UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Análise e classificação de rumores em redes sociais

Nícolas Roque dos Santos

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

ICMC USP
SÃO CARLOS

Nícolas Roque dos Santos

Análise e classificação de rumores em redes sociais

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Ciências de Computação e Matemática Computacional. *EXEMPLAR DE DEFESA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientadora: Profa. Dra. Rosane Minghim

USP – São Carlos
Novembro de 2019

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

R786a

Roque dos Santos, Nícolas
Análise visual e classificação de rumores em redes sociais / Nícolas Roque dos Santos; orientadora Rosane Minghim. -- São Carlos, 2019.
84 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2019.

1. Rumor. 2. Análise visual. 3. Classificação supervisionada. 4. Aprendizado de Máquina. 5. Visualização de Dados. I. Minghim, Rosane, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Nícolas Roque dos Santos

Visual analysis and classification of rumors in social networks

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master Program in Computer Science and Computational Mathematics. *EXAMINATION BOARD PRESENTATION COPY*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Profa. Dra. Rosane Minghim

USP – São Carlos
November 2019

Este trabalho é dedicado a todos que pensaram em desistir, mas persistiram até o fim.

AGRADECIMENTOS

Agradeço aos meus pais, Benedito e Márcia, e minha irmã Tainá, que não deixaram de acreditar em mim, mesmo quando eu já não acreditava mais no meu potencial, e sempre estiveram presentes nos momentos que mais precisei de amparo e de um ombro amigo.

Agradeço a Prof. Dra. Rosane Minghim pela orientação, por ter me ensinado inúmeras coisas, pela paciência, pela compreensão e pelos ocasionais puxões de orelha.

Agradeço ao Prof. Dr. Evangelos Milios, Prof. Dr. Abidalrahman Moh'd e Anh Dang pela orientação e pelo suporte fornecido durante minha estadia em Halifax.

Agradeço aos amigos Diego Cintra, Henry Heberle e Gladys Hilasaca pelo companheirismo, momentos de risada e conselhos.

Agradeço aos colegas do VICG que contribuíram para o desenvolvimento desta pesquisa e pela convivência no dia a dia.

Agradeço a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo financiamento durante este mestrado e ao Emerging Leaders in the Americas Program (ELAP) pelo financiamento do estágio em pesquisa realizado na Dalhousie University.

“Se algo é importante o suficiente,
faça-o mesmo que as chances não estejam a seu favor.”
(Elon Musk)

RESUMO

SANTOS, N. R. **Análise e classificação de rumores em redes sociais**. 2019. 84 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

O aumento da quantidade de pessoas com acesso à internet nos últimos anos contribuiu para o aumento da quantidade de usuários de redes sociais. Entretanto, a falta de monitoramento do que é publicado nas redes sociais pode levar ao surgimento de rumores, que são informações cuja veracidade, no momento de seu surgimento, não pode ser comprovada ou negada. A Análise de Redes Sociais é uma tarefa que envolve esforço de diferentes áreas, como a Ciência da Computação, Matemática e Psicologia, para investigar os usuários e as relações entre eles, e a disseminação de informações. A Visualização de Dados e o Aprendizado de Máquina são subáreas da Ciência da Computação que permitem a descoberta de padrões e anomalias em um conjunto de dados. Neste trabalho de mestrado foram utilizados conceitos de ambas subáreas e da Análise de Redes Sociais na realização de duas análises visuais e uma classificação supervisionada. A primeira análise visual tem como objetivo a comparação entre o *Reddit* e o *Twitter* no contexto de propagação de rumores. Essa análise possibilitou a identificação de semelhanças e diferenças existentes entre as duas redes sociais. A segunda análise visual tem como finalidade a identificação dos pontos similares entre um rumor verdadeiro e um rumor falso, e os pontos nos quais eles diferem. Uma classificação supervisionada foi também realizada com o objetivo de detectar se um usuário acredita no rumor que ele está propagando. Para isto, parte do conjunto de dados coletado foi anotado manualmente, classificado e avaliado. Os resultados obtidos mostram que a utilização de duas classes (positivo e negativo) na classificação atingiu resultados satisfatórios, ao contrário do que ocorreu quando três classes (positivo, neutro e negativo) foram utilizadas. Em conjunto, essas tarefas buscaram fornecer elementos para novas estratégias de identificação de rumores.

Palavras-chave: Rumor, Análise Visual, Visualização de Dados, Aprendizado de Máquina, Redes Sociais, Classificação Supervisionada.

ABSTRACT

SANTOS, N. R. **Visual analysis and classification of rumors in social networks**. 2019. 84 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

The extense access to internet contributed to a spike in social network users. However, the lack of control in what is published may lead to the spread of rumors, which are unverified information. Social Network Analysis is a task that involves effort from different areas, such as Computer Science, Mathematics and Psychology, to investigate users and the relations between them, and information dissemination. Data Visualization and Machine Learning are Computer Science subareas that allow the discovery of patterns and anomalies of a dataset. Concepts from both subareas and Social Network Analysis were employed to perform two visual analysis and one supervised classification in this Master's research work. The goal of the first visual analysis is the comparison between Reddit and Twitter in the context of rumor propagation. This analysis allowed the identification of the existing similarities and differences between posts in either social network. The goal of the second visual analysis is the identification of similarities and differences between a true rumor and a false rumor. A supervised classification was performed to detect if a user believes in the rumor that he or she is propagating. In order to do so, part of the collected dataset was manually annotated, classified and measured. The results show that the use of two classes (positive and negative) in the classification achieved satisfactory results, as opposed to when three classes (positive, neutral and negative) were used. Together, these tasks seek to provide elements for new rumor identification strategies.

Keywords: Rumor, Visual Analysis, Data Visualization, Machine Learning, Social Networks, Supervised Classification.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

CAPÍTULO

1

INTRODUÇÃO

1.1 Contextualização

Redes sociais, como *Twitter*[1], *Facebook*[2] e *Reddit*[3], estão cada vez mais inseridas no cotidiano das pessoas. De acordo com (STATISTA, 2019), em Janeiro de 2019 o Facebook possuía mais de 2.2 bilhões de usuários, o Reddit possuía 330 milhões de usuários e o Twitter possuía 326 milhões de usuários. Entre os motivos que levam pessoas e organizações a utilizarem as redes sociais, pode-se destacar: obter informações, participar de debates, divulgar produtos e serviços, socializar, criar novas amizades e comunicar-se com amigos e familiares (BRANDTZÆG; HEIM, 2009). Entretanto, um problema que as redes sociais possuem é a falta de um sistema eficiente para o monitoramento das informações que são propagadas pela rede, contribuindo com o surgimento e com a disseminação de rumores (WEBB *et al.*, 2016) (ZUBIAGA *et al.*, 2017).

Rumor é uma informação que está em circulação cuja veracidade não foi comprovada ou negada. O conceito de veracidade é utilizado para definir um rumor como falso, verdadeiro ou não resolvido (DIFONZO; BORDIA, 2007). Rumores são gerados principalmente por notícias de última hora, como furos de reportagem, que são publicadas por portais de notícias que não verificaram adequadamente suas fontes. Tais portais atualizam constantemente estas notícias, podendo, a cada atualização, fornecer uma visão diferente da notícia ao usuário (ZUBIAGA *et al.*, 2017). Outras fontes de rumores existem mas têm menor poder de propagação.

Um problema que ocorre com a existência de um rumor é a capacidade que ele possui em alterar o comportamento da parcela da sociedade que entra em contato com ele. Em 2009 o governo norte-americano precisou criar um website para desmentir rumores relacionados à Gripe Suína (MOROZOV, 2009) (CHEW; EYSENBACH, 2010). Após o terremoto de 2011 no

[1] <www.twitter.com>
[2] <www.facebook.com>
[3] <www.reddit.com>

Japão, vários rumores atingiram a sociedade japonesa. Um deles afirmava que a gasolina estava acabando, levando a população a comprar e estocar gasolina (TAKAYASU *et al.*, 2015) (HASHIMOTO; KUBOYAMA; SHIROTA, 2011). As ações do mercado financeiro norte-americano sofreram uma queda após uma informação falsa ter sido publicada no Twitter oficial da agência de notícias *Associated Press*[4]. Esta publicação afirmava afirmava que o presidente Barack Obama estava ferido devido à uma explosão que ocorreu dentro da Casa Branca (ELBOGHDADY, 2013).

Existem esforços para criar sistemas que identificam rumores a fim de inibi-los ou diminuir seu impacto (RATKIEWICZ *et al.*, 2011) (QAZVINIAN *et al.*, 2011). Entretanto, existem alguns problemas em relação a esta abordagem. Primeiramente, as características de um rumor variam entre redes sociais e entre rumores. Por exemplo, rumores a respeito de celebridades podem diferir consideravelmente entre si, dependendo da celebridade ser um político ou uma atriz famosa. Outro empecilho neste tipo de trabalho é que o comportamento dos usuários envolvidos em um rumor varia de um rumor para o outro (DANG *et al.*, 2016). Além disso, não é trivial determinar o momento em que um rumor iniciou, pois um rumor é sujeito a períodos de dormência, a ser resolvido ou a desaparecer (GEORGE; JONES, 2007).

A Análise de Redes Sociais é uma tarefa que procura identificar e entender o que os usuários e suas relações implicam na composição da rede. Além disso, outro objetivo desta tarefa é a avaliação da mudança sofrida por uma rede social ao longo do tempo (WASSERMAN; FAUST, 1994). Esta tarefa envolve esforços de diferentes áreas científicas, sendo uma delas a Ciência da Computação. Na Ciência da Computação, a Análise de Redes Sociais emprega conceitos de subáreas como a Visualização de Dados e o Aprendizado de Máquina.

A Visualização de Dados é uma subárea que fornece suporte à analise de dados, principalmente aqueles com alta complexidade ou de natureza exploratória, cujos modelos de comportamento não estão bem definidos. Usando o fato de que a visão é um sentido chave para o entendimento de informações, essas técnicas apresentam informações por meio de representações gráficas geradas por um computador. São apresentadas aos analistas visões que revelam diferentes aspectos, como anomalias e padrões estruturais dos dados aos quais as técnicas de visualização foram aplicadas (WARD; GRINSTEIN; KEIM, 2010).

Aprendizado de Máquina é uma subárea da Ciência da Computação que estuda modelos estatísticos e algoritmos que são capazes de detectar padrões em dados automaticamente e, partir desses padrões, predizer novos dados (MURPHY, 2012). Além disso, é possível utilizar os métodos desta área para o auxílio em tarefas de tomadas de decisão.

Alguns trabalhos que fizeram uso de conceitos da Análise de Rede Social, Aprendizado de Máquina ou da Visualização de Dados para estudar a disseminação de rumores obtiveram os seguintes resultados: categorização de rumores (BOYANDIN *et al.*, 2011), detecção da

[4] <https://twitter.com/AP>

veracidade de rumores disseminados no *Reddit* (DANG *et al.*, 2019) (ZUBIAGA *et al.*, 2016), modelagem do comportamento dos usuários envolvidos na disseminação de rumores (BAO *et al.*, 2013) (COLLARD *et al.*, 2015), categorização do comportamento dos usuários disseminadores de rumor (MADDOCK *et al.*, 2015) (PROCTER; VIS; VOSS, 2013) e detecção de rumores (KWON; CHA; JUNG, 2017) (ZHANG *et al.*, 2015). Esses e outros trabalhos são descritos no Capítulo 3.

Conforme pode ser visto, existem trabalhos que buscam estudar rumores por meio de diferentes abordagens. Entretanto, não existe nenhum trabalho que compara as redes sociais no contexto de disseminação de rumores. Ainda, não existem trabalhos que apresentam as semelhanças e diferenças entre rumores de veracidades diferentes. Além disso, a literatura carece de uma abordagem para a predição da crença de um usuário em relação a um rumor. Esses dois aspectos da disseminação de rumores podem servir futuramente para inspirar estratégias de tratamento de rumores.

1.2 Objetivo e Contribuições

A partir do contexto introduzido acima, o objetivo deste trabalho de mestrado foi analisar, com o uso de visualizações, rumores disseminados no *Reddit* e no *Twitter*, a fim de identificar quais são os aspectos similares de ambas as redes sociais e quais os pontos em que uma diverge da outra. Ainda, foi realizada uma análise visual com o intuito de verificar as semelhanças e diferenças de um rumor verdadeiro e um falso. Ambas as análises foram realizadas com o uso do sistema de visualização *RumourFlow* (cf. Seção 4.2). Este sistema foi desenvolvido pela orientadora deste trabalho em colaboração com pesquisadores colaboradores. O *RumourFlow* utiliza técnicas de mineração de dados, visualizações, teorias de ciências sociais, análise de sentimentos e modelos de difusão de informação para permitir a análise de rumores.

Além das análises visuais, uma classificação supervisionada foi desenvolvida e avaliada. Esta classificação teve como objetivo predizer se um usuário acredita ou não no rumor que ele está disseminando. Para tal fim, parte do conjunto de dados coletado foi manualmente anotado e classificado por seis classificadores.

Em resumo, as contribuições desta pesquisa são:

- Análise visual das semelhanças e diferenças de rumores disseminados no *Reddit* e no *Twitter* e identificação dos principais pontos em que as redes sociais diferem uma da outra no contexto de disseminação de rumores;
- Comparação visual entre rumores de diferentes veracidades e elucidação das características dos rumores analisados;
- Definição, implementação e avaliação de uma abordagem para a classificação da crença que os usuários possuem em relação a rumores propagados em redes sociais.

O trabalho aqui descrito foi desenvolvido em colaboração com os pesquisadores Prof. Dr. Evangelos Milios, Prof. Dr. Abidalrahman Moh'd e Anh Dang do laboratório *Machine Learning and Networked Information Spaces* (MALNIS), da Faculdade de Ciência da Computação da *Dalhousie University*, Halifax, Canadá. Esta colaboração resultou na submissão do seguinte artigo:

- DANG, A.; SANTOS, N. R.; MOH'D, A.; MILIOS, E. E.; MINGHIM, R. RumorFlow: A Visual Analysis Framework for Studying Rumor Spread in Digital Social Media. **Information Visualization. (em revisão)**

1.3 Organização do Texto

Esta dissertação está organizada da seguinte maneira:

- No Capítulo 2 são apresentados conceitos de Aprendizado de Máquina e de Análise de Redes Sociais que são essenciais para a compreensão do trabalho apresentado nesta dissertação.
- No Capítulo 3 são apresentados os trabalhos relacionados.
- No Capítulo 4 são descritas as análises visuais realizadas.
- No Capítulo 5 é apresentada a abordagem proposta para a detecção da crença dos usuários envolvidos na disseminação de rumores, assim como os resultados obtidos.
- Por fim, a conclusão desse trabalho de mestrado é apresentada no Capítulo 6.

CAPÍTULO

2

CONCEITOS FUNDAMENTAIS

Neste capítulo são apresentados conceitos de Aprendizado de Máquina e Análise de Redes Sociais que foram essenciais no desenvolvimento deste trabalho.

2.1 Aprendizado de Máquina

Nesta seção são descritos conceitos de Aprendizado de Máquina utilizados na classificação supervisionada realizada neste trabalho. Vale resaltar que aprenas uma breve descrição dos conceitos será apresentada. Maiores explicações sobre cada um dos conceitos podem ser encontradas nas referências.

2.1.1 Máquina de Vetores de Suporte (SVM)

O SVM é um algoritmo supervisionado originalmente desenvolvido para classificação binária. Entretanto, ele foi estendido para também ser utilizado em tarefas de regressão e classificação multiclasse (MURPHY, 2012). A estensão feita para a classificação multiclasse consiste na utilização da abordagem *one-vs-rest*, onde o número de classificadores treinados é o número de classes existentes, de forma que a classe analisada é tratada como a classe *positivo* e as outras classes são tratadas como *negativo*. Com isso, a classe final é obtida por meio da Equação 2.1.

$$y(x) = argmax_c f_c(x) \tag{2.1}$$

Este algoritmo usa vetores de suporte para o cálculo do hiperplano de corte que divide o espaço de atributos em subespaços. Existem dois tipos de corte aceitos pelo SVM: linear, cujo tempo de treinamento é $O(N)$ (JOACHIMS, 2006), e polinomial, cujo tempo mínimo de treinamento é $O(N^2)$ por meio da utilização do algoritmo *Sequential Minimal Optimization* (SMO), que foi utilizado neste trabalho (PLATT, 1999).

2.1.2 Naive Bayes

O *Naive Bayes* é um algoritmo de classificação probabilístico que emprega o Teorema de Bayes em tarefas de classificação. Este algoritmo tem como característica a baixa ocorrência de sobreajuste, pois ele possui apenas $O(CD)$ parâmetros, onde C é a quantidade de classes e D representa a quantidade de atributos (MURPHY, 2012). Sobreajuste é o termo usado quando um classificador se ajusta a um conjunto de dados e produ bons resultados, porém quando ele é utilizado para classificar outros conjuntos de dados, ele tem um baixo desempenho.

O Teorema de Bayes é usado para calcular a probabilidade de um evento ocorrer dado que outro evento já ocorreu. Este teorema é calculado por meio da Equação 2.2, onde P(B|A) é a probabilidade de B ocorrer dado que A aconteceu, P(A) é a probabilidade de A ocorrer e B é a probabilidade de B ocorrer.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \tag{2.2}$$

2.1.3 K-vizinhos mais próximos

O *K-vizinhos mais próximos* (KNN) é um algoritmo de classificação supervisionado não paramétrico que pode ser utilizado tanto para tarefas de classificação, quanto para tarefas de regressão. Para classificar um novo elemento, os K vizinhos mais próximos dele são calculados por meio de uma função de distância e a classe majoritária entre esses vizinhos é atribuída a ele.

2.1.4 Árvores de Decisão

Árvores de Decisão são ferramentas de suporte de decisão comumente utilizadas em aprendizado de máquina. Na classificação baseada em árvores de decisão utiliza-se uma árvore para definir a classe de um novo elemento que está sendo classificado, de forma que cada nó interno dessa árvore representa um teste feito em um dos atributos de entrada, cada ramo representa o resultado de um teste e cada folha representa uma classe, que é obtida após todos os testes dos nós internos terem sido computados. As regras de classificação são dadas pelos possíveis caminhos entre a raiz e uma folha da árvore.

2.1.5 Comitê de Classificadores

Comitês de Classificadores têm como objetivo combinar as predições feitas por diferentes estimadores para melhorar a robustes de um estimador e, consequentemente, o resultado de uma classificação. Neste trabalho foram utilizados os comitês Floresta Aleatória, Votação Majoritária e Votação Majoritária Ponderada.

2.1.5.1 Floresta Aleatória

O método Floresta Aleatória, proposto por (BREIMAN, 2001), é um comitê que utiliza árvores de decisão como estimadores. Neste método, cada árvore de decisão é construída com uma amostra aleatória do conjunto de dados e a classe final é dada pela junção das classes preditas por cada árvore de decisão. Este método ajuda a reduzir a variância do estimador, visto que a floresta é construída por meio de uma randomização.

2.1.5.2 Votação Majoritária

O comitê de classificadores Votação Majoritária tem como intuito combinar diferentes classificadores e, por meio de um sistema de votos, predizer a classe de cada elemento sendo classificado. Este método está disponível nas versões simples e ponderada. Na versão simples, a classe atribuída a um elemento é a classe majoritária entre as classes preditas pelos diferentes classificadores. Por exemplo, suponha que os classificadores A, B e C sejam utilizados e as predições feitas por eles sejam:

- Classificador A: classe 2
- Classificador B: classe 1
- Classificador C: classe 2

Como a classe majoritária é a classe 2, ela será atribuída ao elemento sendo classificado. No caso de empate, a classe atribuída é a primeira na ordem crescente das classes. Por exemplo, se no exemplo anterior fosse adicionado o classificador D e a predição dele fosse a classe 1, haveria um empate e, como a classe 1 é a primeira na ordem crescente das classes, ela seria atribuída ao elemento sendo classificado.

Já a versão ponderada atribui a classe cuja média ponderada das probabilidades preditas é a maior. Por exemplo, suponha que tenhamos os classificadores A, B e C, as classes $1, 2$ e 3 e os pesos $P1 = 1$, $P2 = 1$ e $P3 = 1$. Um exemplo do cálculo da classe de um novo elemento pode ser visto na Tabela 1.

Tabela 1 – Exemplo do cálculo feito por um comitê de Votação Majoritária ponderada para definir a classe de um novo elemento.

Classificador	**Classe 1**	**Classe 2**	**Classe 3**
Classificador A	$P1 \cdot 0.4$	$P1 \cdot 0.5$	$P1 \cdot 0.5$
Classificador B	$P2 \cdot 0.7$	$P2 \cdot 0.8$	$P2 \cdot 0.2$
Classificador C	$P3 \cdot 0.3$	$P3 \cdot 0.2$	$P3 \cdot 0.4$
Média Ponderada	0.47	0.5	0.37

Fonte: Elaborada pelo autor.

A classe 2 será atribuída ao novo elemento, pois ela possui a maior média ponderada. No caso de empate, aplica-se a mesma estratégia executada na versão simples.

2.1.6 Medidas de Avaliação

Nesta subseção são definidas as medidas utilizadas na avaliação do desempenho dos classificadores utilizados neste trabalho. Para entender as equações apresentadas abaixo, é necessário conhecer os seguintes termos:

- **Verdadeiro Positivo (VP):** exemplo positivo classificado como positivo.
- **Verdadeiro Negativo (VN):** exemplo negativo classificado como negativo.
- **Falso Positivo (FP):** exemplo negativo classificado como positivo.
- **Falso Negativo (FN):** exemplo positivo classificado como negativo.

2.1.6.1 Acurácia

A acurácia é utilizada para medir a porcentagem de acerto de um classificador. Esta medida é calculada por meio da Equação 2.3.

$$Acurácia = \frac{VP+VN}{VP+VN+FP+FN} \tag{2.3}$$

2.1.6.2 Precisão

A Precisão, dada pela Equação 2.4, mede a quantidade de elementos positivos que, de fato, são positivos.

$$Precisão = \frac{VP}{VP+FP} \tag{2.4}$$

2.1.6.3 Revocação

A Revocação é a porcentagem de elementos positivos que foram classificados como positivo. O cálculo da revocação é feito por meio da Equação 2.5.

$$Revocação = \frac{VP}{VP+FN} \tag{2.5}$$

2.1.6.4 F1

A F1 é uma média harmônica da Precisão e da Revocação indicada para a avaliação de um conjunto de dados desbalanceado. Esta medida é calculada por meio da Equação 2.6.

$$F1 = 2 \times \frac{Precisão \times Revocação}{Precisão + Revocação} \tag{2.6}$$

Tabela 2 – Exemplo de uma Matriz Documento-Termo com m documentos e n termos.

	t_1	t_2	t_3	...	t_n
d_1	a_{11}	a_{12}	a_{13}	...	a_{1n}
d_2	a_{21}	a_{22}	a_{23}	...	a_{2n}
...	...	...	...	...	...
d_m	a_{m1}	a_{m2}	a_{m3}	...	a_{mn}

Fonte: Elaborada pelo autor.

2.1.7 Medidas de Concordância

Nesta subseção são definidas as medidas que foram utilizadas para mensurar a concordância entre os anotadores responsáveis por anotar o conjunto de dados usado na classificação desenvolvida neste trabalho de mestrado.

2.1.7.1 Krippendorff's Alpha (α)

O *Krippendorff's Alpha* é um coeficiente de confiabilidade criado para medir concordância entre anotadores (KRIPPENDORFF, 2004). Este coeficiente é calculado por meio da Equação 2.7, onde D_o é a discordância entre os anotadores e D_e é a chance da discordância esperada. A versão nominal deste coeficiente foi utilizada neste trabalho de mestrado.

$$\alpha = 1 - \frac{D_o}{D_e} \tag{2.7}$$

2.1.7.2 Fleiss' Kappa (κ)

O *Fleiss' Kappa* é um medida utilizada para medir a concordância entre anotadores (FLEISS; COHEN, 1973). Esta medida é descrita pela Equação 2.8, onde $\bar{P}$ é a concordância entre os anotadores e $\bar{P}_e$ é a probabilidade de chance de concordância.

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e} \tag{2.8}$$

2.1.8 Bag-of-words e TF-IDF

O *Bag-of-words* é um modelo de representação textual comumente utilizado em trabalhos de classificação, análise de sentimentos e recuperação de informação. A estrutura utilizada para representar este modelo é a Matriz Documento-Termo (SALTON; MCGILL, 1986), que é uma matriz $m \times n$, onde n é a quantidade de termos presentes em um conjunto textual, m é a quantidade de documentos deste conjunto e o valor de cada célula a_{ij} ($1 \leq i \leq m$ e $1 \leq j \leq n$) da matriz recebe o valor 1, se o termo t_j está presente em um documento d_i. Caso contrário, o valor da célula recebe o valor 0. Um exemplo desta matriz pode ser visto na Tabela 2.

Uma outra alternativa da abordagem binária, é a utilização do peso de cada termo ao invés da frequência dele na construção da matriz, desta forma diminui-se o impacto que termos muito frequentes exercem no resultado final. A medida Frequência do Termo-Inverso da Frequência dos Documentos (TF-IDF) é bastante utilizada em trabalhos da literatura para calcular o peso de cada termo (ROBERTSON, 2004). Esta medida é calculada por meio da Equação 2.9, onde $f_{t,d}$ é a frequência do termo t no documento d, $max_x f_{t_x,d}$ é a maior frequência de todos os termos t_x que aparecem no documento d, m é o número de documentos e m_t é o número de documentos em t está presente.

$$TF-IDF_{t,d} = \frac{f_{t,d}}{max_x f_{t_x,d}} \times \log \frac{m}{m_t} \tag{2.9}$$

2.1.9 Método de Trigramas do Google (Google Trigram Method - GTM)

O Método de Trigramas do Google é um método estatístico não supervisionado que calcula a similaridade entre dois textos por meio da similaridade de palavras utilizando tri-gramas do *Google* (ISLAM; MILIOS; KESELJ, 2012) (BRANTS; FRANZ, 2006).

A similaridade de palavras utilizando tri-gramas entre o par de palavras w_a e w_b é calculado por meio da Equação 2.10, onde C é a frequência máxima possível entre todos os uni-gramas do *Google*, $c(w)$ é a frequência de w nas uni-gramas do *Google* e $c(w_a w_i w_b)$ é a frequência da tri-grama $w_a w_i w_b$ nas tri-gramas do *Google*. Além disso, $\mu(w_a, n_1, w_b, n_2)$ é a frequência média das n_1 tri-gramas que começam com w_a e terminam com w_b e das n_2 tri-gramas que começam com w_b e terminam com w_a (Equação 2.11).

$$Sim(w_a, w_b) = \begin{cases} \frac{\log \frac{\mu(w_a,n_1,w_b,n_2)C^2}{c(w_a)c(w_b)min(c(w_a),c(b))}}{-2\times\log\frac{min(c(w_a),c(w_b))}{C}}, & \text{se } \frac{\mu(w_a,n_1,w_b,n_2)C^2}{c(w_a)c(w_b)min(c(w_a),c(b))} > 1 \\ \frac{\log 1.01}{-2\times\log\frac{min(c(w_a),c(w_b))}{C}}, & \text{se } \frac{\mu(w_a,n_1,w_b,n_2)C^2}{c(w_a)c(w_b)min(c(w_a),c(b))} \leq 1 \\ 0, & \text{se } \mu(w_a,n_1,w_b,n_2) = 0 \end{cases} \tag{2.10}$$

$$\frac{\sum_{i=1}^{n_1} c(w_a w_i w_b) + \sum_{i=1}^{n_2} c(w_b w_i w_a)}{2} \tag{2.11}$$

O cálculo da similaridade entre dois textos é dado pela Equação 2.12, onde os textos de entrada P e R possuem m e n palavras, respectivamente (P = $p_1, p_2, ..., p_m$, R = $r_1, r_2, ..., r_n$ e $n \geq m$). O somatório desta equação é obtido da seguinte forma:

1. Remove-se as δ palavras iguais presentes em P e R
2. Cria-se uma matriz de tamanho $(m-\delta)(n-\delta)$, de forma que o valor de cada elemento da matriz é a similaridade entre as palavras representadas pelo elemento.

- O cálculo da similaridade é feito por meio da Equação 2.10.

3. Calcula-se a média e o desvio padrão de cada linha da matriz.

4. Computa-se o conjunto de elementos cuja similaridade é maior que a soma da média e do desvio padrão da linha a qual os elementos pertencem.

5. Por fim, o somatório é obtido pela soma da média de todos conjuntos que satisfizeram a condição acima.

$$S(P,R) = \frac{(\delta + \sum_{i=1}^{m-\delta} \mu(A_i)) \times (m+n)}{2mn} \tag{2.12}$$

2.1.10 Vinculação de Entidades (Entity Linking)

A tarefa de vinculação de entidades consiste em ligar as entidades (i.e. palavras) presentes em um texto com a sua entrada em uma base de conhecimento (ex. Wikipedia), fornecendo um melhor entendimento de um corpus. Por exemplo, suponha que a base de conhecimentos utilizada seja o Wikipedia. Dado o texto "O Santos é o melhor time do mundo", se a palavra "Santos"for ligada com a sua entrada na base de conhecimentos (<https://pt.wikipedia.org/wiki/Santos_Futebol_Clube>), saberemos que o texto refere-se a um time de futebol brasileiro e, desta forma, teremos mais informações sobre o texto de entrada. A vinculação de entidades pode ser utilizada em diferentes tarefas, tais como o enriquecimento de textos, busca semântica e anotação semântica.

Neste trabalho, a ferramenta de vinculação de entidades *Dexter* (TRANI *et al.*, 2014) foi utilizada para a extração de tópicos a partir de um conjunto de dados de rumores. A base de conhecimento utilizada pela *Dexter* é o Wikipedia.

2.2 Análise de Redes Sociais

A análise de redes sociais é um processo que surgiu no século XX e tem como objetivo a identificação de padrões e o que eles implicam na composição de uma rede social (WASSERMAN; FAUST, 1994). Este processo foi criado a partir de esforços de pesquisadores da Psicologia e da Sociologia, porém atualmente pesquisadores de áreas como a Matemática, Estatística e Computação combinam esforços com as áreas supracitadas para produzir avanços científicos na análise.

A análise de redes sociais pode ser dividida da seguinte forma: **análise de usuários**, **análise de relações** e **análise de rede** (KERREN; PURCHASE; WARD, 2014). As análises de usuários e de rede foram empregadas na análise visual apresentada na Seção 4.2.

2.2.1 Análise de Usuários

A análise de usuários busca investigar os indivíduos, grupos de indivíduos ou corporações que compõem uma rede social. É possível, por meio desta análise, predizer a personalidade dos usuários de uma rede social (GOLBECK *et al.*, 2011). Além disso, pode-se identificar quais são as preferências pessoais de um indivíduo para utilizá-las em sistemas de recomendação (JAWAHEER; SZOMSZOR; KOSTKOVA, 2010).

2.2.2 Análise de Relações

A análise de relações tem como objetivo caracterizar as relações existentes entre usuários e predizer as propriedades dessas relações. Por exemplo, no sistema criado por (ZHOU *et al.*, 2012), um ranque é construído a partir das propriedades das relações de um usuário. As propriedades utilizadas são: a comunicação do usuário com os seus contatos, amigos em comum e os tópicos das discussões em que ambos participaram.

2.2.3 Análise de Rede

O objetivo desta análise é a identificação de padrões e características a partir da estrutura de uma rede social. Esta análise é dividida em dois tipos: **Análise Baseada em Grafos** e a **Análise Baseada em Sociogramas** (KERREN; PURCHASE; WARD, 2014).

- **Análise Baseada em Grafos**: esta análise pode ser realizada de diferentes formas, indo desde a análise de pares de usuários até a análise que emprega medidas de centralidade para identificar o nível de importância que um usuário exerce sobre a rede.
- **Análise Baseada em Sociogramas**: é a análise que utiliza o Diagram Nó-aresta (Figura 4). Nesta estratégia os nós representam os usuários e as arestas as relações entre os usuários. Esta análise auxilia na identificação de padrões estruturais da rede. Um problema comum para esta análise é o fato de que a grande quantidade de nós e arestas na rede podem geram uma confusão visual, dificultando a compreensão da rede em questão (KERREN; PURCHASE; WARD, 2014).

2.2.4 Técnicas, Estruturas e Métricas

Nesta subseção são apresentadas as técnicas, estruturas e métricas que são utilizadas na análise de redes sociais e que foram empregadas neste trabalho.

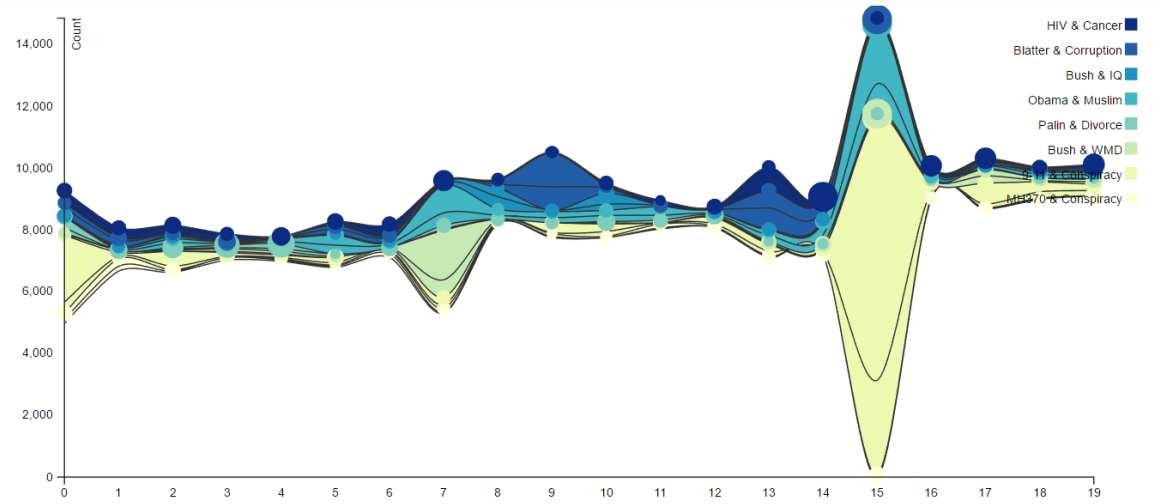
2.2.4.1 Técnicas para a Visualização Temporal

A técnica *ThemeRiver* é utilizada para visualizar mudanças temporais que uma coleção de documentos textuais sofre. Esta técnica utiliza a metáfora de um rio para apresentar os dados,

onde segmentos da coleção são mapeados nas camadas desse rio. O eixo horizontal desta técnica representa um intervalo de tempo e o eixo vertical representa alguma informação da coleção que varia com o tempo.

(DANG *et al.*, 2016) utilizaram a *ThemeRiver* para representar a evolução temporal de um conjunto de rumores (Figura 1). Cada rumor foi mapeado para uma corrente do rio e o eixo vertical foi utilizado para representar o número de usuários que estavam participando da discussão de um dos rumores em cada instante de tempo.

Figura 1 – Exemplo de uma ThemeRiver extraída do sistema RumourFlow (DANG *et al.*, 2016). Temos nesta *ThemeRiver* uma visão da temporalidade de um conjunto de rumores, onde cada corrente representa um dos rumores e o rio representa o conjunto de rumores. A altura de cada corrente representa a quantidade de usuários que estavam discutindo algo relacionado com aquele rumor no instante de tempo em questão.



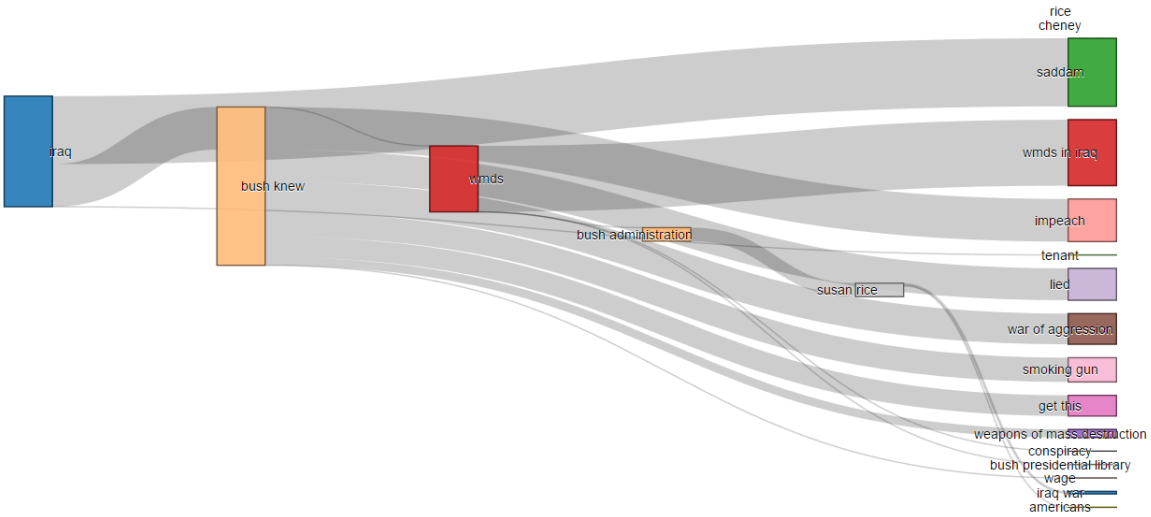
Fonte: Elaborada pelo autor.

O Diagrama de Sankey é um diagrama que tem como finalidade a representação de fluxos. A Figura 2 apresenta um Diagrama de Sankey criado para representar a evolução temporal dos tópicos mencionados em um rumor (DANG *et al.*, 2016). Cada barra horizontal deste diagrama representa um tópico e a altura desta barra indica a quantidade de usuários que estão discutindo sobre aquele tópico. Dada as postagens contínuas $P_1 = T_1 1, T_1 2, T_1 3, ..., T_1 m$ e $P_2 = T_2 1, T_2 2, T_2 3, ..., T_2 n$, uma conexão entre os tópicos $T_1 i$ da postagem P_1 e $T_2 j$ da postagem P_2 é criada se a similaridade semântica entre os tópicos ou a quantidade de usuários em comum entre as duas postagens for maior que um limiar.

2.2.4.2 Representações de Texto

Nuvens de Palavras são representações visuais utilizadas para apresentar as palavras que possuem maior importância em um conjunto de texto, fornecendo uma visão geral do que é discutido no conjunto. Nesta representação visual, quanto maior a importância da palavra, maior será o tamanho da fonte utilizada para representá-la na Nuvem de Palavras (SEIFERT *et al.*,

Figura 2 – Exemplo de um Diagrama de Sankey utilizado para representar a evolução temporal dos tópicos de um rumor. Pode-se observar que no início da discussão do rumor apenas o tópico "iraq"era mencionado, porém novos tópicos foram surgindo ao longo do tempo e no final do ciclo de vida do rumor 16 tópicos eram mencionados.



Fonte: Elaborada pelo autor.

2008). Uma das maneiras utilizadas para a geração das Nuvens de Palavras consiste em utilizar a frequência das palavras no conjunto de texto para representar a importância de cada palavra. Um exemplo de uma Nuvem de Palavras pode ser visto na Figura 3.

Figura 3 – Nuvem de Palavras dos tópicos extraídos do rumor *"9-11 & Conspiracies"*. É possível notar que as palavras *inside job* e *conspiracy theorist* são as palavras que possuem maior frequência no conjunto de textos sobre o rumor em questão, visto que o tamanho da fonte dos tópicos representa a frequência deles.

Fonte: Elaborada pelo autor.

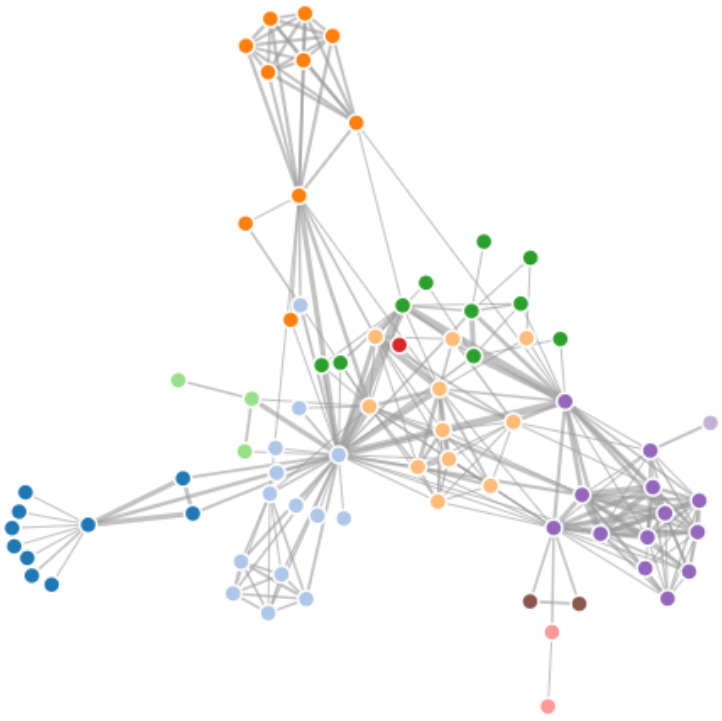
2.2.4.3 Grafo

Um grafo é uma estrutura empregada para representar redes e comumente utilizada em trabalhos de análise de redes sociais (DANG *et al.*, 2016) (LIU *et al.*, 2016). Um grafo G é composto por nós e arestas, onde E(G) representa as arestas e V(G) os nós. $|V|$ representa a quantidade de nós de um Grafo e $|E|$ representa a quantidade de arestas. Denota-se por $(u,v) \in E(G)$ a aresta que conecta os nós $u,v \in V(G)$.

Um grafo cujas arestas possuem direção é chamado de grafo dirigido. Já o grafo cujas arestas possuem peso é chamado de grafo ponderado. O número de arestas que incidem em um nó é chamado de grau do nó. $\delta(G)$ indica o grau mínimo do grafo e Δ o grau máximo.

Entre as diferentes formas de representar um grafo, a mais tradicional é o Diagrama Nó-aresta (LANDESBERGER *et al.*, 2011), que representa o grafo com a utilização de nós e arestas (Figura 4). Na análise de redes sociais, o nós do diagrama representam usuários e as arestas representam uma relação entre os usuários.

Figura 4 – Exemplo de um grafo desenhado na forma Nó-aresta.

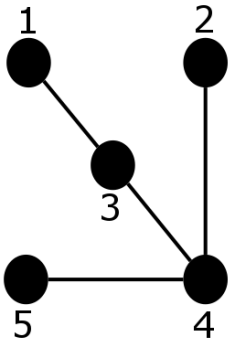


Fonte: Bostock (2017).

Grafos também podem representados por uma matriz de adjacência. Uma matriz de adjacências é uma matriz quadrada $v \times v, v \in V(G)$, onde cada linha e cada coluna representa um nó do grafo. A matriz é preenchida da seguinte forma: se existe uma aresta entre um nó u e um nó v, o elemento da matriz localizado na linha u e coluna v receberá o valor 1; caso contrário, ele receberá 0. Vale notar que se o grafo for ponderado, o valor atribuído ao elemento

que representa a relação receberá o peso da aresta (u,v). A Figura 5 apresenta um exemplo da matriz de adjacências.

Figura 5 – Exemplo de um grafo com 5 vértices e 4 arestas, e sua matriz de adjacências, onde cada célula da matriz é preenchida com 1, se os vértices representados pela linha e pela coluna da célula estão ligados por uma aresta. Caso contrário, a célula é preenchida com 0.



	1	2	3	4	5
1	0	0	1	0	0
2	0	0	0	1	0
3	1	0	0	1	0
4	0	1	1	0	1
5	0	0	0	1	0

Fonte: Elaborada pelo autor.

2.2.4.4 Métricas para Análise de Redes Sociais

Medidas de centralidade são avaliações de importância de vértices em um grafo. Elas são aplicadas na análise de redes sociais, pois elas permitem a identificação de usuários que possuem maior influência sobre o resto da rede (KERREN; PURCHASE; WARD, 2014). Neste trabalho foram utilizadas as centralidades de Intermediação (*betweenness*) e de Proximidade (*closeness*).

A Centralidade de Proximidade mede o quão próximo um nó está de todos os nós da rede. A Proximidade C de um usuário v_i pode ser obtida por meio da Equação 2.13. Esta equação nos fornece a inversa do somatório de cada distância d do nó v_i para todos os outros v_j nós da rede, onde $1 \leq j \leq V$ e V é a quantidade total de nós.

$$C(v_i) = [\sum_{j=1}^{V} d(v_i, v_j)]^{-1} \tag{2.13}$$

Pode-se observar que o valor máximo que pode ser obtido com a utilização da equação acima depende da quantidade total de nós (V), o que dificulta a comparação de medidas de Proximidade de nós que estão em redes distintas. Portanto, temos a normalização da Proximidade (Equação 2.14). Esta normalização consiste na multiplicação de $V-1$ por $\frac{1}{C(v_i)}$, onde V é a quantidade total de nós.

$$C(v_i) = \frac{V-1}{[\sum_{j=1}^{V} d(v_i, v_j)]^{-1}} \tag{2.14}$$

A Centralidade de Intermediação mede a quantidade de vezes que um nó v_i está no caminho mínimo existente entre outros dois nós. O cálculo se dá pela divisão da quantidade de

vezes que um nó v_i está nos caminhos mínimos entre os nós v_a e v_b pela quantidade de caminhos mínimos existentes entre esse par de nós (Equação 2.15).

$$B(v_i) = \sum_{i \neq a \neq b} \frac{\sigma_{ab}(v_i)}{\sigma_{ab}} \tag{2.15}$$

A Equação 2.15 também possui sua versão normalizada, para que seja possível comparar a Intermediação de dois nós presentes em redes diferentes. Em redes dirigidas, a normalização é obtida pela divisão da Equação 2.15 por $V-1 \times V-2$ e em redes não-dirigidas divide-se a Equação 2.15 por $\frac{V-1 \times V-2}{2}$.

2.3 Considerações Finais

Neste capítulo, foram apresentados os conceitos essenciais para o trabalho desenvolvido. Os métodos de classificação e o modelo *Bag-of-words* aqui apresentados são de suma importância para a etapa de classificação, assim como as medidas de avaliação, que foram empregadas para que a abordagem proposta nesta etapa fosse avaliada. O método GTM e a tarefa de Vinculação de Entidades foram necessários para o pré-processamento dos dados visualizados por meio da ferramenta de visualização de rumores *Rumour Flow*. Além disso, os conceitos de análise de redes sociais aqui apresentados foram extremamente importantes para as análises visuais realizadas por meio do *RumourFlow*.

CAPÍTULO

3

TRABALHOS RELACIONADOS

Neste Capítulo é apresentado um levantamento bibliográfico de trabalhos sobre visualização de redes sociais e análise de rumores propagados em redes sociais, os dois temas centrais à proposta deste trabalho de mestrado. No final deste capítulo algumas considerações sobre os trabalhos citados e suas relações com a proposta deste trabalho são apresentadas.

3.1 Visualização em Redes Sociais

Apesar dos esforços aplicados na análise de redes sociais, o progresso de trabalhos nessa área é dificultado pela falta de mecanismos para obter dados de interesse dos analistas a partir de mídias sociais. Ainda que trabalhos como o de (BOSCH *et al.*, 2013) consigam obter postagens relevantes, eles ainda apresentam problemas, de acordo com (LIU *et al.*, 2016), que afirma que trabalhos nessa área não combinam as três dimensões de uma rede social (postagens, usuários e *hashtags*). Além disso, os autores afirmam que os métodos apresentados até o momento não consideravam o que eles chamam de incerteza (*uncertainty*), que são inseridas quando os dados são coletados, transformados e analisados. Com isso, foi criado o *MutualRanker*, ferramenta cujo propósito é permitir a identificação de incertezas introduzidas pelos algoritmos de análise e também coletar dados de interesse (LIU *et al.*, 2016).

O *MutualRanker* extrai uma lista limitada de dados de uma rede social e os classifica de acordo com sua relevância. Em seguida, um grafo de reforço mútuo é gerado. Neste grafo, postagens, usuários e *hashtags* são conectadas por suas relações. Por exemplo, se um usuário é autor de uma *hashtag*, haverá uma ligação entre este usuário e a *hashtag* em questão. Um analista, ao interagir com o sistema, pode modificar o *score* de um item. Caso isso ocorra, o sistema irá atualizar incrementalmente a classificação dos itens, visto que um item pode ter influência sobre outros. Um modelo de incerteza é utilizado para calcular o valor de incerteza de cada item (postagem, usuário ou *hashtag*), além de calcular a propagação dessa incerteza. Por fim, as informações são apresentadas ao usuário por visões que apresentam o grafo de reforço mútuo,

um glifo indicando a incerteza do item e um glifo para representar a mudança que ocorreu com um item, caso um analista tenha mudado seu ranque.

Alguns trabalhos na análise de redes sociais envolvem a visualização de coautorias. Um deles é o trabalho de (RADVANSKý *et al.*, 2013), que afirmam que visualizar as dinâmicas desse tipo de rede é útil, pois a evolução dela e de suas comunidades podem apontar para informações importantes que podem ser extraídas por um analista. Os pesquisadores desenvolveram uma abordagem para a apresentação de redes dinâmicas que é baseada na projeção de Sammon, que é uma técnica de projeção multidimensional, e em aproximações lineares. Para a análise de coautorias, este trabalho e o trabalho de (ALSUKHNI; ZHU, 2012), que será apresentado a seguir, utilizaram a Digital Bibliography and Library Project (DBLP[1]), que é um repositório bibliográfico de Ciência da Computação.

Seguindo o mesmo contexto, (ALSUKHNI; ZHU, 2012) apresentaram uma abordagem para a melhoria da técnica de projeção multidimensional, onde neste caso a projeção multidimensional não métrica foi utilizada. A melhoria apresentada pelos autores consiste na utilização de uma estrutura onde uma lista ligada foi empregada para guardar a estrutura do grafo. Os nós nesta lista são conectados por ponteiros, de modo que cada nó tenha uma referência para o próximo nó ao qual ele está conectado. Além disso, o grafo foi dividido em células a fim de facilitar a computação da força de repulsão. Cada célula, neste caso, possui um número limitado de nós por lista.

Além da projeção multidimensional, foi criado um grado de coautorias. O grafo de coautorias computado sobre o DBLP é criado de forma que os nós representam os autores e as arestas conectam autores que trabalharam em um mesmo artigo. Para o cálculo da distância dos nós, aplicou-se a distância Euclidiana. Além disso, para apresentar a rede em sua melhor configuração, o estresse entre os nós foi minimizado por meio de métodos de otimização.

(ZHAO *et al.*, 2014) criaram um sistema interativo de visualização para analisar a disseminação de informações anômalas no Twitter. Em um primeiro momento, os dados são coletados por meio da utilização de algoritmos de aprendizado de máquina. Após a coleta dos dados, o módulo de visualização permite, por meio de diferentes visões, a análise das informações identificadas.

Uma das visões consiste no agrupamento hierárquico dos tópicos de cada assunto coletado de acordo com as características presentes em cada tweet. Além disso, é possível visualizar, por meio de uma projeção multidimensional, a distribuição dos assuntos no espaço de características, tornando possível a identificação de *outliers* e a comparação dos assuntos. Por fim, ao selecionar um dos nós que representam um assunto em uma das visões supracitadas, uma visão desse nó é carregada como um glifo, onde é possível analisar a evolução temporal do assunto. Além disso, essa visão possui três módulos que permitem ao usuário explorar os dados em baixo

[1] <www.dblp.uni-trier.de>

nível por meio da visualização da evolução do vetor de características, dos tweets enviados e da transformação da projeção multidimensional ao longo do tempo. A Figura 6 exibe uma visão geral do FluxFlow.

Figura 6 – Visão geral do FluxFlow, onde é possível ver as visualizações supracitadas.

Fonte: Zhao *et al.* (2014).

Em Março de 2011, ocorreu um terremoto no Japão que deixaram as pessoas em alerta. Nessa época, os japoneses buscaram informações sobre a situação do país em diversos meios de comunicação, como a internet. Entretanto, informações cuja veracidade eram duvidosas também estavam sendo enviadas nas redes sociais e, com isso, o comportamento da sociedade do Japão foi influenciado por isso. Com isso em mente, (HASHIMOTO; KUBOYAMA; SHIROTA, 2011) criaram um *framework* para a detecção e visualização de possíveis rumores, onde, após a análise do assunto suspeito, informações sobre esse rumor são buscadas em outros meios de comunicação a fim de confirmar sua veracidade.

Após a coleta dos dados, o *framework* extrai palavras-chave das publicações e calcula seu *score*, com a utilização de medidas utilizadas na mineração de dados. Em seguida, um grafo é construído a partir da relação de coocorrência das palavras-chave obtidas. Nesse grafo, uma palavra com maior frequência nos documentos é posicionada em níveis superiores e, nos níveis abaixo dela, palavras relacionadas são inseridas e conectadas a ela, de acordo com sua frequência no texto.

3.2 Rumores em Redes Sociais

Nesta Seção são apresentados trabalhos que investigaram a propagação de rumores em redes sociais.

3.2.1 Tipos de Rumor

(ZUBIAGA *et al.*, 2016) classifica os rumores de acordo com sua veracidade, que pode ser *verdade*, *falsa* ou *não resolvido*, onde o rumor *não resolvido* é aquele que até o momento de sua coleta não teve sua veracidade confirmada ou negada. Este estudo ainda mostra que rumores verdadeiros tem sua veracidade comprovada mais rápida do que rumores falsos. Além disso, é apresentado que, por mais que portais de notícias se esforçam para trazer informações fundamentadas, elas acabam publicando informações que não foram verificadas, dando surgimento a novos rumores.

A temporalidade de um rumor é outro fator utilizado na categorização de rumores. No trabalho de (BOYANDIN *et al.*, 2011) são apresentadas duas categorias de rumor:

- Rumores que surgem a partir de notícias de última hora: são rumores que surgem a partir da postagem de informações de última hora em portais de notícias.
- Rumores de longa data: são rumores antigos cuja veracidade não foi verificada, por mais que existam esforços para tal verificação.

3.2.2 Modelagem e Características de Rumores

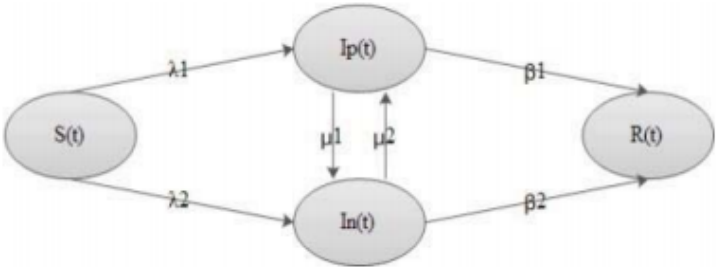
Existem, na literatura, trabalhos que utilizam modelos de difusão de informação para poder identificar rumores. Por exemplo, o trabalho de (BAO *et al.*, 2013) apresentou o modelo SPNR. Este modelo divide os usuários em três estados, que são:

1. S - representa o estado do usuário que ainda não teve contato com o rumor
2. I - estado que representa o usuário que dissemina o rumor, também chamado de usuário infectado. Este estado é dividido nos seguintes subestados:
   - I_p - subestado que representa o usuário que dissemina e apoia o rumor.
   - I_n - subestado que representa o usuário que dissemina e se opõe ao rumor.
3. R - é o estado que representa o usuário que teve contato, porém não acredita mais no rumor.

Além disso, este modelo apresenta algumas propriedades que são enumeradas abaixo e exibidas na Figura 7.

- Um usuário S passará para o estado I se ele tiver contato com um usuário I. A probabilidade dessa transformação ocorrer é dada por γ_1/γ_2.
- Um usuário I_p passará para o estado I_n com a probabilidade μ_1, porém se este usuário tiver contato com um usuário R, ele passará para o estado R com a probabilidade β_1.
- Um usuário I_n passará para o estado I_p com a probabilidade μ_2, porém se este usuário tiver contato com um usuário R, ele passará para o estado R com a probabilidade β_2.

Figura 7 – Propriedades do modelo apresentado SPNR.



Fonte: Bao *et al.* (2013).

(COLLARD *et al.*, 2015) apresentaram um modelo em que um usuário consegue propagar o rumor somente para usuários que estão conectados a ele e afirmam que um rumor possui quatro características. (TRIPATHY; BAGCHI; MEHTA, 2010) apresentaram dois modelos para o controle de um rumor, onde um deles, o *Delayed Start Model* assume que um usuário, chamado de Autoridade, irá propagar na rede um anti-rumor na tentativa de controlar a disseminação deste. Este anti-rumor será buscado em fontes de informação confiáveis antes de ser propagado, o que, segundo os autores, leva um certo tempo, possibilitando que o rumor se propague na rede. Já o outro modelo mantém agentes na rede que, a partir do momento que um rumor é detectado, eles começam a disseminar, automaticamente, anti-rumores.

3.2.3 Usuários envolvidos em Rumores

Quando em contato com uma informação, as pessoas tendem a reagir e expressar a opinião deles sobre o que a informação apresentou a eles. Desta forma, pesquisadores também buscam identificar essas pessoas, além de entender e categorizar o comportamento delas quando possuem contato com rumores.

(COLLARD *et al.*, 2015) apresentaram dois modelos que descrevem o comportamento dos usuários na disseminação do rumor, que são os modelos *ODS Profusion* e *ODS Scarcity*. A sigla ODS representa os três estados de um usuário dentro de um rumor. São eles: mente aberta (O), usuários que ainda não tem conhecimento do rumor, disseminadores (D), que são os responsáveis por propagar o rumor, e os stiflers (S), usuários que tem conhecimento do rumor, mas não o dissemina mais. O *ODS Profusion* tem a premissa de que usuários mente aberta

passarão a transmitir o rumor, isto é, passarão a ser disseminadores, se a maioria de seus vizinhos forem disseminadores. Já no *ODS Scarcity*, usuários mente aberta passarão a transmitir o rumor se uma pequena porção dos seus vizinhos forem disseminadores.

O trabalho apresentado por (PROCTER; VIS; VOSS, 2013) categorizou a reação de usuários do Twitter em relação a rumores disseminados durante os protestos que ocorreram na Inglaterra em 2011. Os autores deste artigo categorizaram as reações dos usuários em usuários que acreditam no rumor, usuários que não acreditam não no rumor, usuários que pedem mais informações antes de dizer algo e usuários que apenas deixaram algum comentário sobre o rumor, mas não expressaram se acreditam ou não no que é dito.

(MADDOCK *et al.*, 2015) analisou quantitativamente e qualitativamente quatro rumores que foram disseminados no Twitter após a explosão de bombas na maratona de Boston. Neste trabalho, os autores apresentam a origem, evolução temporal e relações entre diferentes tipos de comportamento dos usuários em relação ao rumor.

Análises qualitativas e estatísticas foram conduzidas em rumores da rede social chinesa *Sina Weibo*[2] (LIAO; SHI, 2013). As reações aos rumores foram dividas em sete categorias, onde usuários forneciam informações, opinavam, demonstravam afetividade, diziam algo com fundamentação, questionavam, davam uma diretiva ou eram digressivos sobre o rumor. Além disso, os autores categorizaram as pessoas que estavam envolvidas em: celebridades, portais de notícia, responsáveis por pequenas empresas, empresas, websites, estrelas da internet e pessoas comuns.

O estudo realizado por (MENDOZA; POBLETE; CASTILLO, 2010) identificou que a quantidade de pessoas que negam um rumor que é falso é superior à quantidade de pessoas que apoiam rumores verdadeiros. Já (ZUBIAGA *et al.*, 2016) descobriram que usuários de redes sociais tendem a apoiar e disseminar um rumor, independente da sua veracidade. Os dois trabalhos diferem no fato de que (MENDOZA; POBLETE; CASTILLO, 2010) avaliaram todo o ciclo de vida do rumor, enquanto (ZUBIAGA *et al.*, 2016) analisaram as reações iniciais ao rumor.

(CHENG *et al.*, 2013) descobriram, por meio da utilização de um modelo estocástico, que a difusão de rumores depende da força das relações entre os usuários e que rumores tendem a ser disseminados por relações fortes. Segundo (PETTY; CACIOPPO, 1986), a disseminação de um rumor também depende do que é dito por ele e também pela credibilidade de pessoa que o está divulgando. Análises no Twitter identificaram que rumores publicados por usuários com uma larga quantidade de usuários eram os mais disseminados na rede (CHUA *et al.*, 2016). Existem períodos no ciclo de vida de um rumor em que ele deixa de receber a atenção dos usuários que o estavam disseminando, conforme afirmado por (LUKASIK; COHN; BONTCHEVA, 2015).

[2] <www.weibo.com>

3.3 Considerações Finais

É possível apresentar algumas colocações sobre os trabalhos apresentados neste capítulo. De modo geral, os trabalhos apresentados na Seção 3.1 apresentam análises de diferentes tipos de redes sociais, permitindo identificar que a utilização de grafos é a mais comum entre os trabalhos. Entretanto, pouco é explorado da temporalidade das redes sociais.

Sobre os trabalhos apresentados na Seção 3.2, podemos notar que poucos analisam todo o ciclo de vida de um rumor. Além disso, os trabalhos apresentados na Subseção 3.2.3 apresentam pouco sobre os tópicos que levaram o usuário a ter determinada reação e pouco há sobre a categorização de usuários. Um ponto deixado em aberto é o motivo pelo qual usuários deixam de comentar sobre um rumor, levando este rumor a sair das discussões de redes sociais ou até desaparecer.

Trabalhos que estudam usuários buscam identificar os tipos de usuários, quais as reações deles quando diante de um rumor, entre outros. Entretanto, conforme dito por (CHENG *et al.*, 2013), a difusão de rumores depende da força das relações entre os usuários, porém trabalhos que estudam as relações entre os usuários são pouco encontrados na literatura. Além disso, faltam trabalhos que associam os usuários com a evolução do rumor e os tópicos discutidos no rumor.

Os artigos apresentados na Seção 3.2 envolvem, em sua maioria, a utilização de técnicas de Mineração de Dados, porém pouco é utilizado da área de Visualização de Dados. Isto deixa em aberto uma lacuna, visto que é possível, com visualizações, identificar informações que não são vistas somente com a aplicação de técnicas de Mineração de Dados. Portanto, este ponto está no escopo deste trabalho, onde iremos utilizar técnicas de Visualização de Dados e também de Mineração de dados para a exploração de rumores disseminados no *Reddit* e no *Twitter*.

CAPÍTULO

4

ANÁLISE VISUAL DE RUMORES

Neste capítulo são apresentadas as análises visuais realizadas neste trabalho de mestrado. Inicialmente são apresentadas as etapas de coleta e pré-processamento dos dados do Reddit e do Twitter que foram utilizados neste trabalho. Em seguida, são apresentadas as visualizações do sistema de visualização *RumourFlow* e as análises visuais que foram feitas por meio dele.

4.1 Seleção e Coleta dos Rumores

De modo a realizar as análises e a classificação apresentadas neste trabalho, os *sites* *Snopes*[1] e *Politifact*[2] foram utilizados para a seleção dos rumores que foram coletados do *Reddit* e do *Twitter*. O *Snopes* (especializado em rumores gerais) e o *Politifact* (especializado em rumores políticos) são *sites* dedicados a verificar a veracidade de informações que circulam pela mídia e são comumente utilizados em estudos de rumores (LIU *et al.*, 2015) (MA *et al.*, 2016) (DANG *et al.*, 2019).

Ao final da etapa de seleção dos rumores, os seguintes rumores foram escolhidos:

- **Trump & Tax Cut**: surgiu após o presidente dos Estados Unidos, Donald Trump, afirmar que a sua proposta para redução de impostos era a maior da história do país;
- **Cholera & Puerto Rico**: surgiu após Paul Krugman, colunista do *The New York Times*[3], postar em sua conta do Twitter que existiam casos de cólera em Porto Rico após a passagem do furacão Maria pela ilha;
- **Papers & Global Warming**: surgiu quando o colunista James Delingpole publicou um artigo no *Breitbart*[4] afirmando que quatrocentos artigos científicos publicados em 2017

[1] <www.snopes.com>
[2] <www.politifact.com>
[3] <https://www.nytimes.com/>
[4] <https://www.breitbart.com/>

provam que o aquecimento global é um mito;

- **Sandra Bullock & Hillary Clinton**: surgiu após diversos *sites* publicarem que a atriz Sandra Bullock destratou Hillary Clinton em defesa do presidente norte-americano Donald Trump;

- **Sex Offenders & Uber**: surgiu após o senador norte-americano, Thomas Croci, afirmar que havia uma brecha em uma lei do estado de Nova York que permitia que alguns agressores sexuais trabalhassem para serviços como o *Uber* e o *Lyft*;

- **Hydrogen Peroxide & Cancer**: surgiu em 2006 e as pessoas que o discutem afirmam que o Peróxido de Hidrogênio pode curar o câncer.

- **9-11 & Conspiracies**: são rumores e conspirações sobre os quatro ataques terroristas sofridos pelos Estados Unidos em 2001.

Após a seleção dos rumores, um *crawler*[5] foi utilizado para a coleta de postagens do Twitter que estavam relacionados aos rumores selecionados. O *crawler* usado foi desenvolvido com a utilização da linguagem de programação Java e a biblioteca de código aberto twitter4j[6], que facilita a conexão do usuário com as APIs da rede social.

Entre as APIs disponibilizadas pelo Twitter, as de busca e de *streaming* foram utilizadas. A API de busca retorna tweets antigos que correspondem a uma consulta criada pelo usuário, enquanto que a API de *streaming* retorna tweets criados em tempo real e que também correspondem a uma consulta criada pelo usuário. Vale ressaltar que a API de busca está disponível na versão padrão (*standard*), que permite o acesso aos tweets publicados nos últimos sete dias, e na versão paga (*premium*), que permite o acesso a todos tweets existentes.

Inicialmente a API de busca foi utilizada para coletar os tweets criados entre os dias 1 e 7 de Outubro de 2017. Em seguida, a API de *streaming* foi utilizada para a coleta dos tweets criados entre os dias 8 e 15 de Outubro de 2017, durante duas horas nos períodos matutino, vespertino e noturno. Um resumo dos tweets coletados pode ser visto na Tabela 3.

Tabela 3 – Conjunto de dados extraídos do Twitter.

Rumor	**Tweets**	**Usuários**	**Início**	**Status**
Trump & Tax Cut	3621	1916	2017	Negado em 2017
Cholera & Puerto Rico	1458	658	2017	Negado em 2017
Papers & Global Warming	2345	1635	2017	Negado em 2017
Sandra Bullock & Hillary Clinton	1114	665	2017	Negado em 2017

Fonte: Elaborada pelo autor.

[5] Crawler é um programa que extrai informações de um *site* de maneira automática.
[6] <http://twitter4j.org/en/index.html>

As postagens do Reddit foram coletadas por um dos pesquisadores colaboradores. Para a coleta das postagens, foi criado um *crawler* com a utilização da linguagem de programação *Java* e o *wrapper*[7] *jReddit*. A Tabela 4 apresenta um resumo das postagens coletadas. Vale ressaltar que uma das vantagens do Reddit em relação ao Twitter é que o Reddit emprega uma política de dados abertos, permitindo, desta maneira, que os usuários coletem qualquer postagem criada na rede social.

Tabela 4 – Conjunto de dados extraídos do Reddit.

Rumor	**Postagens**	**Usuários**	**Início**	**Status**
Sex Offenders & Uber	402	316	2017	Confirmado em 2017
Hydrogen Peroxide & Cancer	684	371	2006	Negado em 2013
Trump & Tax Cut	360	352	2017	Negado em 2017
Cholera & Puerto Rico	313	327	2017	Negado em 2017
Papers & Global Warming	182	158	2017	Negado em 2017
Sandra Bullock & Hillary Clinton	396	349	2017	Negado em 2017
9-11 & Conspiracies	19793	5190	2001	Inverificável

Fonte: Elaborada pelo autor.

4.2 *RumourFlow*

O sistema de visualização *RumourFlow* (DANG *et al.*, 2016) foi utilizado como sistema base para o desenvolvimento deste trabalho. Utilizando este sistema, duas análises visuais foram realizadas. A primeira delas consistiu em analisar os rumores "*Trump & Tax Cut*", "*Cholera & Puerto Rico*", "*Papers & Global Warming*"e "*Sandra Bullock & Hillary Clinton*"disseminados no *Reddit* e no *Twitter*, com o intuito de identificar semelhanças e diferenças na disseminação de rumores em redes sociais diferentes. Já a segunda análise consistiu em analisar os rumores "*Sex Offenders & Uber*"e "*Hydrogen Peroxide & Cancer*"disseminados no *Reddit*, a fim de identificar as semelhanças e diferenças entre a disseminação de um rumor verdadeiro e a de um rumor falso.

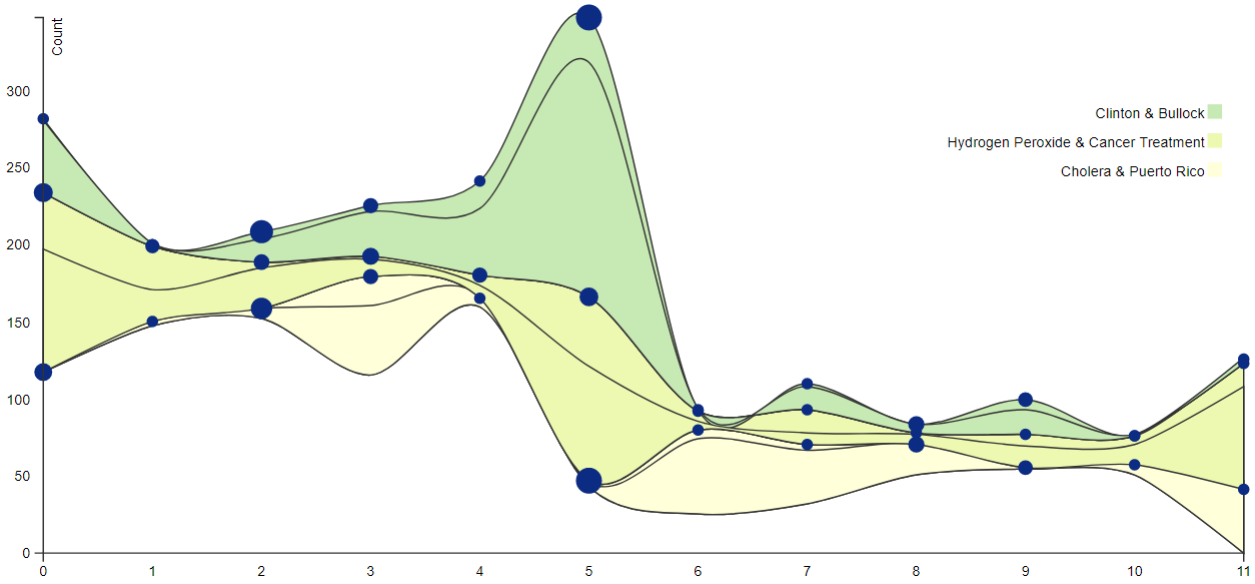
O *RumourFlow* emprega técnicas de mineração de dados, visualizações, teorias de ciências sociais, análise de sentimentos e modelos de difusão de informação para permitir a análise de rumores disseminados em redes sociais. As visualizações deste sistema permitem a análise da evolução temporal de um rumor, a identificação dos usuários que participam ativamente da disseminação de um rumor e os momentos que esses usuários agem, a análise da evolução temporal dos tópicos e a análise das palavras e tópicos mais frequentes em um rumor. Uma descrição da funcionalidade mais relevante do sistema é apresentada a seguir.

[7] Wrapper é uma classe que permite transformar tipos primitivos em objetos e vice-versa.

4.2.1 Evolução Temporal do Rumor

A metáfora de rio (cf. Subsubseção 2.2.4.1) foi empregada na visualização *Rumor Flow* para fornecer uma visão geral da evolução temporal de um rumor ou mais rumores, conforme pode ser visto na Figura 8. Nesta visualização, cada camada do "rio"representa um rumor que está sendo analisado. O eixo horizontal da visualização representa o tempo e o eixo vertical representa o número de comentários criados. Um círculo é utilizado para representar a postagem que recebeu mais comentários em cada instante de tempo. Além disso, o tamanho do círculo indica a controvérsia da postagem, de forma que, quanto maior o círculo, mais controversa a postagem é.

Figura 8 – Evolução temporal dos rumores *Hydrogen Peroxide & Cancer, Sandra Bullock & Hillary Clinton* e *Cholera & Puerto Rico* extraídos do *Reddit*. Cada rumor está representado em uma camada do "rio", conforme indicado pela legenda presente na figura. Além disso, cada círculo azul representa a postagem que recebeu a maior quantidade de comentários no instante de tempo indicado no eixo X. O tamanho do círculo representa o quão controversa é a postagem, de forma que quanto maior o círculo, mais controversa a postagem é.

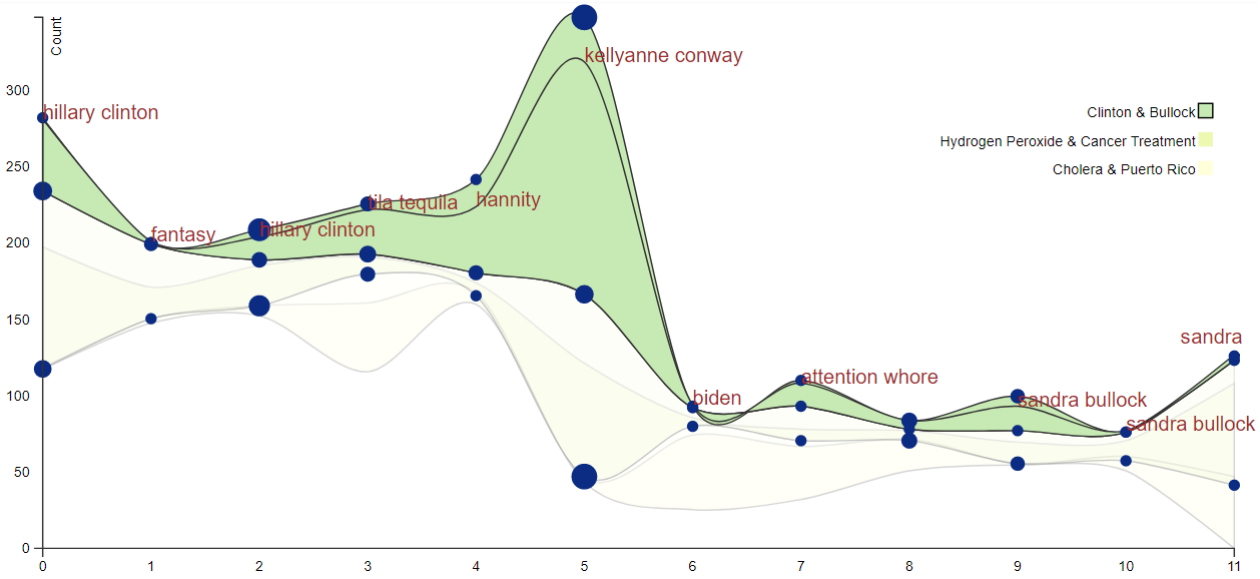


Fonte: Elaborada pelo autor.

Cada camada do "rio"possui uma subcamada que reflete o número de usuários únicos de uma postagem em um determinado instante de tempo. Isto permitiu, por exemplo, que (DANG *et al.*, 2016) identificassem que rumores com um alto número de comentários e alto número de usuários únicos são mais populares que rumores com alto número de comentários e baixo número de usuários únicos. Esta visualização também apresenta o tópico mais frequente de cada postagem e o usuário que mais criou comentários em cada postagem, conforme pode ser visto na Figura 9 e na Figura 10, respectivamente.

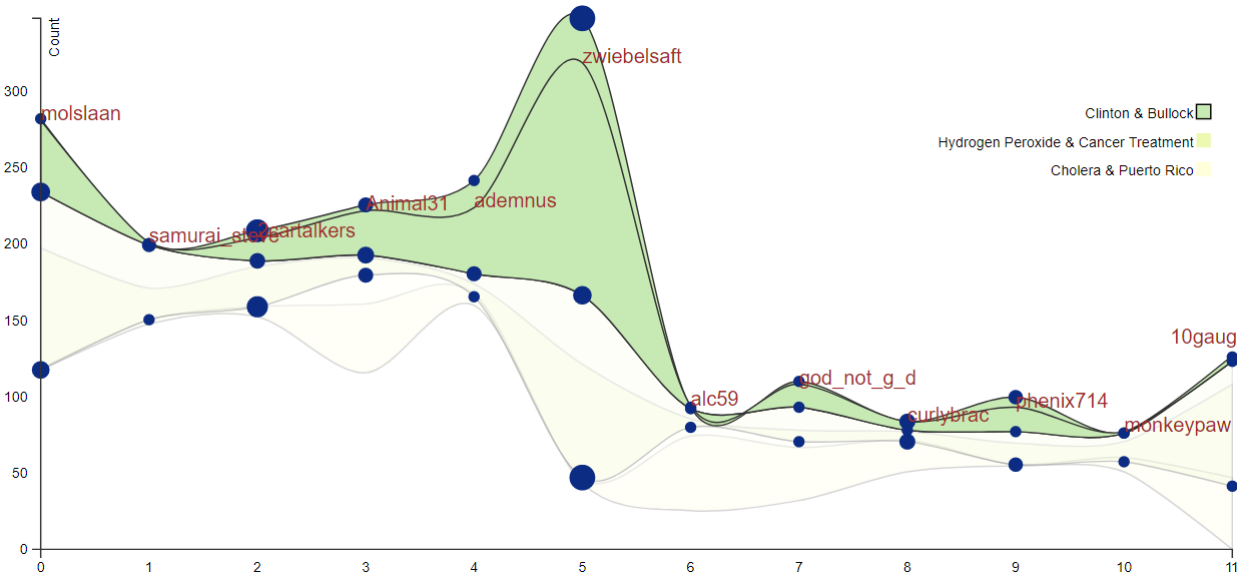
A controvérsia de uma postagem P é calculada por meio da Equação 4.1. onde *Pos* representa o número de comentários positivos recebidos pela postagem. Já *Neg* representa o

Figura 9 – Tópicos mais frequentes das postagens que receberam a maior quantidade de comentários ao longo do ciclo de vida do rumor *Hydrogen Peroxide & Cancer* extraído do *Reddit*.



Fonte: Elaborada pelo autor.

Figura 10 – Usuários que comentaram mais vezes nas postagens que receberam a maior quantidade de comentários ao longo do ciclo de vida do rumor *Hydrogen Peroxide & Cancer* extraído do *Reddit*.



Fonte: Elaborada pelo autor.

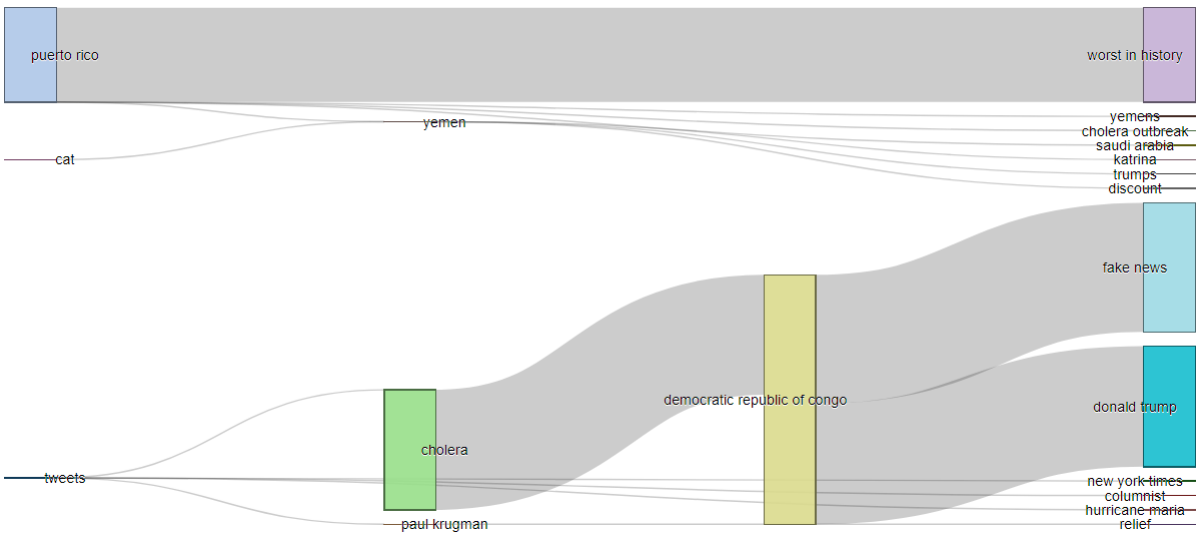
número de comentários com sentimento negativo (DANG-XUAN; STIEGLITZ, 2012). Para o cálculo do sentimento, é utilizado o *framework Stanford CoreNLP*[8], que fornece inúmeras ferramentas de Processamento de Linguagem Natural desenvolvidas em Stanford.

$$Controvérsia(P) = \frac{Pos - Neg}{Pos + Neg} \tag{4.1}$$

4.2.2 Fluxo de Tópicos

As visualizações *Semantic Topic* e *User Topic* são dois Diagramas de Sankey criados para a visualização da evolução temporal dos tópicos de um rumor. Em ambas visualizações, cada barra representa um tópico extraído do rumor. Na *Semantic Topic* (Figura 11), para dois pares de postagens contínuas $P_1 = T_1 1, T_1 2, T_1 3, ..., T_1 m$ e $P_2 = T_2 1, T_2 2, T_2 3, ..., T_2 n$, uma conexão entre os tópicos $T_1 i$ da postagem P_1 e $T_2 j$ da postagem P_2 é criada se a similaridade semântica entre os tópicos for maior que um limiar. Já na *User Topic* (Figura 12), uma conexão é criada entre os tópicos $T_1 i$ da postagem P_1 e $T_2 j$ da postagem P_2 se a quantidade de usuários em comum entre as duas postagens for maior que um limiar.

Figura 11 – Visualização *Semantic Topic* do rumor *Cholera & Puerto Rico* extraído do *Reddit*. Dada as postagens contínuas $P_1 = T_1 1, T_1 2, T_1 3, ..., T_1 m$ e $P_2 = T_2 1, T_2 2, T_2 3, ..., T_2 n$, uma conexão entre os tópicos $T_1 i$ da postagem P_1 e $T_2 j$ da postagem P_2 é criada se a similaridade semântica entre os tópicos for maior que um limiar.

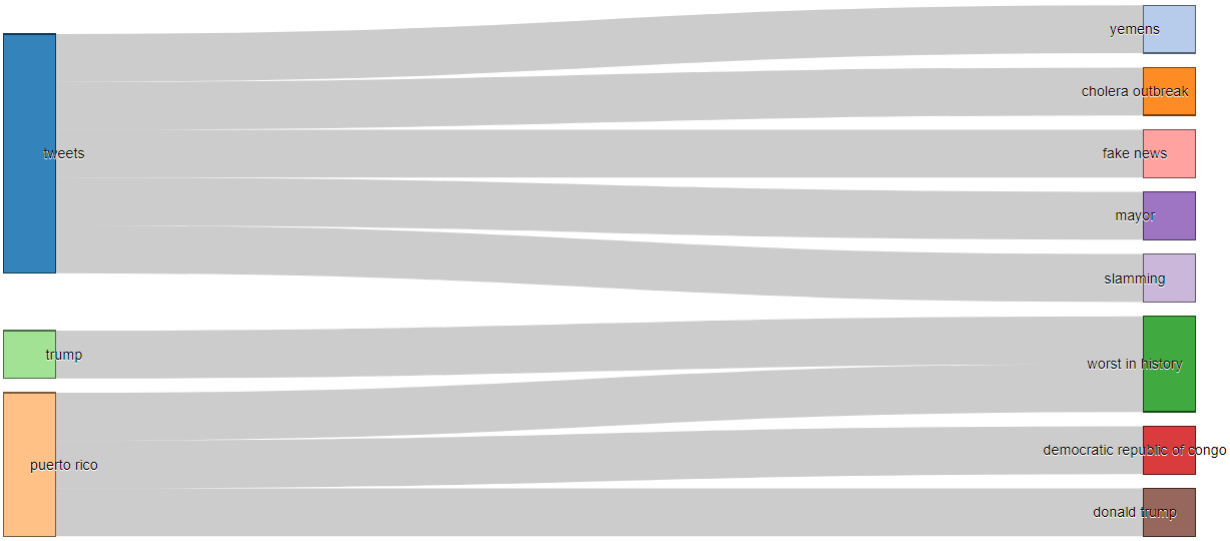


Fonte: Elaborada pelo autor.

A *Semantic Topic* permite a identificação dos tópicos sendo discutidos em diferentes momentos do ciclo de vida de um rumor e a similaridade semântica entre esses tópicos. Já a *User Topic* possibilita a visualização do volume de usuários que migra de um tópico para outro. Os tópicos são obtidos com a utilização da ferramenta de vinculação de entidade (*entity linking*)

[8] <https://stanfordnlp.github.io/CoreNLP/>

Figura 12 – Visualização *User Topic* do rumor *Cholera & Puerto Rico* extraído do *Reddit*. Dada as postagens contínuas $P_1 = T_1 1, T_1 2, T_1 3, ..., T_1 m$ e $P_2 = T_2 1, T_2 2, T_2 3, ..., T_2 n$, uma conexão entre os tópicos $T_1 i$ da postagem P_1 e $T_2 j$ da postagem P_2 é criada se a quantidade de usuários em comum entre as duas postagens for maior que um limiar.



Fonte: Elaborada pelo autor.

Dexter (cf. Subseção 2.1.10) (TRANI *et al.*, 2014). A similaridade semântica foi calculada por meio do Método de Trigramas do *Google* (*Google Trigram Method* - GTM) (ISLAM; MILIOS; KESELJ, 2012) (cf. Subseção 2.1.9).

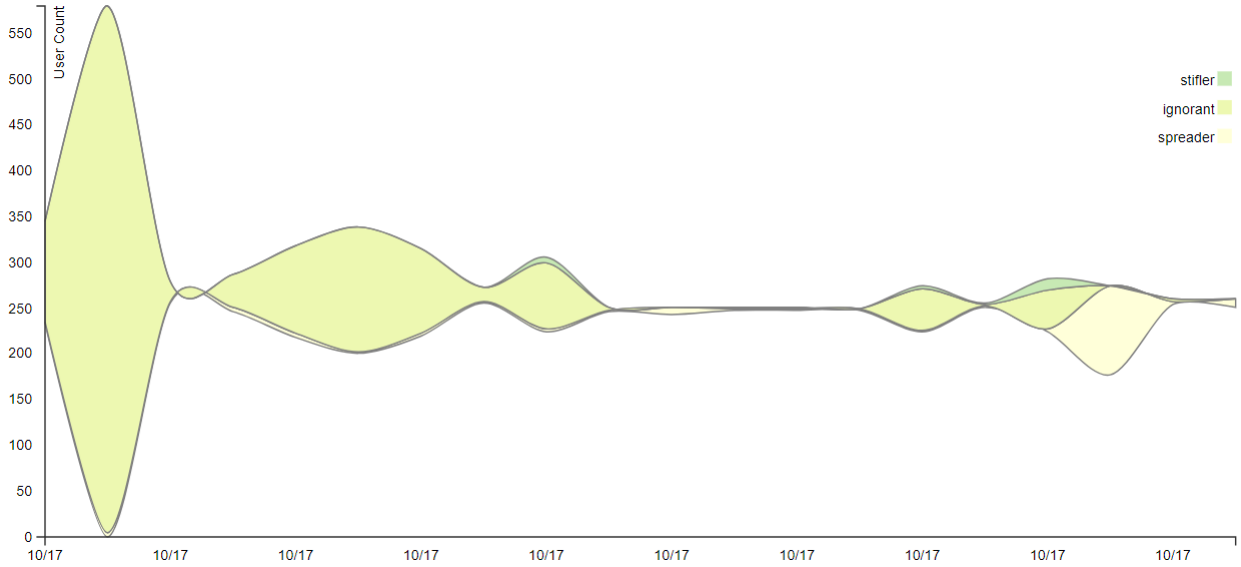
4.2.3 Análise de Usuários

A análise das ações dos usuários é importante para uma análise completa de um rumor, pois eles são os responsáveis pela criação e disseminação do rumor. Para isto, o *RumourFlow* possui três visualizações que permitem a análise da atividade dos usuários e o comportamento desses usuários ao disseminar um rumor.

A visualização *User Spread* (Figura 13) possibilita a análise do comportamento dos usuários ao disseminar um rumor. Para isto, o modelo de disseminação de rumor proposto por (DALEY; KENDALL, 1965) é utilizado. Neste modelo, os N usuários que interagem com o rumor são categorizados em uma das seguintes categorias: *spreader* (S), que é um usuário que dissemina o rumor; *ignorant* (I), que é um usuário que não teve contato com o rumor; e *stifler* (R), que é o usuário que já disseminou o rumor, mas perdeu o interesse em disseminá-lo.

No início da disseminação do rumor, apenas uma pessoa tem conhecimento sobre ele e passa a propagá-lo pelas redes sociais. No decorrer da vida do rumor, a interação dos usuários é descrita de acordo com os seguintes cenários: um usuário *ignorant*, quando entra em contato com um usuário *spreader*, torna-se *spreader* a uma taxa α ($S + I \xrightarrow{\alpha} 2S$). Quando dois *spreaders*

Figura 13 – Visualização *User Spread* do rumor *Trump & Tax Cut* extraído do Twitter. O modelo de disseminação de rumor proposto por (DALEY; KENDALL, 1965) foi empregado para gerar esta visualização. O eixo vertical desta visualização representa a quantidade de usuários e o eixo horizontal representa o tempo. Pode-se observar que há uma predominância de indivíduos do tipo *ignorant*.



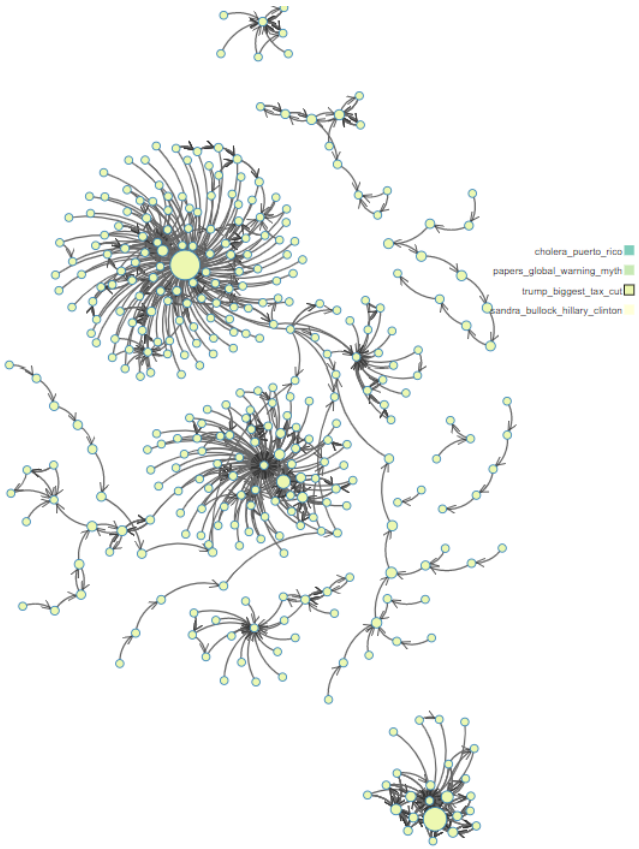
Fonte: Elaborada pelo autor.

entram em contato, um deles é transformado em *spreader* a uma taxa β ($2S \xrightarrow{\beta} S+R$). Um usuário *spreader*, quando entra em contato com um usuário *stifler*, torna-se *stifler* a uma taxa γ ($S+R \xrightarrow{\gamma} 2R$). As taxas α, β e γ são calculadas da seguinte forma:

1. A primeira postagem é criada.
   - S = 1; I = número de usuários no tempo t_0 - 1; R = 0
2. No tempo t_0, todos os usuários que comentaram na primeira postagem são *ignorants*.
3. No tempo t:
   a) Se o usuário j cria uma postagem, ele torna-se um *spreader*.
   b) Se o usuário j possui um comentário:
      i. O usuário j é um *spreader* se ele criou um comentário em $t-1$.
      ii. O usuário j é um *stifler* se ele criou um comentário em algum momento do intervalo $t-2...t_0$.
      iii. O usuário j é um *ignorant* se ele não tiver criado um comentário no intervalo $t-1...t_0$.
   c) O usuário j torna-se um *spreader* após o tempo t.

Um grafo dirigido é criado na visualização *User Graph* (Figura 14) para ilustrar a interação entre os usuários. Neste grafo, os nós representam os usuários e as arestas representam a interação entre os usuários conectados. Uma aresta é criada entre os usuários U_i e U_j, se o usuário U_i comenta em uma postagem ou em um comentário criado pelo usuário U_j. Além disso, as centralidades de intermediação (*betweenness*) e proximidade (*closeness*), e o coeficiente de agrupamento são calculados e codificados no tamanho dos nós, de forma que, quanto maior o valor da centralidade/coeficiente, maior o tamanho do nó. O maior valor possível para as centralidades e o coeficiente é 1.0 e o menor é 0.0. As centralidades e o coeficiente permitem a identificação de usuários importantes para a disseminação de um rumor e grupos que foram formados dentro da rede. Com isso, um analista pode fazer uma análise mais minuciosa desses usuários/grupos.

Figura 14 – Grafo dirigido da interação entre os usuários do rumor *Trump & Tax Cut* extraído do *Reddit*. Os nós do grafo representa os usuários e as arestas representam comentários entre os usuários. O tamanho de cada nó deste grafo está representando a centralidade de intermediação do nó.

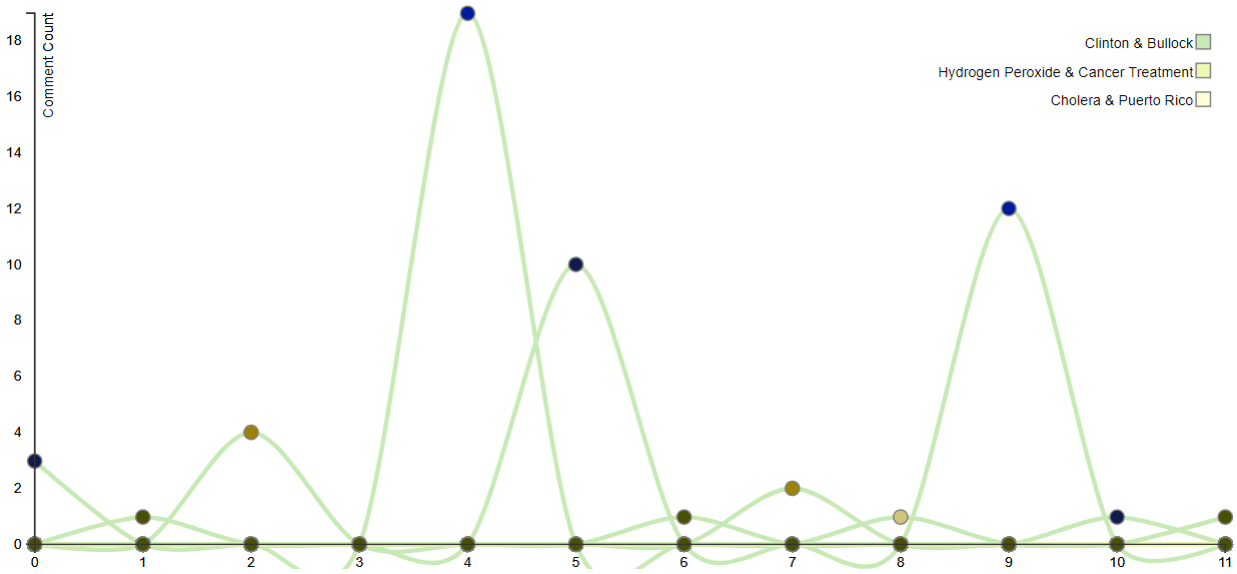


Fonte: Elaborada pelo autor.

A visualização *User activity* (Figura 15) é um gráfico de linhas que possibilita a observação dos períodos em que um usuário esteve ativo durante o ciclo de vida do rumor e, por meio dela, também é possível verificar se um usuário participou da disseminação de mais de um rumor. O eixo Y representa a quantidade de comentários que o usuário criou e o eixo X representa o tempo. Além disso, as linhas do gráfico são coloridas de acordo com o rumor que cada uma

delas representam.

Figura 15 – Visualização da atividade dos usuários mais ativos do rumor *Sandra Bullock & Hillary Clinton* extraído do *Reddit*. O eixo Y desta visualização representa o número de comentários feitos pelo usuário e o eixo X representa o tempo. Além disso, a cor das linhas representam um rumor, de acordo com a legenda presente na figura, e a cor dos círculos representam as postagens que mais receberam comentários ao longo do ciclo de vida do rumor.



Fonte: Elaborada pelo autor.

4.2.4 Nuvens de Palavras

Nuvem de Palavra é uma representação visual de texto amplamente utilizada para fornecer uma visão geral de um conjunto de dados textual. Com isso, duas nuvens de palavras foram implementadas no sistema. A nuvem *Topic Cloud* apresenta os tópicos mais frequentes de todos os rumores e, caso um rumor seja selecionado, permite que os tópicos mais frequentes do rumor selecionado sejam visualizados. Os tópicos são obtidos com a utilização da ferramenta de vinculação de entidade (*entity linking*) Dexter (cf. Subseção 2.1.10) (TRANI *et al.*, 2014). Um exemplo da *Topic Cloud* pode ser visto na Figura 16a. Já a nuvem *Word Cloud* apresenta as cinquenta palavras mais frequentes de uma postagem selecionada pelo usuário. As palavras apresentadas nesta visualização são apenas as palavras cuja frequência está acima de um limiar (valor utilizado neste trabalho: 3). O pré-processamento para esta visualização consiste em colocar todas as palavras em letra minúscula e remover as palavras vazias. Um exemplo da *Word Cloud* pode ser visto na Figura 16b.

Figura 16 – Visualizações (a) *Topic Cloud* e (b) *Word Cloud* do rumor *Hydrogen Peroxide & Cancer* disseminado no *Reddit*. O tamanho da fonte das palavras representa a frequência delas no rumor.

(a) Visualização *Topic Cloud* do rumor *Hydrogen Peroxide & Cancer*.

(b) Visualização *Word Cloud* do rumor *Hydrogen Peroxide & Cancer*.

Fonte: Elaborada pelo autor.

4.3 Análise Visual

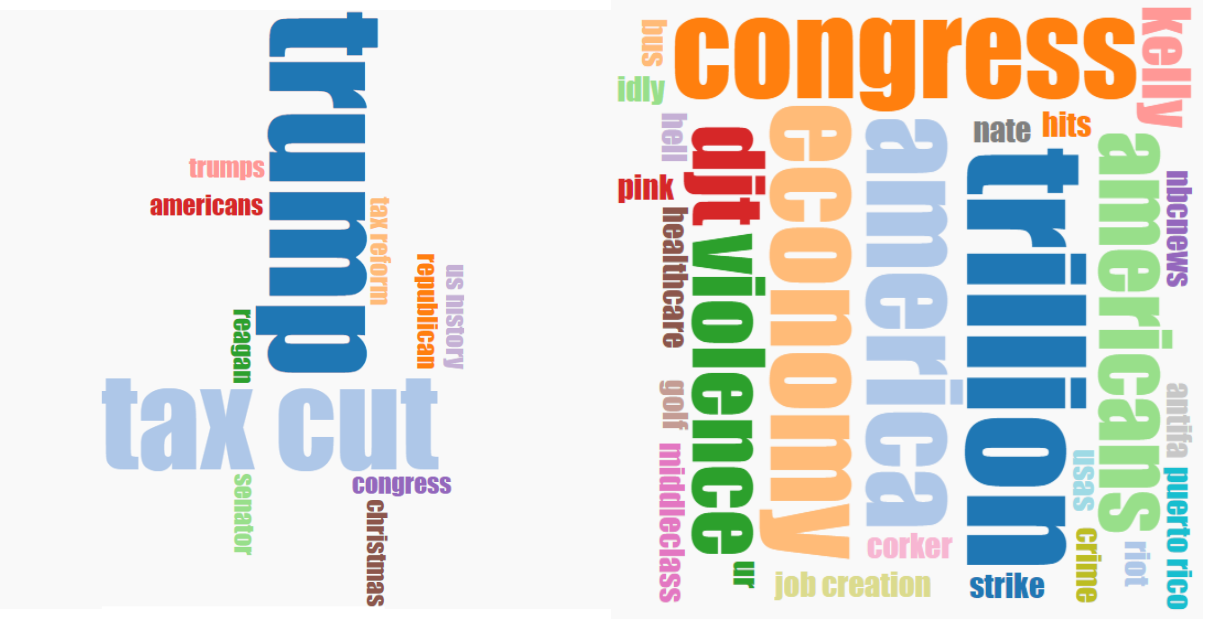
Nesta seção é apresentada a análise visual realizada com o objetivo de identificar as semelhanças e diferenças entre o *Reddit* e o *Twitter* no contexto de disseminação de rumores. Além disso, é apresentada a análise visual que teve como intuito a comparação entre rumores verdadeiros e rumores falsos.

4.3.1 Comparação Reddit - Twitter

A nuvem de palavras *Topic Cloud* foi utilizada para a análise dos tópicos mais frequentes no rumor *Trump & Tax Cut*. Foi possível identificar que os usuários do *Twitter* tendem a mencionar uma grande variedade de tópicos, isto é, mais do que 25 tópicos. Já os usuários do *Reddit* utilizam poucos tópicos quando estão discutindo um rumor, isto é, no máximo 18 tópicos. Pode-se observar na Figura 17a que apenas dois tópicos ("Trump"e "*Tax Cut*") foram os mais discutidos no *Reddit*, enquanto que no *Twitter* (Figura 17b) oito tópicos foram altamente discutidos. São eles: "*congress*", "*economy*", "*djt*", "*violence*", "*america*", "*trillion*", "*americans*"e "*kelly*".

Este comportamento pode ser atribuído ao fato de que o *Twitter* limita a quantidade de caracteres que os usuários podem utilizar para criar um *tweet*, forçando-os a ir "direto ao ponto"e, consequentemente, alguns tópicos tendem a ser repetidos. Isso não ocorre no *Reddit*, visto que esta limitação não existe na rede.

Figura 17 – Nuvens de Palavras para o rumor *Trump & Tax Cut* disseminado no *Reddit* e no *Twitter*, onde o tamanho da fonte das palavras representa a frequência delas.



(a) Nuvem de Palavras do rumor do *Reddit*

(b) Nuvem de Palavras do rumor do *Twitter*

Fonte: Elaborada pelo autor.

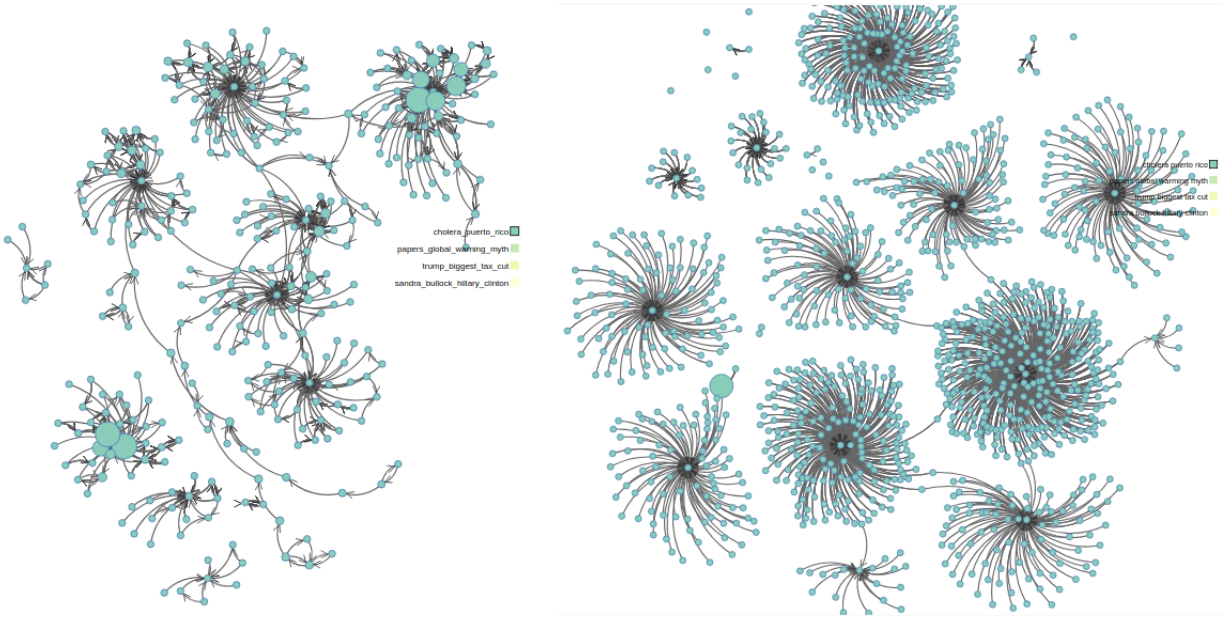
Além disso, foi identificado, por meio da visualização *Word Cloud*, que existem poucas interseções entre as palavras usadas no *Reddit* e as palavras usadas no *Twitter*. Os dois comportamentos identificados pela análise das duas nuvens de palavras são consistentes para todos os rumores da análise.

Utilizando a visualização *User Graph* para analisar o rumor *Cholera & Puerto Rico*, foi possível identificar que a maioria dos agrupamentos (*clusters*) no grafo do *Twitter* (Figura 18b) são disconexos e a centralidade de intermediação dos usuários, que é representada no grafo pelo tamanho do nó, são similares, algo totalmente diferente do que acontece no *Reddit* (Figura 18a). No *Reddit* quase todos os agrupamentos estão conectados e existem mais usuários com uma alta centralidade de intermediação, isto é, usuários cuja centralidade de intermediação está próxima do valor 1.

Isso sugere que os usuários do *Reddit* são mais propensos a se envolver na discussão de rumores, enquanto que os usuários do *Twitter* tendem a participar na discussão dos rumores comentando em apenas uma postagem.

Durante a análise da evolução temporal dos tópicos por meio das visualizações *User Topic* e *Semantic Topic* foi possível identificar que o conjunto de dados de ambas as redes sociais possuíam conjuntos de tópicos disjuntos ao longo do ciclo de vida dos rumores e a maioria das visualizações começava e terminava o fluxo com mais de um tópico. Além disso, foi observado que as visualizações frequentemente continham tópicos relacionados com o assunto

Figura 18 – Grafos de interações entre usuários do rumor *Cholera & Puerto Rico*, onde os nós representam os usuários e as arestas representam comentários entre os usuários conectados. Na Figura 18b é possível observar que boa parte dos agrupamentos são disconexos, indicando que os usuários do *Twitter* tendem a comentar em somente uma postagem durante a discussão de um rumor, ao contrário do que acontece no *Reddit* (Figura 18a).



(a) Grafo do Reddit

(b) Grafo do Twitter

Fonte: Elaborada pelo autor.

do rumor, como *"tax cut"*, *"cholera", "Puerto Rico", "Sandra Bullock"* e *"climate change"*. Essas observações indicam que não existe nenhuma diferença entre o *Reddit* e o *Twitter* em termos de evolução temporal de tópicos.

Por fim, a visualização *User Spread* foi utilizada para a análise do comportamento dos usuários. Foi possível identificar, em todos os rumores, que não existem *stiflers* no *Reddit*, algo que indica que os usuários do *Reddit* são propensos a não deixar de acreditar em um rumor. Além disso, foi identificado que a quantidade de *spreaders* no *Reddit* é baixa, enquanto que no *Twitter* esta quantidade oscila ao longo do tempo.

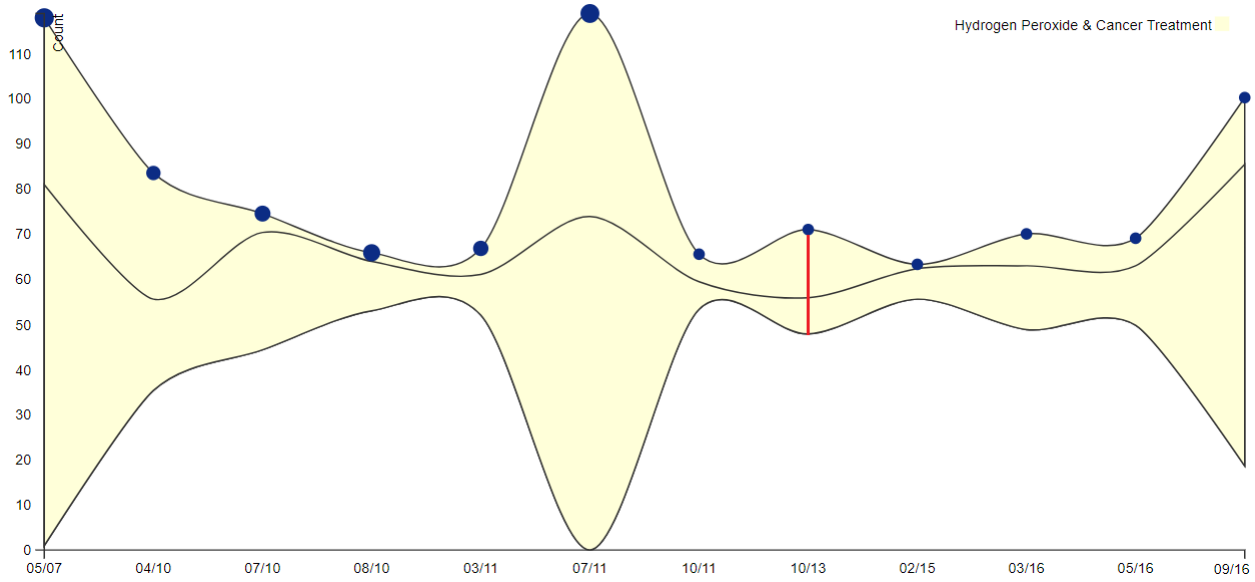
4.3.2 Comparação rumor verdadeiro - rumor falso

Na segunda análise visual empregamos os rumores *Hydrogen Peroxide & Cancer* e *Sex Offenders & Uber*. Enquanto o primeiro surgiu em 2007 e foi negado em 2013, o segundo surgiu em 2015 e foi confirmado em 2017.

Foi identificado, por meio da visualização *Rumor Flow*, que havia pessoas que ainda discutiam o rumor *Hydrogen Peroxide & Cancer* depois dele ser negado, conforme pode ser visto na Figura 19. Por exemplo, a postagem com a maior quantidade de comentários após o rumor ter sido negado possuía 177 comentários. Além disso, observou-se, por meio da árvore

de comentários do rumor, que a maioria dos usuários acreditavam no rumor e traziam mais informações para a discussão, detalhando o ponto de vista deles sobre o rumor. Ao contrário do que aconteceu no rumor *Hydrogen Peroxide & Cancer*, este comportamento não ocorreu no rumor *Sex Offenders & Uber*, conforme pode ser visto na Figura 20.

Figura 19 – Evolução temporal do rumor *Hydrogen Peroxide & Cancer*. Nesta visualização, cada círculo azul representa a postagem que recebeu a maior quantidade de comentários no instante de tempo indicado no eixo X. O tamanho do círculo representa o quão controversa é a postagem, de forma que quanto maior o círculo, mais controversa a postagem é. Além disso, a linha vertical vermelha indica o momento em que este rumor foi negado.

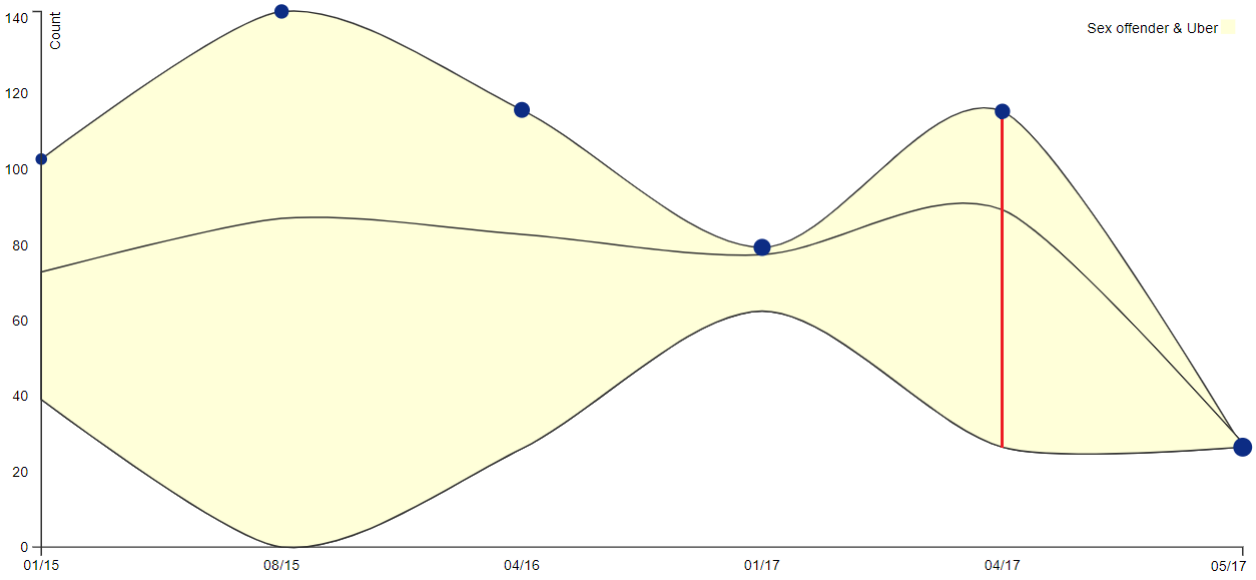


Fonte: Elaborada pelo autor.

Um usuário influente é o usuário cujo número de comentários criados por ele e comentários criados por outros usuários para respondê-lo está acima de um limiar (valor utilizado neste trabalho: 10). Dada essa definição, observou-se que o usuário mais influente no rumor *Hydrogen Peroxide & Cancer* comentou 44 vezes na postagem que era a mais controversa de todas. Já o usuário mais influente no rumor *Sex Offenders & Uber* criou 43 comentários na postagem menos controversa do rumor. Além disso, foi identificado que as postagens mais controversas nem sempre são as postagens com a maior quantidade de comentários.

Usuários cuja centralidade de intermediação é próxima do valor 1 possuem alta capacidade de intermediar uma conversação. Com isso em mente, os usuários com alta centralidade de intermediação do rumor *Sex Offenders & Uber* foram selecionados na visualização *User Graph* (Figura 21a e Figura 21b). Foi identificado que esses usuários são os usuários mais influentes do rumor. Além disso, foi observado que os usuários com alta centralidade de intermediação criaram comentários no período em que o rumor foi mais discutido no *Reddit*. Isto significa que a participação desses usuários foi essencial para que o rumor fosse disseminado ainda mais. Esta observação pode ser vista na Figura 21c e na Figura 20.

Figura 20 – Evolução temporal do rumor *Sex Offenders & Uber*. Nesta visualização, cada círculo azul representa a postagem que recebeu a maior quantidade de comentários no instante de tempo indicado no eixo X. O tamanho do círculo representa o quão controversa é a postagem, de forma que quanto maior o círculo, mais controversa a postagem é. Além disso, a linha vertical vermelha indica o momento em que este rumor foi confirmado.



Fonte: Elaborada pelo autor.

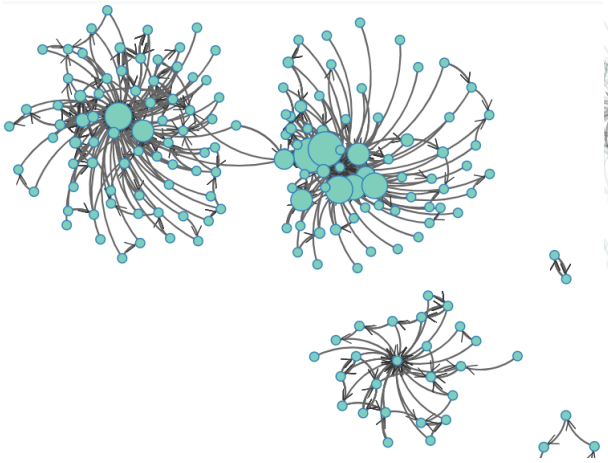
4.3.3 Considerações Finais

Neste capítulo foram descritas as análises visuais realizadas neste trabalho de mestrado por meio do sistema de visualização *RumourFlow*. A visualização *Rumor Flow* possibilitou uma melhor compreensão da evolução temporal dos rumores analisados que foram propagados no *Reddit* e no *Twitter*. Por meio desta visualização foi possível identificar os momentos em que cada rumor recebeu mais ou menos atenção dos usuários de ambas as redes sociais. Além disso, um ponto interessante que esta visualização mostrou é que os usuários pararam de propagar o rumor *Sex Offenders & Uber* após ele ser confirmado, enquanto que o rumor *Hydrogen Peroxide & Cancer* foi discutido após ele ser negado.

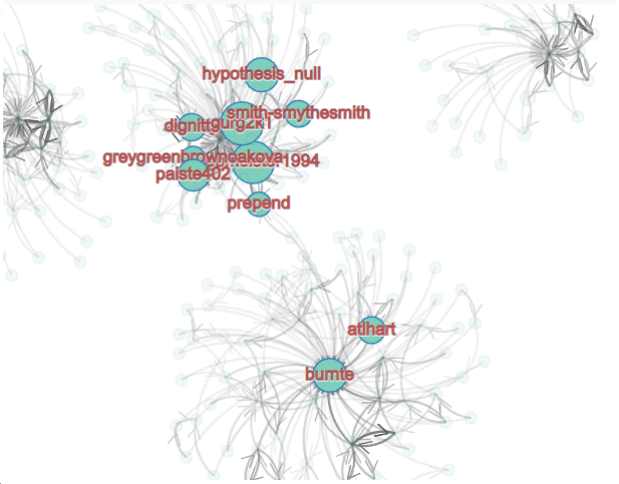
As nuvens de palavras *Topic Cloud* e *Word Cloud* permitiram a identificação dos tópicos e das palavras mais frequentes no conjunto de dados analisado. A *Topic Cloud* foi útil, pois ela possibilitou a identificação de que os usuários do *Twitter* utilizam mais tópicos na discussão de um rumor do que os usuários do *Reddit*. Já a *Word Cloud* permitiu identificar que o vocabulário dos usuários do *Reddit* é diferente do vocabulário dos usuários do *Twitter*, visto que existem poucas interseções entre as palavras usadas em ambas as redes sociais.

Foi possível perceber, por meio do grafo de usuários, a interação existente entre que os usuários do *Twitter* e do *Reddit*. Identificou-se, por exemplo, que os usuários do *Twitter* tendem a comentar somente uma vez ou a apenas compartilhar uma postagem relacionada a um rumor,

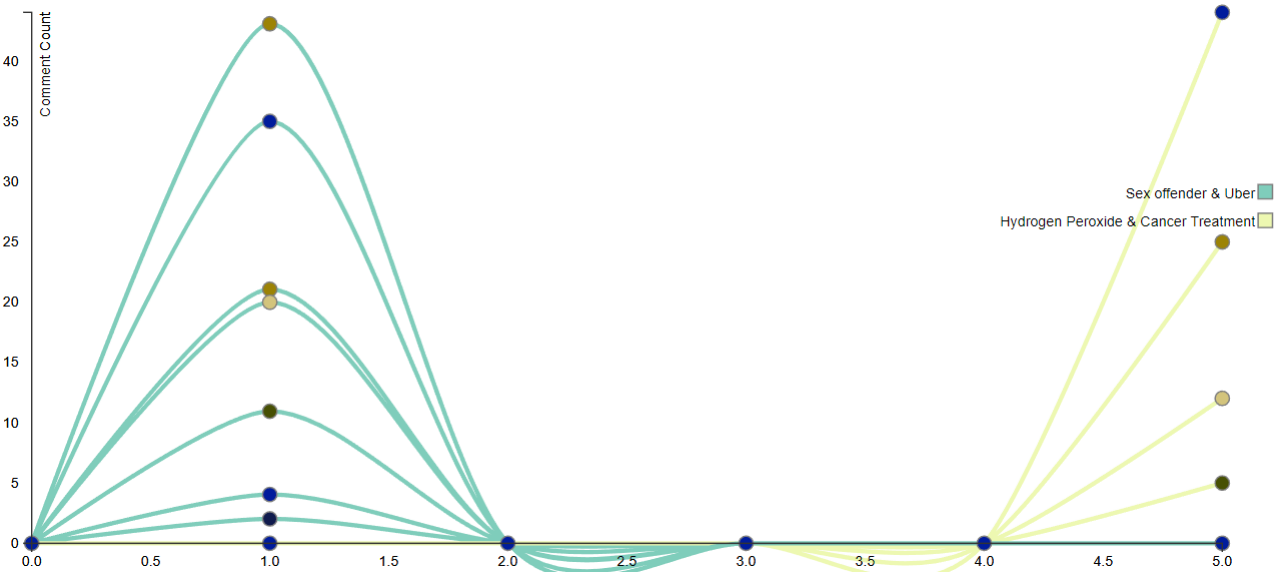
Figura 21 – Seleção dos usuários que discutiam o rumor *Sex Offenders & Uber* que possuíam as maiores centralidades de intermediação no *User Graph* e a visualização *User activity* dos usuários selecionados. Pode-se observar, por meio da Figura 21c e da Figura 20, que esses usuários criaram comentários neste período, contribuindo para que o rumor fosse propagado ainda mais.



(a) *User Graph* antes da seleção dos usuários com maior centralidade de intermediação.



(b) *User Graph* após a seleção dos usuários com maior centralidade de intermediação.



(c) *User activity* dos usuários selecionados, onde cada círculo representa um usuário, o eixo vertical representa a quantidade de comentários criado por um usuário e o eixo horizontal representa o tempo.

Fonte: Elaborada pelo autor.

enquanto que os usuários do *Reddit* são mais propensos a se envolver na discussão. Também foi possível identificar, por meio da visualização *User Spread*, qual o comportamento dos usuários na disseminação de rumores. Percebemos, por exemplo, que não existem usuários do tipo *stifler* no *Reddit*, sugerindo que os usuários desta rede social são propensos a não deixar de acreditar em um rumor.

Por fim, foi possível, por meio dos diagramas de Sankey *User Topic* e *Semantic Topic*, perceber qual o fluxo dos tópicos de cada rumor ao longo do tempo. Identificamos, por exemplo, que todos os diagramas gerados continham tópicos relacionados com o assunto do rumor e todos eles possuíam conjuntos de tópicos disjuntos durante todo o ciclo de vida dos rumores.

Pode-se concluir que o sistema de visualização de rumores *RumourFlow* permite ao usuário identificar semelhanças e diferenças entre rumores que possuem diferentes propriedades. Foi possível, por meio das visualizações disponíveis no sistema, elucidar informações que poderão ser utilizadas para apoiar o controle do impacto que rumores causam na sociedade atingida por ele.

CAPÍTULO

5

CLASSIFICAÇÃO DE RUMORES

Neste capítulo são apresentadas as etapas da classificação multiclasse supervisionada que foi desenvolvida neste trabalho de mestrado. Inicialmente é apresentado o processo de anotação manual dos dados que foram utilizados na classificação. Em seguida, é apresentado o pré-processamento dos dados, onde um conjunto de transformações tornaram os dados processáveis pelos algoritmos de classificação utilizados. Depois, são apresentados os classificadores empregados na classificação e as métricas utilizadas para medir o desempenho dos classificadores. Por fim, são apresentados os resultados obtidos na etapa de classificação.

5.1 Anotação Manual dos Dados

A classificação apresentada neste capítulo teve como objetivo detectar se um usuário acredita no rumor que ele está disseminando. Para isto, foram utilizadas postagens e comentários relacionados aos seguintes rumores: "*Trump & Tax Cut*", "*Cholera & Puerto Rico*", "*Papers & Global Warming*", "*Sandra Bullock & Hillary Clinton*"e "*9-11 & Conspiracies*".

Após uma análise das postagens e comentários, e uma discussão com os pesquisadores colaboradores, optou-se pela utilização das classes "positivo", "negativo"e "neutro"para esta classificação. Foi necessário utilizar a classe "neutro", pois houve casos em que não foi possível identificar se o usuário acreditava ou não no rumor.

Após a definição das classes, todas as postagens e comentários foram anotados manualmente. Em um primeiro momento, a anotação foi realizada apenas pelo autor deste trabalho, para que alguns testes fossem realizados. Esses testes tinham como intuito avaliar a qualidade da abordagem proposta para esta etapa.

Após a realização dos testes, quatro voluntários (acadêmicos de ciência da computação) realizaram a anotação. Os anotadores receberam uma planilha com todas as submissões (postagens e comentários) e receberam intruções em como diferenciar uma submissão que deveria ser

marcada como positivo, negativo ou neutro. Além disso, eventuais dúvidas foram sanadas ao longo do processo de anotação.

A medida *Fleiss' Kappa* (κ) e o coeficiente *Krippendorff's Alpha* (α) nominal e foram utilizados para medir a concordância entre os anotadores (cf. Subseção 2.1.7). O nominal resultou em 0.155 e o *Kappa* resultou em 0.192, indicando que houve uma baixa concordância entre os anotadores, visto que $0 \leq \alpha \leq 1$ e $0 \leq \kappa \leq 1$.

Ocorreram 88 empates entre as 536 submissões anotadas, isto é, houve votos iguais para duas classes diferentes (positivo e neutro, positivo e negativo ou neutro e negativo). Entre os empates, 39 foram entre a classe positivo e a classe neutro, 12 entre a classe positivo e a classe negativo e 37 entre a classe neutro e a classe negativo. Essas submissões foram removidas da classificação, pois eram submissões ambíguas. A composição dos votos de cada anotador e a composição do conjunto de dados final pode ser vista na Tabela 5.

Tabela 5 – Composição dos votos e do conjunto de dados final.

Classe	Anotador 1	Anotador 2	Anotador 3	Anotador 4	Anotador 5	Final
Positivo	133	289	118	45	132	110
Neutro	276	28	338	373	280	238
Negativo	127	219	80	118	124	100

Fonte: Elaborada pelo autor.

5.2 Pré-processamento

O pré-processamento dos dados consiste em um conjunto de transformações que tornam os dados processáveis por algoritmos de classificação. Este conjunto de transformações inclui: limpeza e padronização do texto, extração de atributos, normalização, entre outras. O pré-processamento desse trabalho é descrito a seguir.

5.2.1 Limpeza

Palavras vazias são palavras que não trazem nenhuma contribuição ao problema que está sendo tratado e são frequentes em um conjunto de dados. Alguns exemplos de palavras vazias incluem: "a", "o", "no"e "em", no português, e "*the*", "*is*", "*in*", "*at*"e "*on*", no inglês. O processo de remoção de palavras vazias é comum em trabalhos de classificação e foi aplicado neste trabalho, pois o conjunto de dados utilizado possuía um grande número de palavras vazias. Vale ressaltar que os advérbios de negação foram mantidos, devido ao fato de que a negação do que foi dito por um usuário pode alterar a interpretação da posição dele em relação ao rumor.

Além da remoção das palavras vazias, todas as submissões foram convertidas para caixa baixa, a fim de que palavras como "*President*"e "*president*"fossem mapeadas para a mesma

unidade lexical. Datas e horários também foram removidos, pois não são relevantes ao problema em questão.

5.2.2 Escolha e Extração dos Atributos

Após uma análise do conjunto de dados a ser classificado e a identificação dos principais atributos utilizados em trabalhos da literatura, foram escolhidos 12 atributos para a classificação, divididos em 3 categorias (Tabela 6). Entre os atributos utilizados, 10 são do tipo numérico e apenas 2 são categóricos. Uma descrição dos atributos é apresentada a seguir.

- **Usuário:** os atributos dessa categoria são aqueles que trazem alguma informação sobre o usuário que criou uma submissão. O atributo *Tipo do Usuário* representa o comportamento do usuário em relação ao rumor de acordo com o modelo teórico de disseminação de rumores proposto por (DALEY; KENDALL, 1965). Este atributo é do tipo categórico e pode assumir um dos seguintes valores: *ignorant*, caso o usuário não tenha tido contato com o rumor; *spreader*, caso o usuário seja um disseminador do rumor; ou *stifler*, se o usuário já disseminou o rumor, mas perdeu o interesse no mesmo.

  A rede social *Reddit* possui a medida *Karma* para representar a quantidade de votos (*upvotes* e *downvotes*) que um usuário recebeu em uma postagem ou comentário. Esta medida é dividida em dois tipos: karma de postagem, que representa a quantidade de votos recebidos pelas postagens criadas pelo usuário e karma de comentário, que representa o número de votos que o usuário recebeu em seus comentários. Tanto a karma de postagem, quanto a karma de comentário foram utilizadas como atributos e ambos são do tipo numérico.

  Além disso, utilizamos o número de comentários e número de postagens criadas pelo usuário como atributos. Ambos atributos são do tipo numérico.

- **Conteúdo:** atributos categorizados como *conteúdo* são aqueles que representam algo do texto presente em uma submissão. O atributo *sentimento* será utilizado, pois além dele ser comumente empregado em trabalhos de classificação de rumores, o sentimento influencia as ações e atitudes do usuário em relação ao rumor, além de contribuir na popularidade do rumor (DIFONZO; BORDIA, 2007) (DANG-XUAN; STIEGLITZ, 2012). Este atributo é do tipo categórico e pode assumir os valores *positivo, neutro* e *negativo*.

  A quantidade de tópicos presentes em uma submissão também foi utilizada como atributo, pois os tópicos possuem um papel importante na disseminação de rumores, além de influenciarem na interação dos usuários com o rumor (ROSNOW; FOSTER, 2005) (BORDIA; DIFONZO, 2005). Este atributo é do tipo numérico e pode assumir qualquer valor positivo.

  Uma análise do conjunto de dados permitiu a identificação de submissões sarcásticas, por isso optou-se pela utilização de um atributo numérico para representar o nível de

sarcasmo presente em uma submissão. Este atributo pode assumir qualquer valor no intervalo $-100 \leq x \leq 100$, onde x representa o valor do atributo e, quanto maior o valor de x, mais sarcástica é a submissão.

Para o atributo numérico *TF-IDF*, foi utilizada uma Matriz Documento-Termo (SALTON; MCGILL, 1986) com os pesos de cada palavra presente nas submissões. Esses pesos são calculados por meio da medida TF-IDF, que tem o propósito de medir o quão importante uma palavra é em um documento ou corpus.

- **Grafo e Árvore:** os atributos dessa categoria estão relacionados ao grafo (cf. Subseção 4.2.3) que é construído a partir das interações entre os usuários que criaram as submissões utilizadas na classificação e a árvore de comentários criada a partir de cada postagem utilizada na classificação.

  De acordo com (MENDOZA; POBLETE; CASTILLO, 2010), rumores são disseminados por um pequeno grupo de usuários influenciadores que têm a capacidade de disseminar rumores em um menor tempo quando comparado a usuários comuns. A centralidade de proximidade permite a identificação de usuários que podem propagar rumores para uma grande quantidade de usuários em um curto período de tempo. Já a centralidade de intermediação permite a identificação de usuários capazes de disseminar rumores para uma grande quantidade de usuários e influenciar a popularidade do rumor (DANG *et al.*, 2019). Por conta disso, as centralidades de proximidade e intermediação são calculadas a partir do grafo de interações entre os usuários (cf. Subseção 4.2.3) e utilizadas como atributo do tipo numérico.

  Além das centralidades, utilizamos, como atributo do tipo numérico, o nível que uma submissão se encontra na árvore de comentários mencionada anteriormente.

Tabela 6 – Informações sobre os atributos extraídos para a classificação desenvolvida neste trabalho.

Categoria	**Atributo**	**Tipo**
Usuário	Tipo do Usuário	Categórico
	Karma de Postagem	Numérico
	Karma de Comentário	Numérico
	Número de Comentários	Numérico
	Número de Postagens	Numérico
Conteúdo	Sentimento	Categórico
	Número de Tópicos	Numérico
	Sarcasmo	Numérico
	TF-IDF	Numérico
Grafo e Árvore	Centralidade de Proximidade	Numérico
	Centralidade de Intermediação	Numérico
	Nível do Comentário	Numérico

Fonte: Elaborada pelo autor.

Figura 22 – Exemplo da transformação feita para converter atributos categóricos em atributos numéricos, onde cada possível valor do atributo categórico virou um atributo do tipo binário. Na Figura 22a temos os atributos antes de serem convertidos, enquanto que na Figura 22b temos os atributos após a conversão.

Submissão	Sentimento
Submissão 1	Negativo
Submissão 2	Positivo
Submissão 3	Positivo
Submissão 4	Negativo
Submissão 5	Neutro

(a)

Submissão	Positivo	Neutro	Negativo
Submissão 1	0	0	1
Submissão 2	1	0	0
Submissão 3	1	0	0
Submissão 4	0	0	1
Submissão 5	0	1	0

(b)

Fonte: Elaborada pelo autor.

5.2.3 Normalização

Na última etapa do pré-processamento, os atributos categóricos foram transformados em atributos numéricos e, em seguida, todos os atributos foram normalizados, pois eles estavam em escalas diferentes.

Para transformar um atributo categórico em um atributo numérico, cada possível valor do atributo categórico foi transformado em um atributo binário. Um exemplo desta transformação está ilustrado na Figura 22.

A normalização aplicada aos atributos foi feita por meio da seguinte equação:

$$X_{norm} = \frac{X - Min}{Max - Min} \tag{5.1}$$

onde X é o valor a ser normalizado, Max é o maior valor do atributo e Min é o menor valor do atributo. Com isso, todos os atributos que estavam em escalas diferentes foram trazidos para a mesma escala, deixando-os dentro do intervalo $[0, 1]$

5.3 Classificadores

A classificação feita neste trabalho foi realizada em duas etapas. A primeira etapa teve como objetivo avaliar a qualidade da abordagem proposta, portanto, os dados utilizados foram anotados somente pelo autor deste trabalho. Nesta etapa, foram utilizadas as implementações dos algoritmos e métodos *Naive Bayes*, *SVM*, *KNN*, *Árvore de Decisão* e *Floresta Aleatória* disponíveis na *Weka*[1] (WITTEN *et al.*, 2016). Na segunda etapa, os dados utilizados já tinham sido anotados por todos os anotadores e, além dos algoritmos e métodos utilizados na primeira etapa, o comitê de classificadores Votação Majoritária também foi utilizado, porém as implementações

[1] <https://www.cs.waikato.ac.nz/ml/weka/>

utilizadas foram as que estão disponíveis na biblioteca *scikit-learn*[2] (PEDREGOSA *et al.*, 2011). As definições dos algoritmos e métodos utilizados podem ser verificadas na Seção 2.1.

5.4 Avaliação

É importante medir os resultados obtidos em tarefas de classificação, pois assim é possível identificar a qualidade da abordagem proposta. Portanto, foi utilizado um conjunto de métricas comumente empregadas na literatura para mensurar o desempenho dos classificadores. As métricas utilizadas são: Acurácia, Precisão, Revocação e F1 (cf. Subseção 2.1.6).

5.5 Resultados Obtidos

Nesta seção são apresentados os resultados da classificação desenvolvida neste trabalho de mestrado. Conforme visto no início do capítulo, a classificação foi desenvolvida em duas etapas. Além disso, os métodos de avaliação apresentados na Seção 5.4 foram empregados para avaliar cada um dos algoritmos que foram utilizados nas duas etapas (cf. Seção 5.3).

5.5.1 Primeira Etapa do Processo de Classificação

Na Tabela 7 são apresentados os valores da acurácia, precisão, revocação e F1 para os algoritmos utilizados na primeira etapa, onde a qualidade da abordagem proposta foi avaliada. É possível ver que o SVM foi o melhor classificador, atingindo 61.7% de acurácia, porém a Árvore de Decisão atingiu resultados similares. Por outro lado, o pior desempenho foi o do *Naive Bayes*, que obteve 55.53% de acurácia.

Tabela 7 – Resultados da classificação multiclasse da primeira etapa usando validação cruzada 10-fold.

Classificador	**Acurácia**	**Precisão**	**Revocação**	**F1**
Naive Bayes	55.53%	53%	55.5%	53.9%
SVM	61.7%	59.5%	61.7%	59.4%
KNN	58.61%	52.6%	58.6%	48.2%
Árvore de Decisão	60.41%	57.6%	60.4%	57.7%
Floresta Aleatória	57.58%	49.3%	57.6%	47.2%

Fonte: Elaborada pelo autor.

Após esta avaliação, optou-se pela investigação do desempenho dos algoritmos para cada classe por meio da medida F1. Os valores da F1 para cada classe são apresentados na Tabela 8. A classe *neutro* obteve os melhores resultados em todos os classificadores, enquanto que as outras duas classes atingiram valores baixos, onde o valor máximo da classe *positivo* foi 34.4% e o da

[2] <https://scikit-learn.org/stable/>

negativo foi 47.7%. Os resultados atingidos vão de encontro ao esperado, visto que havia um desbalanceamento no conjunto de dados, onde a classe *neutro* predominava.

Tabela 8 – Resultados da medida F1 calculada para as três classes.

Classificador	**F-Positivo**	**F-Neutro**	**F-Negativo**
Naive Bayes	31.6%	69.5%	31.6%
SVM	34.4%	73%	47.7%
KNN	7.8%	73.5%	18.4%
Árvore de Decisão	32.9%	72.6%	41.4%
Floresta Aleatória	5.7%	72.2%	19.5%

Fonte: Elaborada pelo autor.

O impacto de cada atributo no resultado da acurácia foi avaliado. Para isso, o conjunto de dados foi classificado 12 vezes por classificador, onde um atributo foi removido a cada vez que um modelo era treinado e testado. A Tabela 9 apresenta os resultados obtidos.

A ausência da Matriz Documento-Termo do TF-IDF foi a que mais impactou no valor da acurácia em todos os algoritmos, onde a diferença entre o valor da acurácia com todos os atributos e o valor da acurácia sem o TF-IDF foi de 4.38% no *Naive Bayes*, 2.32% no SVM, 2.06% no KNN, 5.4% na Árvore de Decisão e 1.03% na Floresta Aleatória. Houve casos em que a remoção de um atributo contribuiu para o aumento da acurácia, porém optou-se por não removê-los da segunda etapa da classificação, visto que a diferença entre a acurácia com e sem o atributo não foi maior que 10%.

Tabela 9 – Resultados da acurácia de cada classificador com a remoção de um atributo.

Atributo Removido	**Naive Bayes**	**SVM**	**KNN**	**Árvore de Decisão**	**Floresta Aleatória**
TF-IDF	51.15%	59.38%	56.55%	55.01%	56.55%
Cent. Intermediação	57.06%	61.69%	58.61%	57.58%	57.84%
Cent. Proximidade	55.27%	60.41%	57.06%	60.15%	57.84%
Sentimento	55.78%	60.41%	59.64%	60.15%	59.38%
Sarcasmo	55.78%	61.44%	58.61%	59.12%	58.61%
Num. Tópicos	57.32%	61.69%	58.35%	58.35%	58.09%
Karma	55.78%	61.69%	58.61%	59.64%	58.09%
Nível do Comentário	55.26%	60.92%	59.12%	61.18%	56.04%
Tipo do Usuário	56.55%	62.21%	59.38%	60.15%	58.35%
Num. Comentários	55.01%	61.18%	58.86%	60.41%	58.09%
Num. Postagens	56.29%	61.69%	58.61%	59.89%	58.09%
Nenhum	55.53%	61.7%	58.61%	60.41%	57.58%

Fonte: Elaborada pelo autor.

5.5.2 Segunda Etapa do Processo de Classificação

Na segunda etapa da classificação, o conjunto de dados havia passado pelo processo de anotação descrito na Seção 5.1. Os resultados da classificação feita após todos os anotadores terem concluído a anotação podem ser vistos na Tabela 10. O *Naive Bayes* e o KNN foram os algoritmos que sofreram a maior mudança nos resultados entre uma etapa e outra. Enquanto o *Naive Bayes* teve um ganho de 3.76% na acurácia, 4.9% na precisão, 3.8% na revocação e 5.4% na F1, o KNN sofreu uma perda de 5.61% na acurácia, 9.88% na precisão, 5.6% na revocação e 7.8% na F1.

Usando o comitê de classificadores Votação Majoritária para combinar as predições feitas pelos algoritmos que tiveram o melhor desempenho nesta classificação (*Naive Bayes, SVM e Floresta Aleatória*), obteve-se um resultado inferior ao resultado individual de cada classificador utilizado no comitê.

Tabela 10 – Resultados da classificação multiclasse da segunda etapa usando validação cruzada 10-fold.

Classificador	**Acurácia**	**Precisão**	**Revocação**	**F1**
Naive Bayes	59.29%	57.9%	59.3%	59.3%
SVM	60.64%	62.23%	60.64%	60.64%
KNN	53%	42.72%	53%	40.4%
Árvore de Decisão	58.38%	55.8%	58.38%	58.38%
Floresta Aleatória	60.65%	60.89%	60.65%	60.65%
Comitê de Classificadores (NB SVM FA)	59%	58%	59%	54%

Fonte: Elaborada pelo autor.

Após a identificação das mudanças ocorridas entre as duas etapas de classificação (Tabela 7 e Tabela 10), optou-se pela realização de uma classificação binária, onde as submissões da classe *neutro* foram removidas do conjunto de dados a ser classificado. Com isso, o desbalanço existente nos dados foi reduzido, visto que foram classificadas 110 submissões da classe *positivo* e 100 da classe *negativo*.

Os resultados da classificação binária podem ser vistos na Tabela 11. É possível observar que houve um aumento no desempenho de todos os algoritmos. O *Naive Bayes* e o KNN foram os algoritmos que tiveram a maior mudança, visto que ambos tiveram um ganho de mais de 10% em todas medidas. Os algoritmos baseados em árvore tiveram um aumento significativo na *precisão* (Árvore de Decisão 10.82% e Floresta Aleatória 7.11%), quando comparado ao ganho que tiveram nas outras medidas.

Ao contrário do que ocorreu na classificação multiclasse, o comitê de classificadores obteve os melhores resultados. Este comitê foi composto pelo *Naive Bayes*, SVN e KNN, pois eles foram os classificadores que atingiram os melhores resultados na classificação binária.

Tabela 11 – Resultados da classificação binária da segunda etapa usando validação cruzada 10-fold.

Classificador	Acurácia	Precisão	Revocação	F1
Naive Bayes	71.8%	74.2%	71.8%	71.8%
SVM	68%	69.9%	68%	68%
KNN	68.4%	70.7%	68.4%	68%
Árvore de Decisão	62.2%	66.62%	62.2%	61.9%
Floresta Aleatória	63.2%	68%	63.2%	63%
Comitê de Classificadores (NB SVM KNN)	72.3%	75.4%	72.3%	72%

Fonte: Elaborada pelo autor.

5.5.3 Considerações Finais

Neste capítulo foi descrita a classificação supervisionada desenvolvida neste trabalho de mestrado. Os resultados obtidos pela classificação multiclasse (classes positivo, neutro e negativo) não foram satisfatórios, pois o classificador que obteve o melhor resultado (SVM) atingiu 61.7% de acurácia, 59.5% de precisão, 61.7% de revocação e 59% de F1 na primeira etapa (Tabela 7) e 60.64% de acurácia, 62.23% de precisão, 60.64% de revocação e 60.64% de F1 na segunda etapa (Tabela 10). Já o resultado obtido pela classificação binária foi aceitável, visto que o Naive Bayes atingiu 71.8% de acurácia, 74.2% de precisão, 71.8% de revocação e 71.8% de F1 e o comitê de classificadores obteve 72.3% de acurácia, 75.4% de precisão, 72.3% de revocação e 72% de F1, conforme pode ser visto na Tabela 11.

Após a finalização do desenvolvimento da classificação apresentada neste capítulo, foram identificados pontos que podem ser melhorados para que resultados mais satisfatórios sejam obtidos. Primeiramente, pode-se aplicar técnicas de super-amostragem (*oversampling*) ou sub-amostragem (*undersampling*) para tratar o problema do desbalanceamento dos dados. Além disso, um conjunto de dados maior pode ser coletado, visto que o conjunto empregado neste trabalho era pequeno. Por fim, a utilização de técnicas de anotação semi-automática podem ser utilizadas, pois a anotação realizada neste trabalho foi limitada.

A adição de novos atributos também pode contribuir para a melhoria dos resultados da classificação. Por exemplo, podem ser empregados atributos que contêm informações do perfil do usuário na rede social, como a frequência em que o usuário utiliza a rede, quantidade de seguidores que ele possui, entre outros, e atributos que trazem informações sobre as submissões presentes na árvore de comentários, visto que os usuários influenciam uns aos outros.

CAPÍTULO

6

CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho foram apresentados resultados de análises visuais com o intuito de identificar padrões em rumores. O sistema de visualização *RumourFlow* foi utilizado para que diferentes pontos dos rumores fossem analisados e comparados, conforme descrito no Capítulo 4. Também foi apresentada uma classificação supervisionada que teve como objetivo classificar se um usuário acredita no rumor que ele está disseminando. Para isso, o conjunto de dados utilizado passou por uma série de processos para ser empregado nos classificadores listados, conforme descrito no Capítulo 5. Os resultados obtidos foram avaliados por meio da utilização das métricas apresentadas na Seção 5.4.

Considera-se como contribuição deste trabalho as análises realizadas para destacar características das redes sociais que influenciam no modo como os rumores são disseminados nelas e, além disso, evidenciou-se particularidades de rumores que possuem diferentes veracidades. Através das observações obtidas nesta dissertação, pode-se melhorar o desempenho de detectores de rumores e, consequentemente, apoiar o controle do impacto que informações falsas causam nas sociedades atingidas por elas.

Além das análises, este trabalho apresenta uma abordagem para a detecção da crença de usuários em rumores disseminados em redes sociais. Esta abordagem empregou atributos de diferentes tipos, tais como atributos relacionados a ciências sociais, contribuindo para a diminuição da lacuna existente entre pesquisas experimentais e teorias de ciências sociais.

Algumas limitações existiram no desenvolvimento desta pesquisa. No processo da análise visual, pode-se destacar como limitação a utilização da biblioteca D3.js, que se torna lenta na apresentação de dados com milhares de amostras. Além disso, o processo de coleta de dados do *Reddit* e do *Twitter* é trabalhoso e consome um tempo considerável, tornando imprescindível o desenvolvimento de sistemas de coleta e análise de rumores em tempo real.

Pode-se destacar como limitação da classificação o tamanho pequeno do conjunto de dados classificado. Além disso, o processo de anotação manual é limitado. Por exemplo, um

ambiente web poderia ter sido desenvolvido para facilitar o trabalho manual do anotador e técnicas de anotação automática poderiam ter sido exploradas para diminuir o tempo que a etapa de anotação consumiu.

Esta pesquisa pode ser estendida da seguinte forma:

- Visualização e análises:
  - Desenvolvimento de sistemas que permitam a análise de rumores em tempo real.
  - Análise de diferenças e semelhanças entre outras redes sociais, como o *Facebook, Weibo e Instagram*.
  - Utilização de bibliotecas mais eficientes do que a D3.js, como a WebGL.
- Classificação:
  - Utilização de um conjunto de dados maior.
  - Melhoria do processo de anotação por meio da exploração técnicas de anotação semi-automática guiada pelo usuário.
  - Adição de atributos que possam melhorar a descrição dos dados, como a utilização de mais informações da rede social.
  - Utilização e desenvolvimento de visualizações que possam auxiliar nas etapas de classificação.

REFERÊNCIAS

ALSUKHNI, M.; ZHU, Y. Interactive visualization of the social network of research collaborations. In: **2012 IEEE 13th International Conference on Information Reuse Integration (IRI)**. Las Vegas, NV, USA: IEEE, 2012. p. 247–254. Citado na página 42.

BAO, Y.; YI, C.; XUE, Y.; DONG, Y. A new rumor propagation model and control strategy on social networks. In: **Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on**. Niagara Falls, ON, Canada: IEEE, 2013. p. 1472–1473. Citado nas páginas 25, 44 e 45.

BORDIA, P.; DIFONZO, N. Psychological motivations in rumor spread. **RUMOR MILLS**, 01 2005. Citado na página 69.

BOSCH, H.; THOM, D.; HEIMERL, F.; PüTTMANN, E.; KOCH, S.; KRüGER, R.; WöRNER, M.; ERTL, T. Scatterblogs2: Real-time monitoring of microblog messages through user-guided filtering. **IEEE Transactions on Visualization and Computer Graphics**, v. 19, n. 12, p. 2022–2031, Dec 2013. ISSN 1077-2626. Citado na página 41.

BOSTOCK, M. **Force-Directed Graph**. 2017. Disponível em: <https://bl.ocks.org/mbostock/4062045>. Citado na página 37.

BOYANDIN, I.; BERTINI, E.; BAK, P.; LALANNE, D. Flowstrates: An approach for visual exploration of temporal origin-destination data. **Computer Graphics Forum**, v. 30, p. 971–980, 06 2011. Citado nas páginas 24 e 44.

BRANDTZÆG, P. B.; HEIM, J. Why people use social networking sites. In: OZOK, A. A.; ZAPHIRIS, P. (Ed.). **Online Communities and Social Computing: Third International Conference, OCSC 2009, Held as Part of HCI International 2009, San Diego, CA, USA, July 19-24, 2009. Proceedings**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. p. 143–152. ISBN 978-3-642-02774-1. Citado na página 23.

BRANTS, T.; FRANZ, A. The google web 1t 5-gram corpus version 1.1. **Technical Report**, 2006. Citado na página 32.

BREIMAN, L. Random forests. **Machine Learning**, Kluwer Academic Publishers, v. 45, n. 1, p. 5–32, Oct 2001. ISSN 1573-0565. Citado na página 29.

CHENG, J.-J.; LIU, Y.; SHEN, B.; YUAN, W.-G. An epidemic model of rumor diffusion in online social networks. **The European Physical Journal B**, Springer-Verlag, v. 86, n. 1, p. 29, Jan 2013. Citado nas páginas 46 e 47.

CHEW, C.; EYSENBACH, G. Pandemics in the age of twitter: Content analysis of tweets during the 2009 h1n1 outbreak. **PLOS ONE**, Public Library of Science, v. 5, n. 11, p. 1–13, 11 2010. Citado na página 23.

CHUA, A. Y. K.; TEE, C.-Y.; PANG, A.; LIM, E.-P. The retransmission of rumor-related tweets: Characteristics of source and message. In: **Proceedings of the 7th 2016 International Conference on Social Media & Society**. New York, NY, USA: ACM, 2016. (SMSociety '16), p. 22:1–22:10. ISBN 978-1-4503-3938-4. Citado na página 46.

COLLARD, M.; BRISSON, L.; COLLARD, P.; STATTNER, E. Rumor spreading modeling: Profusion versus scarcity. In: **2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)**. Paris, France: IEEE, 2015. p. 1547–1554. Citado nas páginas 25 e 45.

DALEY, D. J.; KENDALL, D. G. Stochastic rumours. **IMA Journal of Applied Mathematics**, Oxford University Press, v. 1, n. 1, p. 42–55, 1965. Citado nas páginas 16, 55, 56 e 69.

DANG, A.; MOH'D, A.; ISLAM, A.; MILIOS, E. Early Detection of Rumor Veracity in Social Media. In: **Proceedings of the 52nd Hawaii International Conference on System Sciences**. Grand Wailea, Maui: IEEE, 2019. v. 6, p. 2355–2364. ISBN 9780998133126. Citado nas páginas 25, 49 e 70.

DANG, A.; MOH'D, A.; MILIOS, E.; MINGHIM, R. What is in a rumour: Combined visual analysis of rumour flow and user activity. In: **Proceedings of the 33rd Computer Graphics International**. New York, NY, USA: ACM, 2016. (CGI '16), p. 17–20. ISBN 978-1-4503-4123-3. Citado nas páginas 15, 24, 35, 37, 51 e 52.

DANG-XUAN, L.; STIEGLITZ, S. Impact and diffusion of sentiment in political communication-an empirical analysis of political weblogs. In: **International Conference on Web and Social Media**. Dublin, Ireland: AAAI Press, 2012. p. 427–430. Citado nas páginas 54 e 69.

DIFONZO, N.; BORDIA, P. **Rumor psychology: Social and organizational approaches.** Washington, DC, US: American Psychological Association, 2007. x, 292–x, 292 p. ISBN 1-59147-426-4 (Hardcover); 978-159147-426-5 (Hardcover). Citado nas páginas 23 e 69.

ELBOGHDADY, D. **Market quavers after fake AP tweet says Obama was hurt in White House explosions**. 2013. Disponível em: <https://www.washingtonpost.com/business/economy/market-quavers-after-fake-ap-tweet-says-obama-was-hurt-in-white-house-explosions/2013/04/23/d96d2dc6-ac4d-11e2-a8b9-2a63d75b5459_story.html?utm_term=.4868bcf51f70>. Citado na página 24.

FLEISS, J. L.; COHEN, J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. **Educational and Psychological Measurement**, Sage Publications, US, v. 33, n. 3, p. 613–619, 1973. ISSN 1552-3888(Electronic),0013-1644(Print). Citado na página 31.

GEORGE, J.; JONES, G. **Understanding and Managing Organizational Behavior**. [S.l.]: Pearson Prentice Hall, 2007. ISBN 9780132057035. Citado na página 24.

GOLBECK, J.; ROBLES, C.; EDMONDSON, M.; TURNER, K. Predicting personality from twitter. In: **2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing**. Boston, MA, USA: IEEE, 2011. p. 149–156. Citado na página 34.

HASHIMOTO, T.; KUBOYAMA, T.; SHIROTA, Y. Rumor analysis framework in social media. In: **TENCON 2011 - 2011 IEEE Region 10 Conference**. Bali, Indonesia: IEEE, 2011. p. 133–137. ISSN 2159-3442. Citado nas páginas 24 e 43.

ISLAM, A.; MILIOS, E.; KESELJ, V. Text similarity using google tri-grams. In: **Proceedings of the 25th Canadian Conference on Advances in Artificial Intelligence**. Berlin, Heidelberg: Springer-Verlag, 2012. (Canadian AI'12), p. 312–317. ISBN 978-3-642-30352-4. Citado nas páginas 32 e 55.

JAWAHEER, G.; SZOMSZOR, M.; KOSTKOVA, P. Comparison of implicit and explicit feedback from an online music recommendation service. In: **Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems**. New York, NY, USA: ACM, 2010. (HetRec '10), p. 47–51. ISBN 978-1-4503-0407-8. Citado na página 34.

JOACHIMS, T. Training linear svms in linear time. In: **Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: ACM, 2006. (KDD '06), p. 217–226. ISBN 1-59593-339-5. Citado na página 27.

KERREN, A.; PURCHASE, H. C.; WARD, M. O. **Multivariate Network Visualization**. Cham: Springer International Publishing, 2014. 1–9 p. ISBN 978-3-319-06793-3. Citado nas páginas 33, 34 e 38.

KRIPPENDORFF, K. Reliability in Content Analysis. **Human Communication Research**, John Wiley & Sons, Ltd (10.1111), v. 30, n. 3, p. 411–433, jul 2004. ISSN 0360-3989. Citado na página 31.

KWON, S.; CHA, M.; JUNG, K. Rumor detection over varying time windows. **PloS one**, Public Library of Science, v. 12, n. 1, p. e0168344, 2017. Citado na página 25.

LANDESBERGER, T. von; KUIJPER, A.; SCHRECK, T.; KOHLHAMMER, J.; WIJK, J. van; FEKETE, J.-D.; FELLNER, D. Visual analysis of large graphs: State-of-the-art and future research challenges. **Computer Graphics Forum**, Blackwell Publishing Ltd, v. 30, n. 6, p. 1719–1749, 2011. ISSN 1467-8659. Citado na página 37.

LIAO, Q.; SHI, L. She gets a sports car from our donation: Rumor transmission in a chinese microblogging community. In: **Proceedings of the 2013 Conference on Computer Supported Cooperative Work**. New York, NY, USA: ACM, 2013. (CSCW '13), p. 587–598. ISBN 978-1-4503-1331-5. Citado na página 46.

LIU, M.; LIU, S.; ZHU, X.; LIAO, Q.; WEI, F.; PAN, S. An uncertainty-aware approach for exploratory microblog retrieval. **IEEE Transactions on Visualization and Computer Graphics**, v. 22, n. 1, p. 250–259, Jan 2016. ISSN 1077-2626. Citado nas páginas 37 e 41.

LIU, X.; NOURBAKHSH, A.; LI, Q.; FANG, R.; SHAH, S. Real-time rumor debunking on twitter. In: **Proceedings of the 24th ACM International on Conference on Information and Knowledge Management**. New York, NY, USA: ACM, 2015. (CIKM '15), p. 1867–1870. ISBN 978-1-4503-3794-6. Citado na página 49.

LUKASIK, M.; COHN, T.; BONTCHEVA, K. Point process modelling of rumour dynamics in social media. In: **Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)**. Beijing, China: Association for Computational Linguistics, 2015. p. 518–523. Citado na página 46.

MA, J.; GAO, W.; MITRA, P.; KWON, S.; JANSEN, B. J.; WONG, K.-F.; CHA, M. Detecting rumors from microblogs with recurrent neural networks. In: **Proceedings of the Twenty-Fifth**

International Joint Conference on Artificial Intelligence. New York, New York, USA: AAAI Press, 2016. (IJCAI'16), p. 3818–3824. ISBN 978-1-57735-770-4. Citado na página 49.

MADDOCK, J.; STARBIRD, K.; AL-HASSANI, H. J.; SANDOVAL, D. E.; ORAND, M.; MASON, R. M. Characterizing online rumoring behavior using multi-dimensional signatures. In: **Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing**. New York, NY, USA: ACM, 2015. (CSCW '15), p. 228–241. ISBN 978-1-4503-2922-4. Citado nas páginas 25 e 46.

MENDOZA, M.; POBLETE, B.; CASTILLO, C. Twitter under crisis: Can we trust what we rt? In: **Proceedings of the First Workshop on Social Media Analytics**. New York, NY, USA: ACM, 2010. (SOMA '10), p. 71–79. ISBN 978-1-4503-0217-3. Citado nas páginas 46 e 70.

MOROZOV, E. **Swine flu: Twitter's power to misinform**. 2009. Disponível em: <http://foreignpolicy.com/2009/04/25/swine-flu-twitters-power-to-misinform/>. Citado na página 23.

MURPHY, K. P. **Machine learning: a probabilistic perspective**. Cambridge, MA: The MIT Press, 2012. Citado nas páginas 24, 27 e 28.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011. Citado na página 72.

PETTY, R.; CACIOPPO, J. **Communication and persuasion: central and peripheral routes to attitude change**. [S.l.]: Springer-Verlag, 1986. (Social Psychology Series). ISBN 9783540963448. Citado na página 46.

PLATT, J. C. Using analytic qp and sparseness to speed training of support vector machines. In: **Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems II**. Cambridge, MA, USA: MIT Press, 1999. p. 557–563. ISBN 0-262-11245-0. Citado na página 27.

PROCTER, R.; VIS, F.; VOSS, A. Reading the riots on twitter: methodological innovation for the analysis of big data. **International Journal of Social Research Methodology**, v. 16, n. 3, p. 197–214, 2013. Citado nas páginas 25 e 46.

QAZVINIAN, V.; ROSENGREN, E.; RADEV, D. R.; MEI, Q. Rumor has it: Identifying misinformation in microblogs. In: **Proceedings of the Conference on Empirical Methods in Natural Language Processing**. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011. (EMNLP '11), p. 1589–1599. ISBN 978-1-937284-11-4. Citado na página 24.

RADVANSKý, M.; KUDěLKA, M.; HORáK, Z.; SNášEL, V. Visualization of social network dynamics using sammon's projection. In: **2013 Fifth International Conference on Computational Aspects of Social Networks**. Fargo, ND, USA: IEEE, 2013. p. 56–61. Citado na página 42.

RATKIEWICZ, J.; CONOVER, M.; MEISS, M.; GONçALVES, B.; FLAMMINI, A.; MENCZER, F. Detecting and tracking political abuse in social media. In: **Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM)**. Barcelona, Spain: AAAI Press, 2011. Citado na página 24.

ROBERTSON, S. Understanding inverse document frequency: On theoretical arguments for idf. **Journal of Documentation - J DOC**, v. 60, p. 503–520, 10 2004. Citado na página 32.

ROSNOW, R. L.; FOSTER, E. K. Rumor and gossip research. **Psychological Agenda**, v. 19, 01 2005. Citado na página 69.

SALTON, G.; MCGILL, M. J. **Introduction to Modern Information Retrieval**. New York, NY, USA: McGraw-Hill, Inc., 1986. ISBN 0070544840. Citado nas páginas 31 e 70.

SEIFERT, C.; KUMP, B.; KIENREICH, W.; GRANITZER, G.; GRANITZER, M. On the beauty and usability of tag clouds. In: **2008 12th International Conference Information Visualisation**. London, UK: IEEE, 2008. p. 17–25. ISSN 1550-6037. Citado na página 36.

STATISTA. **Most famous social network sites worldwide as of January 2019, ranked by number of active users (in millions)**. 2019. Disponível em: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>. Citado na página 23.

TAKAYASU, M.; SATO, K.; SANO, Y.; YAMADA, K.; MIURA, W.; TAKAYASU, H. Rumor diffusion and convergence during the 3.11 earthquake: A twitter case study. **PLOS ONE**, Public Library of Science, v. 10, n. 4, p. 1–18, 04 2015. Citado na página 24.

TRANI, S.; CECCARELLI, D.; LUCCHESE, C.; ORLANDO, S.; PEREGO, R. Dexter 2.0: An open source tool for semantically enriching data. In: **Proceedings of the 2014 International Conference on Posters & Demonstrations Track - Volume 1272**. Aachen, Germany, Germany: CEUR-WS.org, 2014. (ISWC-PD'14), p. 417–420. Citado nas páginas 33, 55 e 58.

TRIPATHY, R. M.; BAGCHI, A.; MEHTA, S. A study of rumor control strategies on social networks. In: **Proceedings of the 19th ACM International Conference on Information and Knowledge Management**. New York, NY, USA: ACM, 2010. (CIKM '10), p. 1817–1820. ISBN 978-1-4503-0099-5. Citado na página 45.

WARD, M.; GRINSTEIN, G.; KEIM, D. **Interactive Data Visualization: Foundations, Techniques, and Applications**. Natick, MA, USA: A. K. Peters, Ltd., 2010. ISBN 1568814739, 9781568814735. Citado na página 24.

WASSERMAN, S.; FAUST, K. **Social Network Analysis: Methods and Applications**. [S.l.]: Cambridge University Press, 1994. Citado nas páginas 24 e 33.

WEBB, H.; BURNAP, P.; PROCTER, R.; RANA, O.; STAHL, B. C.; WILLIAMS, M.; HOUSLEY, W.; EDWARDS, A.; JIROTKA, M. Digital wildfires: Propagation, verification, regulation, and responsible innovation. **ACM Trans. Inf. Syst.**, ACM, New York, NY, USA, v. 34, n. 3, p. 15:1–15:23, abr. 2016. ISSN 1046-8188. Citado na página 23.

WITTEN, I. H.; FRANK, E.; HALL, M. A.; PAL, C. J. **The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques**. 4th. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2016. Citado na página 71.

ZHANG, Q.; ZHANG, S.; DONG, J.; XIONG, J.; CHENG, X. Automatic detection of rumor on social network. In: **Natural Language Processing and Chinese Computing**. Berlin, Heidelberg: Springer, 2015. p. 113–122. Citado na página 25.

ZHAO, J.; CAO, N.; WEN, Z.; SONG, Y.; LIN, Y. R.; COLLINS, C. Fluxflow: Visual analysis of anomalous information spreading on social media. **IEEE Transactions on Visualization and Computer Graphics**, v. 20, n. 12, p. 1773–1782, Dec 2014. ISSN 1077-2626. Citado nas páginas 42 e 43.

ZHOU, M.; ZHANG, W.; SMITH, B.; VARGA, E.; FARIAS, M.; BADENES, H. Finding someone in my social directory whom i do not fully remember or barely know. In: **Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces**. New York, NY, USA: ACM, 2012. (IUI '12), p. 203–206. ISBN 978-1-4503-1048-2. Citado na página 34.

ZUBIAGA, A.; AKER, A.; BONTCHEVA, K.; LIAKATA, M.; PROCTER, R. Detection and resolution of rumours in social media: A survey. **CoRR**, abs/1704.00656, 2017. Citado na página 23.

ZUBIAGA, A.; LIAKATA, M.; PROCTER, R.; HOI, G. W. S.; TOLMIE, P. Analysing how people orient to and spread rumours in social media by looking at conversational threads. **PLOS ONE**, Public Library of Science, v. 11, n. 3, p. 1–29, 03 2016. Citado nas páginas 25, 44 e 46.


ICMC
USP
SÃO CARLOS