

# Uma Revisão Sistemática acerca das Técnicas de Identificação e Análise de Tendências

Caio C. Trucolo<sup>1</sup>, Luciano A. Digiampietri<sup>1</sup>

<sup>1</sup>Escola de Artes Ciências e Humanidades – Universidade de São Paulo (USP)  
Caixa Postal 03828-000 – São Paulo – SP – Brasil

{trucolo, digiampietri}@usp.br

***Abstract.** This paper consists of developing a systematic review about methods and techniques used for identifying trends in different application types. The first goal is to understand what are the main techniques being used and the way that these techniques are organized by the researchers. The second goal consists of understanding which of these techniques could be applied to social networks.*

***Resumo.** Este artigo consiste no desenvolvimento de uma revisão sistemática acerca dos métodos e técnicas utilizados para se identificar tendências em diferentes tipos de aplicação. O primeiro objetivo é entender quais são as principais técnicas utilizadas e a forma com que essas técnicas são organizadas pelos pesquisadores. O segundo objetivo consiste em entender quais dessas técnicas poderiam ser aplicadas em redes sociais.*

## 1. Introdução

Entender a evolução de tendências ao longo do tempo é crítica para que se possa analisar o comportamento de grupos de pessoas que façam parte de algum círculo social e produzam conhecimento. Esse conhecimento pode ser pensado como sendo desde opiniões sobre produtos em redes sociais digitais até artigos científicos produzidos por pesquisadores. Como o comportamento das pessoas muda com o tempo, diversos serviços procuram identificar e/ou prever tendências para conseguir melhorar a qualidade de processos ou aumentar a competitividade em negócios.

Existem diversos métodos e técnicas propostos para viabilizar a identificação de tendências. Desta forma, este artigo propõe apresentar uma revisão sistemática acerca desses métodos e técnicas. O objetivo desse trabalho é investigar o estado da arte das técnicas de identificação e análise de tendências em aplicações técnico sociais, ou seja, aplicações que estão dentro de um contexto de interação social em um âmbito tecnológico (blogs, notícias, redes sociais digitais, mercado de ações, artigos científicos, etc.).

O restante desse artigo está organizado da seguinte forma: na seção 2 será feita uma contextualização sobre os principais temas da revisão, a seção 3 apresentará a metodologia utilizada por esse trabalho. A condução da revisão será descrita na seção 4 e as considerações finais serão discutidas na seção 5.

## 2. Contextualização

Para que fique mais claro o entendimento ao que diz respeito à identificação de tendências, sua importância e o que vem sendo produzido nessa área ao longo dos últimos anos, esta seção propõe descrever alguns conceitos relevantes.

A análise de tendências consiste em coletar informações para tentar determinar um padrão de comportamento de algum grupo de informações ao longo do tempo. Informações que não possuem variação relacionada ao tempo e em que todo o contexto semântico pode ser interpretado sem a presença da dimensão temporal não necessitam de técnicas de análise de tendências. Entretanto, existem diversos tipos de informação que possuem essa variação. Um desses casos é a análise de documentos científicos. O tipo de conteúdo gerado por esses documentos varia ao longo dos anos como, por exemplo, os temas abordados (BOLELLI et al, 2009). Com o interesse em se entender essa variação e, mais ainda, tentar prever um comportamento, começou-se a estudar a análise de tendências.

A análise de tendências pode ser aplicada em diversas áreas do conhecimento como engenharia, biologia ou economia. Este artigo, no entanto, objetiva analisar os trabalhos que possuem aplicações dentro de um contexto de interação social em um âmbito tecnológico. Mesmo dentro dessa subdivisão, pode ser notado que as técnicas para identificação de tendências variam de acordo com algumas vertentes.

A análise de documentos é uma das aplicações para técnicas de identificação de tendências. *Emerging Trending Detection* (ETD) é um modelo de aplicação de identificação de tendências que identifica tópicos novos ou que estão aumentando em importância ao longo do tempo dentro do corpo de documentos. Para aumentar a eficiência e diminuir a utilização do fator humano dentro dos processos de identificação de tendências por ETD são comumente utilizadas técnicas de *Topic Trending Detection* (TDT), em que algoritmos buscam agrupar informações similares de documentos em tópicos para servir como entrada. Este tipo de solução ajudou na implementação de modelos totalmente automáticos de ETD (KONTOSTATHIS; GALITSKY; POTTENGER, 2004).

Em alguns dos trabalhos analisados pela presente revisão foi identificado outro fator de bastante importância para as técnicas de identificação de tendências. Este fator corresponde à estrutura da rede que gera informação, ou seja, o grafo de conexões entre os geradores de informação. Isto é, além de se analisar as informações dos documentos individualmente, começou-se a se observar que a forma com que os nós geradores de informação se relacionam também influi nos resultados da identificação de tendências.

## 3. Metodologia

Foi realizada uma revisão sistemática para identificar artigos relevantes relacionados ao estado da arte na área. Para iniciar a revisão foi necessário identificar algumas palavras-chave para a busca dos artigos. Essas palavras-chave foram determinadas a partir de uma pesquisa exploratória inicial. As palavras-chave utilizadas foram: *trend analysis*, *tendency analysis*, *trend identification*, *tendency identification*, *trend recognition*, *tendency recognition*, *trend detection*, *tendency detection*, *trend prediction*, *tendency prediction* e *trending*. A busca dos artigos foi realizada entre 07/10/2013 e 11/10/2013 em duas das principais bibliotecas digitais de artigos científicos na área de computação: ACM Digital Library e IEEEExplore. Essas duas bases de dados foram escolhidas por

causa do foco deste trabalho em revisar apenas artigos com aplicações de interação social em âmbito tecnológico.

Duas strings de busca foram utilizadas, uma para cada biblioteca digital:

ACM Digital Library: *Owner:ACM(Keywords:"trend analysis" OR Keywords:"tendency analysis" OR Keywords:"trend identification" OR Keywords:"tendency identification" OR Keywords:"trend recognition" OR Keywords:"tendency recognition" OR Keywords:"trend detection" OR Keywords:"tendency detection" OR Keywords:"trend prediction" OR Keywords:"tendency prediction" OR Keywords:"trending")*

IEEEExplore: *(((((("Author Keywords":"trend analysis") OR "Author Keywords":"tendency analysis") OR "Author Keywords":"trend identification") OR "Author Keywords":"tendency identification") OR "Author Keywords":"trend recognition") OR "Author Keywords":"tendency recognition") OR "Author Keywords":"trend detection") OR "Author Keywords":"tendency detection") OR "Author Keywords":"trend prediction") OR "Author Keywords":"tendency prediction") OR "Author Keywords":"trending")*

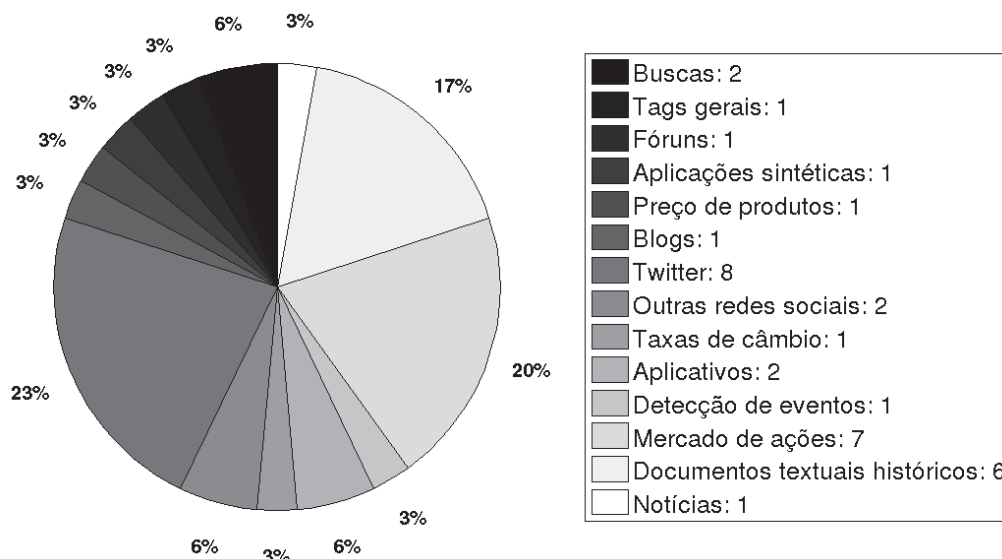
A primeira etapa da seleção dos artigos foi baseada na leitura de seus resumos (*abstract*). Na etapa posterior, todos os artigos restantes foram lidos na íntegra. A seleção dos artigos foi realizada de acordo com critérios de inclusão e exclusão preestabelecidos para a condução da revisão sistemática. Os critérios foram os seguintes: *Inclusão 1*: Trabalhos que apresentam técnicas de análise de tendências; *Exclusão 1*: Trabalhos que não apresentem os métodos utilizados; *Exclusão 2*: Trabalhos secundários; *Exclusão 3*: Trabalhos duplicados; *Exclusão 4*: Trabalhos não disponíveis integralmente; *Exclusão 5*: Trabalhos que apresentem aplicações fora do contexto sociotécnico.

A partir dos trabalhos selecionados após as duas etapas, foram extraídos dados para a construção da análise apresentada na seção 4. Todas as etapas da revisão sistemática foram avaliadas pelo autor e por outros dois pesquisadores a fim de se aumentar a confiabilidade das decisões tomadas.

#### 4. Condução

Na busca realizada nas duas bibliotecas digitais foram retornados 164 artigos. Na primeira etapa de seleção, que consistiu da leitura dos *abstracts*, foram selecionados 62 artigos, sendo 26 da ACM Digital Library e 36 da IEEEExplore. Todos esses 62 artigos foram lidos na íntegra para a segunda etapa de seleção. Nesta etapa foram selecionados 35 artigos para a condução dessa revisão.

Em primeiro lugar, os artigos foram classificados de acordo com sua aplicação. Conforme pode ser visto pela Figura 1, três tipos de aplicações se destacam: Twitter, documentos textuais históricos e mercado de ações. A alta porcentagem de utilização do Twitter como aplicação dos artigos pode ser explicada pelo alto número de usuários, pela enorme quantidade de informação gerada diariamente e a facilidade de se obter uma parcela desses dados. Os documentos científicos também têm uma alta taxa de adoção para as pesquisas de análise de tendências por servir como um bom exemplo de dados históricos e possuir algumas bases de dados amplas. Já as aplicações relacionadas ao mercado de ações são bastante utilizadas por servirem bem para técnicas que analisam séries temporais.



**Figura 1. Gráfico da distribuição das aplicações**

Existem alguns problemas relacionados à utilização de dados de redes sociais como o Twitter como aplicação. O primeiro problema acontece na obtenção dos dados. A maioria dos pesquisadores utiliza a API do Twitter para essa ação e esta API possui alguns limites quanto ao número de chamadas por tempo, o que prejudica a importação desses dados<sup>1</sup>. O outro problema consiste em não se poder replicar os experimentos por causa do dinamismo dos dados (quando os autores não disponibilizam exatamente os dados que foram utilizados, pois estes não podem ser facilmente consultados no Twitter). Por outro lado, o Twitter apresenta o benefício de apresentar dados contemporâneos.

Com a análise dos artigos, observou-se que as técnicas de análise de tendências, em sua maioria, podem ser segmentadas em dois grupos: ETD e análise de séries temporais. Os artigos de ETD têm como aplicações análise de redes sociais em geral e documentos textuais históricos. Enquanto os artigos de análise de séries temporais têm como principal aplicação o mercado financeiro. Isso pode ser explicado pelo fato de que grande parte da pesquisa relacionada ao mercado financeiro se utiliza apenas dos dados de volume e valores (volume de venda de ações, valor máximo e mínimo diário, por exemplo) e tempo. Já pesquisas de ETD necessitam identificar informações relevantes dentre uma enorme gama de informação (caso do Twitter) e identificar tendências a partir delas. ETD também difere de análise de séries temporais por focar em tópicos emergentes e muitas vezes não identificar o comportamento de tópicos cujas tendências não são positivas. Alguns outros artigos analisados tiveram como objetivo apenas o TDT, cujas técnicas podem ser utilizadas para um sistema completo de análise de tendências.

Como as técnicas analisadas nos artigos foram utilizadas de diferentes formas e algumas vezes com diferentes objetivos, foi criada uma abordagem classificatória de tais técnicas. Três classes foram determinadas: Identificação de tópicos, Detecção de tendências emergentes e Reconhecimento de padrões. As técnicas de identificação de tópicos são técnicas que buscam agrupar termos por algum tipo de similaridade ou mesmo entender a distribuição dos termos e criar tópicos a partir disso. As técnicas de

<sup>1</sup> <https://dev.twitter.com/docs/rate-limiting/1.1>

detecção de tendências emergentes são baseadas em contagem e conseguem identificar aumentos súbitos no aparecimento de termos em determinados períodos. Por fim, as técnicas de reconhecimento de padrões procuram aprender o comportamento dos tópicos ou termos e auxiliar na predição de tais comportamentos. Algumas técnicas utilizadas nos artigos não se encaixam em nenhum destes três grupos principais e não foram contabilizadas porque foram utilizadas com o objetivo de auxiliar as técnicas principais. Um exemplo são as técnicas de séries simuladas que serviram como forma de suavizar séries temporais e aumentar a taxa de predição dos modelos em alguns trabalhos.

A distribuição das técnicas pode ser visualizada nas figuras 2, 3 e 4.

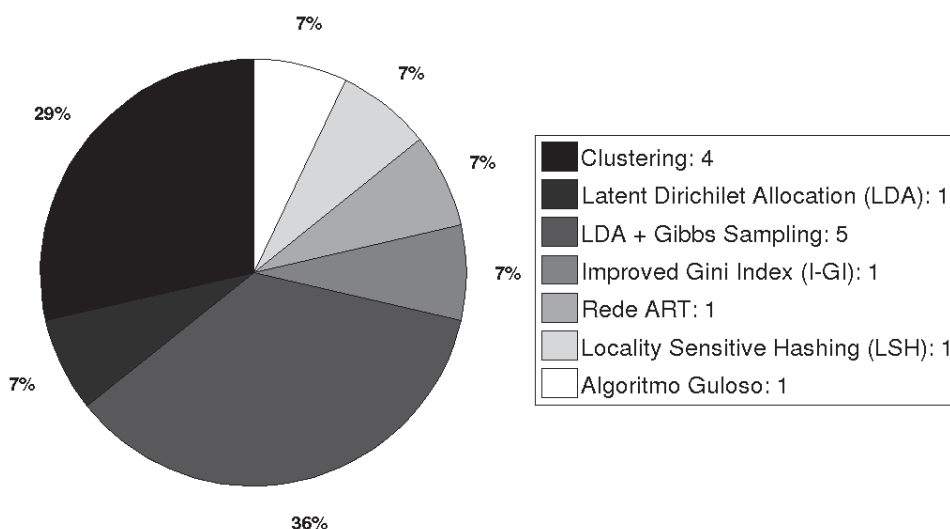


Figura 2. Gráfico de distribuição das técnicas de identificação de tópicos

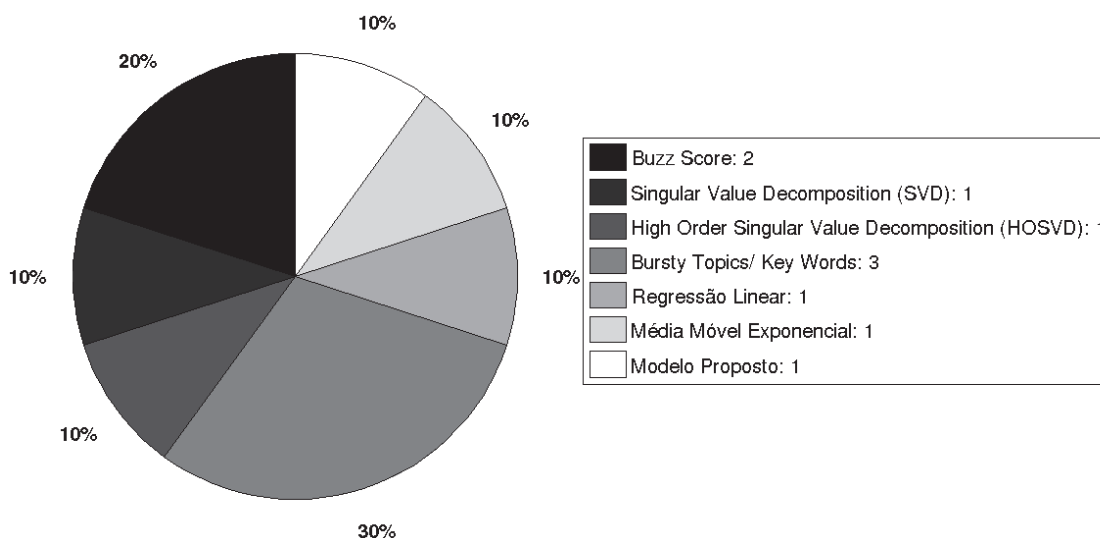
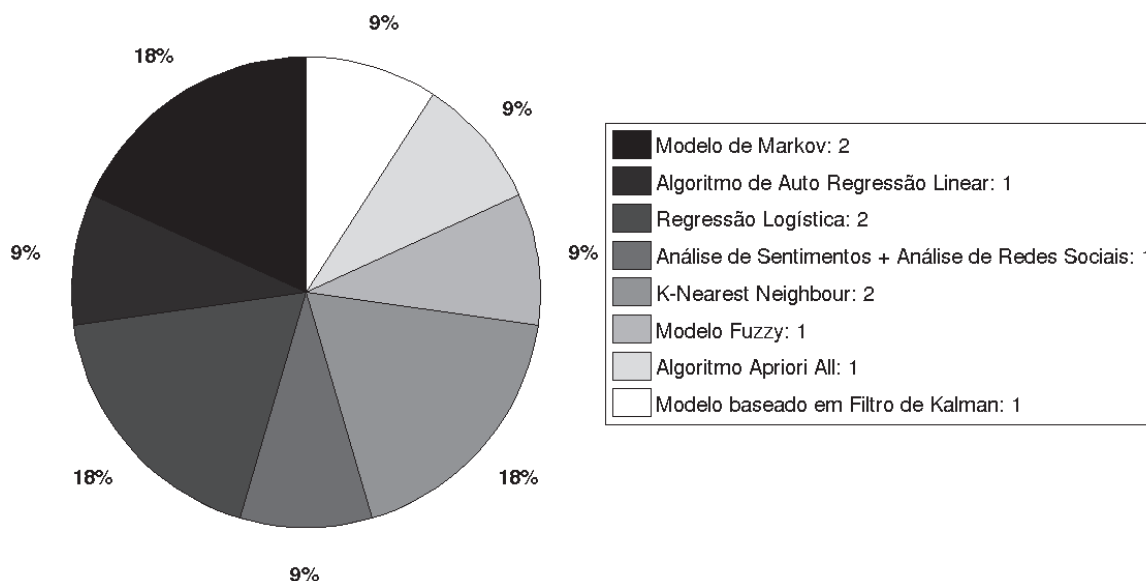


Figura 3. Gráfico de distribuição das técnicas de tendências emergentes

Nota-se, pelas figuras, que as técnicas estão bastante distribuídas pelos trabalhos. Entretanto, algumas poucas técnicas fogem desse padrão. São elas LDA com *Gibbs Sampling* e *Clustering* que são utilizadas pela maioria dos trabalhos, como mostrado na figura 2. Em relação às técnicas de reconhecimento de padrões há um ponto a se ressaltar. Alguns dos autores decidiram utilizar determinadas técnicas pela capacidade de se adaptar a fluxos de dados grandes e dinâmicos com uma complexidade baixa. É a explicação, por exemplo, da utilização de KNN em alguns dos trabalhos.



**Figura 4. Gráfico de distribuição das técnicas de reconhecimento de padrões**

Os artigos que usam LDA com *Gibbs Sampling* foram os artigos de Bolelli et al. (2009), Dey et al. (2009), Kawamae (2010), Kawamae (2012) e Martie et al. (2012). No trabalho de Bolelli et al. (2009), LDA e *Gibbs Sampling* foram utilizados conjuntamente com a ordem temporal dos documentos para a criação de um modelo generativo que aprende as distribuições de autor, tópico e palavra. Em uma aplicação sintética, houve uma taxa de acerto de aproximadamente 72%. Dey et al. (2009) propuseram um modelo de detecção de eventos utilizando também técnicas de *clustering* para entender a correlação dos eventos com a variação do mercado financeiro. Os autores apresentam alguns gráficos onde é possível perceber estas correlações. O trabalho de Kawamae (2010) consistiu em tentar prever as distribuições dos tópicos em artigos científicos levando o tempo em consideração. Com a mesma ideia, Kawamae (2012) ainda estabeleceu uma diferença entre tópicos estáveis (não possuem variação significativa ao longo do tempo) e tópicos dinâmicos, tentando rebater outros modelos que apenas levam em consideração as explosões de tópicos (aumentos súbitos de aparecimento de tópicos em determinados períodos). Com uma aplicação diferenciada, Martie et al. (2012) apresentaram uma abordagem de identificação de tendências nas discussões sobre falhas em um fórum sobre o desenvolvimento de um dos projetos de código aberto da plataforma Android. Com as relações apresentadas entre as tendências identificadas e as atualizações das versões do Android foi possível especular sobre a correção de alguns erros durante essas atualizações. Chen et al. (2013) apenas utilizaram LDA para identificar tópicos e realizaram manualmente uma verificação para analisar quais dos tópicos eram tendências.

Técnicas de *clustering* para identificação de tópicos foram utilizadas nos trabalhos de Al Bawab et al. (2012), Cvijikj et al. (2011), Pervin et al. (2013) e Voigt et al. (2013). Al Bawab et al. (2012) apresentaram um modelo que identifica os termos mais buscados pelos usuários utilizando *buzz score*; determina por afinidade a localização espacial desses termos utilizando entropia; e agrupa esses termos em tópicos levando em consideração a taxa de busca e a localização. O objetivo foi melhorar a ferramenta *Trending Now* do *Yahoo!*. O trabalho gerou um aumento de 5,91% de cliques nos tópicos por localização em relação apenas aos tópicos globais. No artigo de Cvijikj et al. (2011), textos de status do *Facebook* foram tidos como aplicação. Os termos mais importantes foram identificados por TF-IDF e posteriormente agrupados. Com a análise, foi estabelecida por eles uma classificação dos tópicos em eventos disruptivos, tópicos populares e rotinas diárias. Pervin et al. (2013) utilizaram uma abordagem invertida em que eles primeiro agruparam as palavras hierarquicamente para então utilizar a técnica de *burst topic* e encontrar os tópicos mais citados no *Twitter*. Voigt et al. (2013) apresentaram um modelo que apenas identifica tópicos a partir da análise de notícias. Kaleel et al. (2013) utilizaram *Locality Sensitive Hashing* (LSH) para fazer agrupamento. Diferentemente dos anteriores, Gollapudi e Sivakumar (2004) propuseram um modelo que não necessita de reorganização dos dados em grande escala. Todo o processo de agrupamento é feito em uma etapa de pré-processamento com a utilização de métricas de atributo adequadas, selecionadas pela aproximação realizada por algoritmos de árvores métricas e algum conhecimento prévio sobre o domínio. O modelo utiliza o modelo de Markov para a previsão de tendências.

Barbagallo et al. (2011) e Mathioudaki e Koudas (2010) utilizaram técnicas de *bursty topics / key words* para identificar os tópicos mais citados no *Twitter* em períodos determinados. O segundo integrou um algoritmo guloso para a identificação de tópicos para que a complexidade do modelo fosse baixa se adaptando com mais facilidade a um fluxo alto de dados.

Abe e Tsumoto (2009) selecionaram termos de importância por TF-IDF e coeficiente de Jaccard utilizando regressão linear posteriormente para detectar tendências emergentes. Todas as tendências detectadas foram confirmadas como tendências reais por especialistas dos domínios.

Golbandi et al. (2013) e Kamath e Caverlee (2011) também utilizaram técnicas de detecção de tendências emergentes. Golbandi et al. (2013) apresentaram um modelo com um algoritmo de auto regressão linear para prever o comportamento das buscas dos usuários que serviu como entrada para o algoritmo de *Buzz score*, que identifica as tendências emergentes. Já Kamath e Caverlee (2011) propuseram um modelo em que ‘frases’ (conjunto de um ou mais termos) recebem pontuações para cada tempo e então se determina um limiar para indicar o que é e o que não é tendência.

Zhu et al. (2011) propuseram um algoritmo para a interpretação do comportamento da variação de preços de telefones celulares em algumas lojas virtuais. Park et al. (2011) apresentaram uma abordagem baseada em *Improved Gini Index* (I-GI) para seleção de atributos de tópicos escolhidos e utilizaram SVM para avaliação. Utilizando *F1-measure*, o resultado chegou a 98%. No trabalho de Jayashri e Chitra (2012) foi proposto um modelo com rede ART para identificar tópicos e detectar tendências através dos picos dos tópicos extraídos de documentos científicos.

Os trabalhos que utilizaram técnicas para análise de séries temporais foram os de Christiansen et al. (2012), Heshan e Qingshan (2008), Huang et al. (2011), Kato et al. (2010), Teixeira e Oliveira (2009), Wu et al. (2012), Yonghong e Wenyang (2009) e Gong e Sun (2009). Christiansen et al. (2012) apresentaram uma abordagem que consiste em segmentar séries temporais de acordo com variações bruscas, agrupar esses segmentos por similaridade e descrever cada tópico como um conjunto desses segmentos. Por fim, um modelo de Markov foi utilizado para fazer a previsão de tendência dos tópicos no mês seguinte e foi obtida uma média de acerto de 92,9% para uma parte da base de dados do *Google Trends*. Heshan e Qingshan (2008) utilizaram uma técnica de séries simuladas chamada *Ascending Triangle* para a entrada do algoritmo KNN enquanto Huang et al. (2011) utilizaram uma técnica de suavização de séries temporais chamada *Price Percentage Oscillator* (PPO) com um modelo Fuzzy para predição. Kato et al. (2010) utilizaram *Exponential Moving Average* (EMA) para detectar os altos e baixos das taxas de câmbio e alimentaram um algoritmo genético com essa informação para otimizar o ganho de compra e venda de moedas. Teixeira e Oliveira (2009) propuseram um modelo com um algoritmo de KNN para prever os preços de ações cujas informações de entrada foram obtidas de ferramentas técnicas de mercado financeiro. No final, foi comparado o resultado do modelo proposto com a estratégia de mercado chamada *Buy and Hold*. Simulando transações financeiras, observou-se que a estratégia proposta obteve um melhor resultado para compra e venda em 12 das 15 ações do experimento. Wu et al. (2012) utilizaram *K-means* para agrupar segmentos de séries temporais e o algoritmo *AprioriAll* para buscar por padrões. Yonghong e Wenyang (2009) propuseram um modelo que adota a matriz de ganho de Kalman para calcular automaticamente a estimativa de probabilidade máxima de um fluxo de dados. Por fim, Gong e Sun (2009) utilizaram regressão logística para a previsão de tendências em mercado de ações.

Além de todos os modelos apresentados, existem ainda abordagens que levam em consideração a estrutura das conexões entre as fontes das informações analisadas e a importância individual delas nessa estrutura. Chi et al. (2006) compuseram três vetores de informação para a análise de tendências: vetor de tendências (aspecto temporal), vetor de autoridade (importância das fontes de informação) e vetor de *hub* (modelagem das fontes de informação como um estrutura de grafo). A técnica *High-Order Singular Value Decomposition* (HOSVD) foi utilizada e os resultados mostraram que o acréscimo da informação de localização das fontes em relação à estrutura de conexões aumenta a precisão do modelo. Sha et al. (2012) realizaram um estudo sobre tendências em um aplicativo específico de recomendação de fotos. Os usuários do aplicativo foram classificados entre as classes *trend makers* e *trend spotters* e a importância e a relação entre eles foram consideradas para a detecção de tendências. Por fim, regressão linear foi utilizada para prever se uma determinada foto será ou não tendência baseada nos usuários relacionados à foto. Gloor et al. (2009) desenvolveram um conceito chamado *Web buzz index* para tópicos minerados por diversas fontes na internet. Esse índice é baseado inteiramente em técnicas de análise de redes sociais incluindo análise de sentimentos. O gráfico do *Web buzz index* foi plotado juntamente a uma série temporal financeira e pôde ser observado que as duas têm um comportamento bastante parecido.

Alguns artigos analisados tinham como foco apenas classificar de alguma forma as informações de tendências em redes sociais digitais. Cheong e Lee (2009) utilizaram rede SOM para verificar a distribuição dos *trending topics* do *Twitter* em relação à localização dos usuários. Khan et al. (2012a) propuseram classificar *tweets* que



realmente sinalizavam um surto de doenças em alguma localidade utilizando *Naive Bayes* para classificação. Khan et al. (2012b) também utilizaram *Naive Bayes*, mas com o objetivo de colocar *tags* automáticas em *tweets*. Zubiaga et al. (2011) utilizaram SVM para classificar os *tweets* em notícias, eventos correntes, *memes* ou comemorações.

## 5. Considerações Finais

Este trabalho apresentou uma revisão sistemática acerca dos métodos e técnicas utilizados por diferentes trabalhos para a detecção de tendências. Esses métodos foram categorizados de acordo com a função exercida por cada um com o objetivo comum de identificação de tendências. Com essa revisão ficou claro que grande parte dos trabalhos consiste da experimentação de intercalação de técnicas em diferentes etapas do processo. Além disso, é importante considerar que o domínio das aplicações também influi na utilização de técnicas diferentes. Os métodos utilizados em séries temporais, por exemplo, se diferenciam bastante das técnicas utilizadas em aplicações de documentos históricos.

Uma dificuldade ainda existente nessa área é relacionada a forma de avaliação das técnicas. Ainda não existe um padrão de avaliação que facilite a comparação de resultados entre os trabalhos e isso dificulta a determinar se um método leva vantagem sobre outro em relação a alguma variável global. Outro problema, como já discutido, é a replicação dos experimentos. Grande parte dos trabalhos utiliza dados dinâmicos para os experimentos, como dados do *Twitter*, e não é possível obter os mesmos dados para replicar os experimentos. Apenas em 23% dos artigos analisados na revisão é possível fazer a replicação dos experimentos. Uma alternativa para solucionar esse problema seria os autores armazenarem e disponibilizarem esses dados de alguma forma para que outras pessoas conseguissem utilizar os mesmo dados.

Por fim, de todos os trabalhos analisados, ainda são poucos os que utilizam a estrutura das fontes de informação como variável nos experimentos. Há, ainda, um grande campo a ser explorado sobre a utilização da teoria de redes sociais na detecção e análise de tendências.

## Agradecimentos

O trabalho apresentado neste artigo foi parcialmente financiado pela CAPES (bolsa de mestrado) e pelo CNPq (Projeto Universal e Bolsa Produtividade em Pesquisa).

## Referências

ABE, H.; TSUMOTO, S. Evaluating a method to detect temporal trends of phrases in research documents. In: IEEE INTERNATIONAL CONFERENCE ON COGNITIVE INFORMATICS, 8th, 2009, Hong Kong. **Proceedings...** Hong Kong: IEEE, 2009. p. 378–383.

BARBAGALLO, D. et al. Semi-automated Methods for the Annotation and Design of a Semantic Network Designed for Sentiment Analysis of Social Web Content. In: INTERNATIONAL WORKSHOP ON DATABASE AND EXPERT SYSTEMS APPLICATIONS, 22nd, 2011, Toulouse. **Proceedings...** Toulouse: IEEE, 2011. p. 222–226.

BAWAB, Z. AL; MILLS, G. H.; CRESPO, J.-F. Finding trending local topics in search queries for personalization of a recommendation system. In: ACM SIGKDD

INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING – KDD, 18th, 2012, New York. **Proceedings...** New York: ACM Press, 2012. p. 397-405.

BOLELLI, L. et al. Finding topic trends in digital libraries. In: JOINT INTERNATIONAL CONFERENCE ON DIGITAL LIBRARIES – JCDL, 2009, New York. **Proceedings...** New York: ACM Press, 2009. p. 69-72.

CHEN, L.; ZHANG, C.; WILSON, C. Tweeting Under Pressure: Analyzing Trending Topics and Evolving Word Choice on SinaWeibo. In: ACM CONFERENCE ON ONLINE SOCIAL NETWORKS – COSN, 1st, 2013, New York. **Proceedings...** New York: ACM Press, 2013. p. 89–100.

CHEONG, M.; LEE, V. Integrating web-based intelligence retrieval and decision-making from the twitter trends knowledge base. In: ACM WORKSHOP ON SOCIAL WEB SEARCH AND MINING – SWSM, 2nd, 2009, New York. **Proceedings...** New York: ACM Press, 2009. p. 1–8.

CHI, Y.; TSENG, B. L.; TATEMURA, J. Eigen-Trend: Trend Analysis in the Blogosphere based on Singular Value Decompositions. In: ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT – CIKM, 15th, 2006, New York. **Proceedings...** New York: ACM Press, 2006. p. 68–77.

CHRISTIANSEN, L. et al. Modeling topic trends on the social web using temporal signatures. In: INTERNATIONAL WORKSHOP ON WEB INFORMATION AND DATA MANAGEMENT – WIDM, 12th, 2012, New York. **Proceedings...** New York: ACM Press, 2012. p. 3–10.

CVIJKJ, I. P.; MICHAHELLES, F. Monitoring Trends on Facebook. In: IEEE INTERNATIONAL CONFERENCE ON DEPENDABLE, AUTONOMIC AND SECURE COMPUTING, 9th, 2011, Sydney. **Proceedings...** Sydney: IEEE, 2011. p.895–902.

DEY, L.; MAHAJAN, A.; HAQUE, S. M. Document Clustering for Event Identification and Trend Analysis in Market News. 2009 Seventh In: INTERNATIONAL CONFERENCE ON ADVANCES IN PATTERN RECOGNITION, 7th, 2009, Kolkata. **Proceedings...** Kolkata: IEEE, 2009. p.103–106.

GLOOR, P. A. et al. Web Science 2.0: Identifying Trends through Semantic Social Network Analysis. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL SCIENCE AND ENGINEERING, 2009, Vancouver. **Proceedings...** Vancouver: IEEE, 2009. p. 215–222.

GOLBANDI, N. G. et al. Expediting search trend detection via prediction of query counts. In: ACM INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING – WSDM, 6th, 2013, New York. **Proceedings...** New York: ACM Press, 2013. p. 295-304.

GOLLAPUDI, S.; SIVAKUMAR, D. Framework and algorithms for trend analysis in massive temporal data sets. In: ACM CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT – CIKM, 13th, 2004, New York. **Proceedings...** New York: ACM Press, 2004. p. 168-177.

GONG, J.; SUN, S. A New Approach of Stock Price Prediction Based on Logistic Regression Model. In: INTERNATIONAL CONFERENCE ON NEW TRENDS IN

INFORMATION AND SERVICE SCIENCE, 2009, Beijing. **Proceedings...** Beijing: IEEE, 2009. p. 1366–1371.

HESHAN GUAN; QINGSHAN JIANG. Pattern matching of time series and its application to trend prediction. In: INTERNATIONAL CONFERENCE ON ANTI-COUNTERFEITING, SECURITY AND IDENTIFICATION, 2nd, 2008, Guiyang. **Proceedings...** Guiyang: IEEE, 2008. p. 41–44.

HUANG, H.; PASQUIER, M.; QUEK, C. Financial Market Trading System With a Hierarchical Coevolutionary Fuzzy Predictive Model. **IEEE Transactions on Evolutionary Computation**, v. 13, n. 1, p. 56–70, 2009.

JAYASHRI, M.; CHITRA, P. Topic clustering and topic evolution based on temporal parameters. In: INTERNATIONAL CONFERENCE ON RECENT TRENDS IN INFORMATION TECHNOLOGY, 2012, Chennai. **Proceedings...** Chennai: IEEE, 2012. p. 559–564.

JOHANSSON, F. et al. Detecting Emergent Conflicts through Web Mining and Visualization. In: EUROPEAN INTELLIGENCE AND SECURITY INFORMATICS CONFERENCE, 2011, Athens. **Proceedings...** Athens: IEEE, 2011. p. 346–353.

KALEEL, S. B.; ALMESHARY, M.; ABHARI, A. Event detection and trending in multiple social networking sites. In: COMMUNICATIONS & NETWORKING SYMPOSIUM ARTICLE NO. 5, 16th, 2013, San Diego. **Proceedings...** San Diego: Society for Computer Simulation International, 2013. p. 1–5.

KAMATH, K. Y.; CAVERLEE, J. Discovering trending phrases on information streams. In: ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT – CIKM, 20th, 2011, New York. **Proceedings...** New York: ACM Press, 2011. p. 2245-2248.

KATO, D.; YATA, N.; NAGAO, T. Evolutionary trend prediction using plural technical indicators for foreign exchange transaction. In: SICE ANNUAL CONFERENCE, 2010, Taipei. **Proceedings...** Taipei: IEEE, 2010. p. 1170-1175.

KAWAMAE, N. Theme Chronicle Model: Chronicle Consists of Timestamp and TopicalWords over Each Theme. In: ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT – CIKM, 21st, 2012, New York. **Proceedings...** New York: ACM Press, 2012. p. 2065-2069.

KAWAMAE, N.; HIGASHINAKA, R. Trend detection model. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB – WWW, 19th, 2010, New York. **Proceedings...** New York: ACM Press, 2010. p.1129-1130.

KHAN, M. A. H.; IWAI, M.; SEZAKI, K. Towards urban phenomenon sensing by automatic tagging of tweets. In: INTERNATIONAL CONFERENCE ON NETWORKED SENSING (INSS), 9th, 2012, Antwerp. **Proceedings...** Antwerp: IEEE, 2012a. p. 1–7.

KHAN, M. A. H.; IWAI, M.; SEZAKI, K. A robust and scalable framework for detecting self-reported illness from twitter. In: IEEE INTERNATIONAL CONFERENCE ON E-HEALTH NETWORKING, APPLICATIONS AND SERVICES (HEALTHCOM), 14th, 2012, Beijing. **Proceedings...** Beijing: IEEE, 2012b. p. 303–308.

KONTOSTATHIS, A.; GALITSKY, L.; POTTENGER, W. A survey of emerging trend detection in textual data mining. **Springer**, 2004.

MARTIE, L. et al. Trendy bugs: Topic trends in the Android bug reports. In: IEEE WORKING CONFERENCE ON MINING SOFTWARE REPOSITORIES (MSR), 9th, 2012, Zurich. **Proceedings...** Zurich: IEEE, 2012. p. 120–123.

MATHIOUDAKIS, M.; KOUDAS, N. TwitterMonitor: Trend Detection over the Twitter Stream. In: INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA – SIGMOD, 2010, New York. **Proceedings...** New York: ACM Press, 2010. p. 1155-1157.

PARK, H. et al. Detection and Analysis of Trend Topics for Global Scientific Literature Using Feature Selection Based on Gini-Index. In: IEEE INTERNATIONAL CONFERENCE ON TOOLS WITH ARTIFICIAL INTELLIGENCE, 23rd, 2011, Boca Raton. **Proceedings...** Boca Raton: IEEE, 2011. p. 965–969.

PERVIN, N. et al. Fast, Scalable, and Context-Sensitive Detection of Trending Topics in Microblog Post Streams. **ACM Transactions on Management Information Systems (TMIS)**, v. 3, n. 4, p. 19-43, 2013.

SHA, X. et al. Spotting Trends: The Wisdom of the Few. In: ACM CONFERENCE ON RECOMMENDER SYSTEMS – RECSYS, 6th, 2012, New York. **Proceedings...** New York: ACM Press, 2012. p.51-58.

TEIXEIRA, L. A.; OLIVEIRA, A. L. I. DE. Predicting stock trends through technical analysis and nearest neighbor classification. In: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS, 2009, San Antonio. **Proceedings...** San Antonio: IEEE, 2009. p. 3094–3099.

VOIGT, M.; ALEYTHE, M.; WEHNER, P. Towards topics-based, semantics-assisted news search. In: INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE, MINING AND SEMANTICS – WIMS, 3rd, 2013, New York. **Proceedings...** New York: ACM Press, 2013. p. 1-7.

WU, Y.-P.; WU, K.-P.; LEE, H.-M. Stock Trend Prediction by Sequential Chart Pattern via K-Means and AprioriAll Algorithm. In: CONFERENCE ON TECHNOLOGIES AND APPLICATIONS OF ARTIFICIAL INTELLIGENCE, 2012, Tainan. **Proceedings...** Tainan: IEEE, 2012. p. 176–181.

YONGHONG, Y.; WENYANG, B. Forecasting Model over Random Interval Data Stream Based on Kalman Filter. In: INTERNATIONAL WORKSHOP ON DATABASE TECHNOLOGY AND APPLICATIONS, 1st, 2009, Wuhan. **Proceedings...** Wuhan: IEEE, 2009. p. 448–451.

ZHU, Q. et al. Commodities Price Dynamic Trend Analysis Based on Web Mining. In: INTERNATIONAL CONFERENCE ON MULTIMEDIA INFORMATION NETWORKING AND SECURITY, 3rd, 2011, Shanghai. **Proceedings...** Shanghai: IEEE, 2011. p.524–527.

ZUBIAGA, A. et al. Classifying Trending Topics: A Typology of Conversation Triggers on Twitter Arkaitz. In: ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT – CIKM, 20th, 2011, New York. **Proceedings...** New York: ACM Press, 2011. p. 2461-2464.