades na principal concorrente da área.

Terceiro trimestre. Neste trimestre, o par de citações mais relevante foi CELG-GILD, que aparece no mês de setembro. Essa é a amostra menos significativa entre os trimestres, pois como vemos na Figura 8 os grafos de julho, agosto e setembro são altamente conectados. Assim, temos apenas dois candidatos à análise aprofundada, sendo eles CELG-GILD (365 citações) e AEZS-GBSN (346 citações). O par BBRY-BB.CA representa a mesma empresa, Blackberry Inc.⁸, portanto não é um par elegível para a análise.

Por ser o par mais citado do terceiro trimestre, vamos estudar as ações CELG-GILD. A empresa Celgene Corporation representada pelo *ticker* CELG, é do ramo biofarmacêutico e em conjunto com suas subsidiárias, a companhia tem como objetivo a descoberta, desenvolvimento e comercialização de terapias para o tratamento de câncer e doenças inflamatórias. Em comparação, o *ticker* GILD representa a empresa Gilead Sciences Inc., também da área biofarmacêutica voltada para pesquisa e que desenvolve e comercializa medicamentos nas áreas de necessidades médicas ainda sem cura, como Vírus da Imunodeficiência Humana/Síndrome da Imunodeficiência Adquirida (HIV/AIDS), doença hepática, câncer, doenças respiratórias e cardiovasculares.

Em 2015, o par selecionado CELG-GILD não tem uma variação muito expressiva na quantidade de citações por mês, como pode ser visto na Figura 9(a). Além disso, o mês de setembro não é o que possui maior número de citações como nos outros pares que estudamos. Ainda assim, é possível observar comportamentos interessantes no gráfico de

⁷Sacks: <http://bit.ly/pbmd-decolou-270-porcento> - Acesso em 01/2018

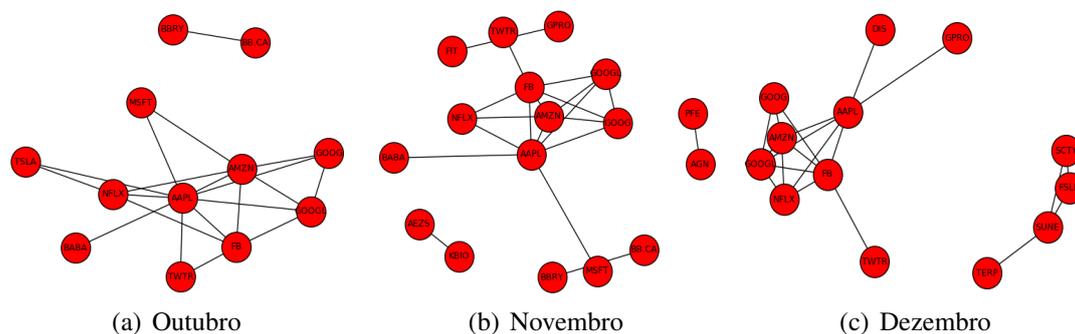
⁸BB.CA:TSE (Toronto Stock Exchange); BBRY:NYSE (New York Stock Exchange)



(a) Janeiro a Dezembro de 2015

(b) Setembro de 2015

Figura 9. CELG (laranja), GILD (azul) - Variação de preço em contraste com número de citações



(a) Outubro

(b) Novembro

(c) Dezembro

Figura 10. Quarto trimestre - Rede de ativos TOP20.

ações das empresas CELG e GILD, já que o mês de setembro apresentou a queda mais expressiva no preço das duas ações. Esse é um resultado diferente daqueles vistos no primeiro e segundo trimestre, pois está associado à diminuição do valor da ação. Em detalhes, conseguimos observar na Figura 9(b) a variação do preço e número de citações do mês de setembro. No dia 21 de setembro de 2015 houve um pico no número de citações e queda abrupta no valor das ações do par CELG-GILD. Esse foi exatamente o dia em que Hillary Clinton postou um *tweet* prometendo tomar alguma atitude em relação aos altíssimos preços praticados pela indústria farmacêutica⁹. Sendo assim, podemos observar que houve grande impacto do comentário nas ações CELG e GILD, que são da indústria farmacêutica

Quarto trimestre. No último trimestre, o par de ações com o maior número de citações é o SCTY-SUNE, que representam as empresas SolarCity Corporation e SunEdison Inc., respectivamente. Esse par teve 938 citações no mês de dezembro e ambas são empresas do setor energético norte-americano. Entre as análises realizadas nos trimestres anteriores, o mês de dezembro é o que tem quantidade mais expressiva em número de citações.

Sobre as ações aqui analisadas, apresentamos a SolarCity Corporation que oferece serviços relacionados à energia limpa. Entre eles, energia solar, eficiência energética e desenho de veículos elétricos, monitoramento e manutenção em ambiente residencial, escolar e governamental nos Estados Unidos. Paralelamente, a SunEdison Inc. opera em dois segmentos: materiais semicondutores e energia solar. A companhia produz e vende

⁹CNN Money: <http://bit.ly/tweet-hillary-clinton> - Acesso em 03/2018

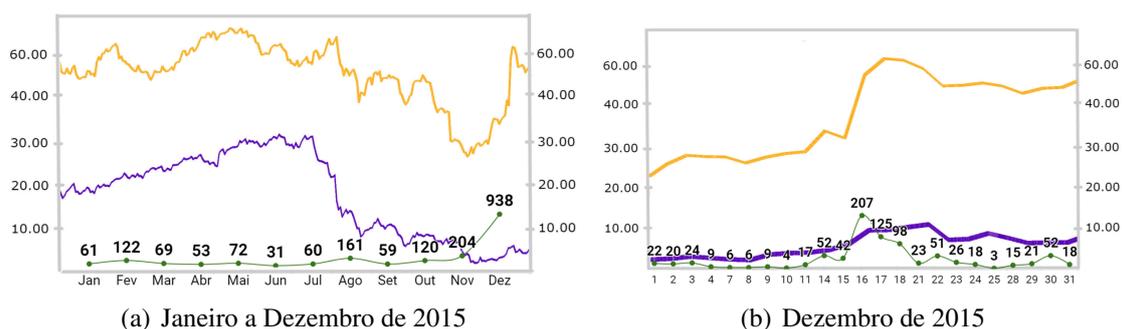


Figura 11. SCTY (laranja), SUNE (azul) - Variação de preço em contraste com número de citações

wafers e produtos relacionados à semicondutores e à indústria de energia solar.

As notícias publicadas no mês de dezembro tiveram impacto claro na variação de preço e citações do par de SCTY-SUNE no mês de dezembro. Primeiro, ocorreu, no início de dezembro, a Conferência do Clima Paris 2015 (COP21) que reuniu quase 200 governos mundiais os quais concordaram em limitar a emissão de gases estufa e investir no desenvolvimento de energias renováveis.¹⁰ Em seguida, no dia 15 de dezembro os rumores sobre a possível aprovação de incentivos fiscais para indústrias do setor energético se espalha¹¹ e no dia 18 de Dezembro as expectativas se concretizam. O governo norte-americano aprova o incentivo milionário¹². Essa sequência de notícias comparadas à Figura 11(b) mostra que a expectativa de mercado fez o preço das ações aumentarem. Em seguida, a quantidade de citações estabilizou novamente, após a sequência de novidades.

Após analisar todos os trimestres de 2015, aprofundando em quatro pares de ações, é possível responder à pergunta de pesquisa P2. Os ativos são citados juntos quando são do mesmo setor, ou seja, concorrentes. Em todos os exemplos, essa é uma constante e mostra que os usuários do StockTwits estão atentos às empresas que competem entre si.

5. Conclusão

Este trabalho apresentou indícios de que pares de ativos podem ser utilizados no auxílio à tomada de decisão no mercado financeiro. Até o momento, não identificamos pesquisas nesse sentido, com foco na análise da correlação entre ativos citados juntos através de uma modelagem em grafos. Os resultados mostraram que os pares de ativos mais citados refletem o comportamento de mercado dos mesmos, em relação a preços e notícias financeiras. Em resposta a pergunta P1, as características da rede de ativos revelaram que há uma tendência na formação de triângulos e pontes na citação de ativos, bem como mostraram a presença de ativos que são citados com muita frequência. Em geral, tais ativos não trazem muita informação nova para análise de mercado. Sobre a questão P2, observou-se que o principal motivo para ativos serem citados juntos repetidas vezes em uma mesma postagem é a divulgação de notícias que impactam as ações em questão. Em todos os casos, as ações são impactadas por mudanças políticas ou econômicas. Com essa informação, conclui-se que ações que atuam no mesmo setor são as mais citadas juntas,

¹⁰Yahoo Finance: <http://bit.ly/porque-alta-setor-energetico> - Acesso 02/2018

¹¹Forbes: <http://bit.ly/investimentos-industria-energetica> - Acesso em 03/2018

¹²Forbes: <http://bit.ly/novidades-congresso-US> - Acesso em 03/2018

mostrando que os investidores no StockTwits estão atentos à competição entre empresas.

Finalmente, ainda há muito a ser explorado e existem diversos trabalhos futuros. Por exemplo, gerar uma nuvem de palavras para cada par de ativos mais citados e entender o que está sendo dito. Também analisar nós que aparecem e saem do componente conectado a cada mês, não apenas os pares isolados. Outra possibilidade é estudar a criação e recomendação de portfólios de investimentos através da análise dos pares de ações.

Agradecimentos. Este trabalho foi parcialmente financiado pelo Instituto Nacional de Ciência e Tecnologia para a Web (grant no. 573871/2008-6), MASWeb (grant FAPEMIG/PRONEX APQ-01400-14), CAPES, CNPq e Fapemig.

Referências

- Atkins, A., Niranjana, M., and Gerding, E. (2018). Financial news predicts stock market volatility better than close price. *The Journal of Finance and Data Science*.
- Bar-Haim, R., Dinur, E., Feldman, R., Fresko, M., and Goldstein, G. (2011). Identifying and following expert investors in stock microblogs. In *Procs. of EMNLP*, pages 1310–1319, Edinburgh, United Kingdom.
- Ciotec, S., Dascalu, M., and Trausan-Matu, S. (2014). A comprehensive study of twitter social networks. In *RoEduNet Conference 13th Edition*, pages 1–7. IEEE.
- Cortez, P., Oliveira, N., and Ferreira, J. a. P. (2016). Measuring user influence in financial microblogs: Experiments using stocktwits data. In *Procs. of WIMS*, pages 23:1–23:10, Nîmes, France.
- David, E. and Foray, K. (2010). *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, New York, USA.
- Li, Q. and Shah, S. (2017). Learning stock market sentiment lexicon and sentiment-oriented word vector from stocktwits. In *CoNLL*, pages 301–310, Vancouver, Canada.
- Oh, C. and Sheng, O. (2011). Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. In *Procs. of ICIS*, page 57–58, Shanghai, China.
- Oliveira, N., Cortez, P., and Areal, N. (2013). On the predictability of stock market behavior using stocktwits sentiment and posting volume. volume 854, page 355–365, Berlin, Germany.
- Santos, H. S., Laender, A. H., and Pereira, A. C. (2015). Uma visão do mercado brasileiro de ações a partir de dados do twitter. In *BraSNAM*, Recife, Brazil.
- Sprenger, T. O., Tumasjan, A., Sandner, P. G., and Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, 20(5):926–957.
- Tu, W., Cheung, D. W., Mamoulis, N., Yang, M., and Lu, Z. (2016). Investment recommendation using investor opinions in social media. In *SIGIR*, pages 881–884, Pisa, Italy.
- Wang, G., Wang, T., Wang, B., Sambasivan, D., Zhang, Z., Zheng, H., and Zhao, B. Y. (2015). Crowds on wall street: Extracting value from collaborative investing platforms. In *CSCW*, pages 17–30, Vancouver, Canada.