UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

SOBRE AS MÉTRICAS SUBJACENTES ÀS REDES DE DISSEMINAÇÃO DE NOTÍCIAS FALSAS EM MICROBLOGS

Felipe de Oliveira Sampaio

Orientadoras
Adriana Cesário de Faria Alvim
Vânia Maria Félix Dias

RIO DE JANEIRO, RJ – BRASIL
SETEMBRO DE 2017

SOBRE AS MÉTRICAS SUBJACENTES ÀS REDES DE DISSEMINAÇÃO DE NOTÍCIAS FALSAS EM MICROBLOGS

Felipe de Oliveira Sampaio

DISSERTAÇÃO APRESENTADA COMO REQUISITO PARCIAL PARA OBTENÇÃO DO TÍTULO DE MESTRE PELO PROGRAMA DE PÓSGRADUAÇÃO EM INFORMÁTICA DA UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO (UNIRIO). APROVADA PELA COMISSÃO EXAMINADORA ABAIXO ASSINADA.

Aprovada por:

_____

Adriana Cesário de Faria Alvim, Dsc. - UNIRIO

_____

Vânia Maria Félix Dias, Dsc. - UNIRIO

_____

Simone Bacellar Leal Ferreira, Dsc. - UNIRIO

_____

Paulo Eustáquio Duarte Pinto, Dsc. - UERJ

RIO DE JANEIRO, RJ – BRASIL
SETEMBRO DE 2017

Sampaio, Felipe de Oliveira.
Sxxx Sobre as métricas subjacentes às redes de disseminação de notícias falsas em microblogs. / Felipe Sampaio. - 2017.
xx f.

Orientadoras: Profª. Drª Adriana C.F. Alvim e Profª. Drª Vânia M. Félix Dias.
Dissertação (Mestrado em Informática) – Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2017.

1. Redes Sociais – Microblogs – Notícias falsas. I. Alvim, Adriana C.F.. II. Dias, Vânia M.F. III. Universidade Federal do Estado do Rio de Janeiro. IV. Análise da Influência das Redes Tradicionais de Comunicação Brasileiras nas Redes Sociais de Microblogs.

À minha família.

AGRADECIMENTOS

Um trabalho de dissertação sempre envolve alguma dificuldade natural, por ser considerado um dos primeiros contatos com a metodologia científica e por abordar temas que muitas vezes não são de familiaridade do aluno. Essas dificuldades são mais desafiantes quando se somam a outras de ordem pessoal que, pela natureza da vida, acontecem sem aviso prévio.

Agradeço primeiramente à minha esposa, pela paciência em preservar meus momentos de estudo mesmo com as demandas familiares que são parte do bom convívio.

Agradeço também, especialmente, às minhas orientadoras Adriana e Vânia, por sua paciência e dedicação em direcionar corretamente os trabalhos, principalmente quando estavam em estágio inicial e necessitavam de muitos ajustes de rota para se enquadrarem num padrão adequado de qualidade.

Agradeço também a todos do Programa de Pós-graduação da UNIRIO por investirem importantes recursos na minha formação. O que aprendi, além de fazer parte de mim em definitivo, será certamente utilizado em benefício de todos aos quais puder ajudar de agora em diante.

RESUMO

O presente trabalho tem como objetivo investigar os mecanismos de disseminação de notícias falsas em redes sociais de microblogs. As notícias falsas são um problema antigo, possivelmente desde o próprio nascimento da comunicação entre indivíduos, entretanto os avanços na tecnologia aumentam seu potencial de danos. Para explorar os mecanismos de disseminação de notícias falsas, foram coletadas postagens da rede social Twitter relacionadas a um grupo de dez dessas notícias, divulgadas entre março e junho de 2017 no Brasil. A partir dessas postagens foram utilizadas métricas da teoria dos grafos e análise de redes sociais que trouxeram um esclarecimento das questões propostas. Com relação às origens da disseminação das notícias falsas, não foi possível identificar um indivíduo ou grupo específico que tivesse responsabilidade frequente na divulgação das notícias falsas selecionadas. As redes representativas das postagens, em geral, demonstraram ter modelo similar ao previsto para as redes de livre escala. A análise da evolução da audiência atingida pelas postagens demonstrou claramente os riscos envolvidos, mostrando em geral um crescimento muito rápido do número de indivíduos alcançados, proporcionado pelo poder das interconexões entre os usuários do Twitter.

Palavras-chave: notícias falsas, análise de redes sociais, Twitter.

ABSTRACT

The present work aims to investigate the mechanisms of dissemination of false news in social networks of microblogs. False news is an old problem, possibly since the very birth of communication between individuals, however advances in technology increase its potential for harm. In order to explore the mechanisms for the dissemination of false news, postings from Twitter were collected related to a group of ten of these news, published between March and June 2017 in Brazil. From these posts were used graph theory metrics and social networks analysis that brought a clarification of the proposed questions. With regard to the origins of the dissemination of false news, it was not possible to identify a specific individual or group that had frequent responsibility for disclosing the selected false news stories. The representative networks of postings have generally been shown to have a model similar to that foreseen for scale-free networks. The analysis of the evolution of the audience reached clearly demonstrated the risks involved, showing in general a very rapid growth in the number of individuals reached, provided by the connections among Twitter users.

Keywords: false news, social networks analysis, Twitter.

ÍNDICE

I - LISTA DE FIGURAS 9

II - LISTA DE EQUAÇÕES 12

III - LISTA DE TABELAS 13

1. Introdução 15

1.1. Objetivo 17

1.2. Estrutura do trabalho 18

2. Revisão Teórica e Bibliográfica 20

2.1. Teoria dos Grafos nas Redes Sociais 20

2.2. Análise de Redes Sociais 22

2.3. Modelos de Redes 28

2.4. Revisão Bibliográfica 31

3. Metodologia 37

3.1. Seleção de notícias falsas 37

3.2. Coleta dos dados 37

3.3. Preparação dos dados 41

3.3.1. Planilha “Arestas” 41

3.3.2. Planilha “Vértices” 42

3.3.3. Postagens órfãs e duplicadas 42

3.4. Desenho dos grafos e métricas analisadas 43

4. Análise dos Dados 47

4.1. Notícia: Vacina da febre amarela tem veneno mortal 47

4.2. Notícia: Morte do cantor Arlindo Cruz 50

4.3. Notícia: Taxista que xingou Jesus morre em acidente 54

4.4. Notícia: Tapioca pode matar 58

4.5. Notícia: Pink Floyd anuncia turnê no Brasil 62

4.6. Notícia: Gritaram “Prendam Lula!” na encenação da peça teatral “A Paixão de Cristo”..... 66
4.7. Notícia: Esposa do juiz Sérgio Moro está ligada a esquema de corrupção na APAE..... 70
4.8. Notícia: Delegado que investigava a morte de Teori Zavascki foi assassinado.. 75
4.9. Notícia: Homossexuais terão cota em concursos públicos..... 79
4.10. Notícia: Andreas Von Richthofen encontrado na “Cracolândia”..... 84
4.11. Análise do conjunto das notícias selecionadas..... 89
5. Conclusões..... 96
REFERÊNCIAS 101

I - LISTA DE FIGURAS

Figura 1 – Diâmetro de um grafo (fonte: autor). 21
Figura 2 – Grafo com duas componentes conexas (fonte: autor). 22
Figura 3 – A maior centralidade de grau dos vértices A e F não espelha a importante influência do vértice I para a interconexão da rede (fonte: autor). 23
Figura 4 – Componente máxima (vértices A a G) e vértices isolados (vértices M e N) (fonte: autor). 25
Figura 5 – Exemplos de cálculo de coeficiente de agrupamento local (Coef) para o vértice A das redes hipotéticas (a), (b), (c) e (d) (fonte: autor). 25
Figura 6 – Mesmo grupo de vértices conectados com algoritmo que (a) insere arestas randomicamente até que uma quantidade *n* seja atingida, formando um modelo de rede randômica, e (b) insere novos vértices sequencialmente conectando-os com os vértices já existentes na rede dando maior probabilidade de escolha para os de maior grau, formando um modelo de rede do mundo real (CALLAWAY et. al., 2000). 29
Figura 7 – Exemplo de histograma de distribuiçãode graus (LAMBERSON, 2015). ... 30
Figura 8 – Desenho da rede total da coleta “vacina febre amarela veneno mortal” (fonte: autor)..... 48
Figura 9 – Distribuição de graus de entrada e saída da coleta “vacina febre amarela veneno mortal” (fonte: autor). 49
Figura 10 – Gráfico de evolução da audiência da coleta “vacina febre amarela veneno mortal” (fonte: autor). 50
Figura 11 – Desenho da rede total da coleta “Arlindo Cruz morre” (fonte: autor). 51
Figura 12 – Distribuição de graus de entrada e saída da coleta “Arlindo Cruz morre” (fonte: autor). 52
Figura 13 – Gráfico de evolução da audiência da coleta “Arlindo Cruz morre” (fonte: autor)..... 53
Figura 14 – Desenho da rede total da coleta “Taxista que xingou Jesus morre em acidente” (fonte: autor). 55
Figura 15 – Distribuição de graus de entrada e saída da coleta “Taxista que xingou Jesus morre em acidente” (fonte: autor). 56
Figura 16 – Gráfico de evolução da audiência da coleta “Taxista que xingou Jesus morre em acidente” (fonte: autor). 57

Figura 17 – Desenho da rede total da coleta “Tapioca pode matar” (fonte: autor). 59
Figura 18 – Distribuição de graus de entrada e saída da coleta “Tapioca pode matar” (fonte: autor). 60
Figura 19 – Gráfico de evolução da audiência da coleta “Tapioca pode matar” (fonte: autor)..... 61
Figura 20 – Desenho da rede total da coleta “Pink Floyd anuncia turnê Brasil” (fonte: autor)..... 63
Figura 21 – Distribuição de graus de entrada e saída da coleta “Pink Floyd anuncia turnê Brasil” (fonte: autor)..... 64
Figura 22 – Gráfico de evolução da audiência da coleta “Pink Floyd anuncia turnê Brasil” (fonte: autor)..... 65
Figura 23 – Desenho da rede total da coleta “Gritaram Prendam Lula Paixão de Cristo” (fonte: autor). 67
Figura 24 – Distribuição de graus de entrada e saída da coleta “Gritaram Prendam Lula Paixão de Cristo” (fonte: autor). 68
Figura 25 – Gráfico de evolução da audiência da coleta “Gritaram Prendam Lula Paixão de Cristo” (fonte: autor)..... 69
Figura 26 – Desenho da rede total da coleta “Mulher Moro esquema corrupção APAE” (fonte: autor). 71
Figura 27 – Distribuição de graus de entrada e saída da coleta “Mulher Moro esquema corrupção APAE” (fonte: autor). 73
Figura 28 – Gráfico de evolução da audiência da coleta “Mulher Moro esquema corrupção APAE” (fonte: autor). 74
Figura 29 – Desenho da rede total da coleta “delegado inquérito Teori Zavascki assassinado” (fonte: autor). 76
Figura 30 – Distribuição de graus de entrada e saída da coleta “delegado inquérito Teori Zavascki assassinado” (fonte: autor). 77
Figura 31 – Gráfico de evolução da audiência da coleta “delegado inquérito Teori Zavascki assassinado” (fonte: autor). 78
Figura 32 – Desenho da rede total da coleta “Homossexuais terão cota concursos” (fonte: autor). 80
Figura 33 – Distribuição de graus de entrada e saída da coleta “Homossexuais terão cota concursos” (fonte: autor). 82

Figura 34 – Gráfico de evolução da audiência da coleta “Homossexuais terão cota concursos” (fonte: autor). 83
Figura 35 – Desenho da rede total da coleta “Andreas Von Richthofen encontrado Cracolândia” (fonte: autor). 85
Figura 36 – Distribuição de graus de entrada e saída da coleta “Andreas Von Richthofen encontrado Cracolândia” (fonte: autor). 87
Figura 37 – Gráfico de evolução da audiência da coleta “Andreas Von Richthofen encontrado Cracolândia” (fonte: autor). 88
Figura 38 – Gráfico com a audiência final atingida pelas notícias falsas selecionadas (fonte: autor). 89
Figura 39 – Modelo de desenho observado nas redes representativas das notícias falsas analisadas, feito a partir da rede da coleta “Vacina febre amarela veneno mortal” (fonte: autor)..... 91

II - LISTA DE EQUAÇÕES

Equação 1 – fórmula para o cálculo da centralidade de intermediação (BRANDES, 2001). 23
Equação 2 – fórmulas para o cálculo da centralidade de proximidade, em suas versões não-normalizada (a) e normalizada (b). BAVELAS (1950)..... 24
Equação 3 – fórmula para o cálculo do coeficiente de agrupamento local para grafos direcionados (WATTS e STROGATZ, 1998)..... 25
Equação 4 – fórmula para o cálculo do coeficiente de agrupamento local para grafos não direcionados (WATTS e STROGATZ, 1998)..... 26
Equação 5 – fórmula para o cálculo da centralidade de autovetor (fonte: autor). 26
Equação 6 – fórmula para o cálculo da métrica Pagerank (CHO e ROY, 2004). 27
Equação 7 – padrão de formação das probabilidades na distribuição de grau de redes de livre escala. 30

III - LISTA DE TABELAS

Tabela 1 – Quadro resumo com os estudos relacionados aos temas abordados nesta dissertação. 36
Tabela 2 – Métricas básicas da rede total da coleta “vacina febre amarela veneno mortal”. 47
Tabela 3 – Usuários mais influentes por ordem de PageRank da coleta “vacina febre amarela veneno mortal”. 50
Tabela 4 – Métricas básicas da rede total da coleta “Arlindo Cruz morre”. 51
Tabela 5 – Usuários mais influentes por ordem de PageRank da coleta “Arlindo Cruz morre”. 54
Tabela 6 – Métricas básicas da rede total da coleta “Taxista que xingou Jesus morre em acidente”. 54
Tabela 7 – Usuários mais influentes por ordem de PageRank da coleta “Taxista que xingou Jesus morre em acidente”. 58
Tabela 8 – Métricas básicas da rede total da coleta “Tapioca pode matar”. 58
Tabela 9 – Usuários mais influentes por ordem de PageRank da coleta “Tapioca pode matar”. 61
Tabela 10 – Métricas básicas da rede total da coleta “Pink Floyd anuncia turnê Brasil”. 62
Tabela 11 – Usuários mais influentes por ordem de PageRank da coleta “Pink Floyd anuncia turnê Brasil”. 65
Tabela 12 – Métricas básicas da rede total da coleta “Gritaram Prendam Lula Paixão de Cristo”. 66
Tabela 13 – Usuários mais influentes por ordem de PageRank da coleta “Gritaram Prendam Lula Paixão de Cristo”. 70
Tabela 14 – Métricas básicas da rede total da coleta “Mulher Moro esquema corrupção APAE”. 70
Tabela 15 – Usuários mais influentes por ordem de PageRank da coleta “Mulher Moro esquema corrupção APAE”. 74
Tabela 16 – Métricas básicas da rede total da coleta “delegado inquérito Teori Zavascki assassinado”. 75
Tabela 17 – Usuários mais influentes por ordem de PageRank da coleta “delegado inquérito Teori Zavascki assassinado”. 79

Tabela 18 – Métricas básicas da rede total da coleta “Homossexuais terão cota concursos”. 79
Tabela 19 – Usuários mais influentes por ordem de PageRank da coleta “Homossexuais terão cota concursos”..... 83
Tabela 20 – Métricas básicas da rede total da coleta “Andreas Von Richthofen encontrado Cracolândia”. 84
Tabela 21 – Usuários mais influentes por ordem de PageRank da coleta “Andreas Von Richthofen encontrado Cracolândia”..... 88
Tabela 22 – Resumo das coletas associadas às notícias falsas selecionadas..... 87
Tabela 23 – Coletas ordenadas pela razão entre audiência final e quantidade de arestas das redes representativas das coletas. 90
Tabela 24 – Resumo das notícias falsas selecionadas para estudo..... 97

1. Introdução

A disseminação de notícias falsas é antiga e sua ocorrência remonta a passado distante, desde a formação das primeiras sociedades, como explorado por SHERMER (2011) no livro *The Believing Brain*, que procura explicar os motivos que nos levam a acreditar em informações sem base nos fatos:

> Nós formamos nossas crenças por uma variedade de razões subjetivas, pessoais, emocionais e psicológicas no contexto dos ambientes criados pela família, amigos, colegas, cultura e sociedade como um todo; depois de formadas nossas crenças nós a defendemos, justificamos e a racionalizamos com uma série de razões intelectuais, argumentos convincentes e explicações racionais. Crenças vêm primeiro, explicações para as crenças vêm depois. Eu chamo esse processo de "realismo crença-dependente", onde nossas percepções da realidade são dependentes das crenças que temos sobre elas. A realidade existe independentemente da mente humana, mas nosso entendimento sobre ela depende das crenças que mantemos em qualquer época. (SHERMER, 2011, p. 5. Tradução do autor).

Os problemas trazidos pela divulgação de notícias falsas, apesar de representarem enormes desafios para as sociedades desde tempos remotos, tiveram um maior destaque com a repercussão recente do termo *post-truth* (pós-verdade), eleito como a palavra do ano de 2016 pelo mundialmente conhecido dicionário Oxford da língua inglesa (OXFORD, 2016).

Apesar do termo pós-verdade ter alguma relação com a expressão "notícia falsa" quando se considera que ambos procuram caracterizar os casos em que uma divulgação de informação é feita desconectada de fatos e sem compromisso com a verdade, a diferenciação entre eles está no fato do primeiro ter um objetivo direcionado à formação de opinião na sociedade. A notícia falsa, por sua vez, pode não ter um objetivo claramente definido, ou ter um objetivo desconectado de pretensões de formação de opinião, sendo inclusive comum ver notícias falsas sendo usadas como um perigoso divertimento.

A disseminação de notícias falsas encontra terreno fértil nas *redes sociais,* expressão utilizada neste estudo para representar os sistemas de interações entre usuários criados no ambiente da internet, onde informações sem amparo nos fatos podem prosperar com relativa facilidade, muito devido à agilidade na troca de informações entre indivíduos observada em nossos dias (GOOGLE, 2016). Se, antigamente, notícias falsas demoravam atingir boa parte da população, e na grande maioria das vezes eram limitadas geograficamente, hoje obstáculos de tempo e espaço são pouco perceptíveis.

Outra característica das redes sociais que auxilia a divulgação de notícias falsas é que todo novo usuário, seja uma pessoa física ou um grande grupo de comunicação, é criado na rede com direitos iguais com relação à sua capacidade de disseminação de comentários. A aceitação ou não do conteúdo que esses usuários novatos irão criar provocará ou não a referência de outros usuários da rede a esse conteúdo, dizendo-se neste caso que houve uma conexão entre esses usuários. Essa conexão pode acontecer na prática através da citação, resposta ou repasse do conteúdo original de um usuário por outro usuário. Esse potencial de divulgação de cada usuário, portanto, se traduzirá principalmente no número de conexões estabelecidas a esse usuário, a depender unicamente do interesse dos outros usuários em criar essas conexões. Seguindo esse raciocínio, uma pessoa física, por exemplo, pode ter uma capacidade maior de atingir um grande público que uma multinacional já estabelecida no mercado (PRIMO, 2011; MARCHI, 2012). As redes sociais, sob esse ponto de vista, podem ser consideradas um ambiente democrático e igualitário quando se considera o igual nível de condições que dá a todos os usuários participantes.

Tal igualdade de potencial de divulgação dos usuários nas redes sociais, por mais que seja muito benéfica à sociedade sob diversos aspectos, traz um possível efeito colateral de tornar um emissor costumeiro de notícias falsas num divulgador tão ou mais eficiente que um renomado canal jornalístico. O processo pelo qual uma notícia falsa se dissemina nas redes sociais – de forma geral, mas não única – consistiria em uma primeira divulgação dessa notícia por uma postagem de comentário, algumas vezes fazendo referência a um link externo a um site onde a notícia foi divulgada. O processo continua com o repasse dessa informação falsa a outros usuários que, por sua vez, repetem o ciclo. É fácil constatar como isso pode levar a um crescimento exponencial do número de pessoas recebedoras desse conteúdo em um curto espaço de tempo. Assumindo o pressuposto de que esse mecanismo de disseminação através dos usuários

da rede social seja real, os consumidores das notícias falsas acabariam como fundamentais participantes no processo de disseminação desse tipo de informação. O presente estudo procura esclarecer em que grau essa participação acontece, identificando que existem lacunas de conhecimento a preencher para compreender melhor como se dão os mecanismos de disseminação de notícias falsas nas redes sociais, tema central desse trabalho.

1.1. Objetivo

Essa dissertação aborda especificamente as notícias falsas divulgadas nas redes sociais de microblogs, que serão caracterizadas no contexto desse estudo como notícias recentes, com conteúdo comprovadamente desconectado da realidade ou com incorreções graves que podem levar o leitor a interpretações errôneas sobre os fatos, que tenham sido esclarecidas por posterior checagem em fontes externas confiáveis.

A escolha das redes sociais de microblogs como universo de pesquisa para esse estudo exploratório vem de sua característica principal: espelhar, com relativa rapidez, o pensamento de um número considerável de usuários sobre diversos temas, em textos que, limitados por uma pequena quantidade de palavras, podem ser mais bem analisados de forma automatizada. Algumas das potenciais análises dos conteúdos de postagens, largamente usadas nessa pesquisa, envolvem a identificação de URLs e palavras-chave no texto.

Dentro do universo de aplicativos de rede social de microblogs, a seleção do Twitter como fonte de dados para esse estudo se dá, naturalmente, por sua grande popularidade no Brasil, sendo a referência em nosso país e também no resto do mundo quando se trata de redes sociais com característica de microblog.

Com o objetivo de avaliar os mecanismos de disseminação de um grupo de notícias falsas nas postagens divulgadas em uma rede social de microblogs como o Twitter, o presente trabalho de mestrado propõe explorar três questões principais. Considerando-se uma seleção de notícias falsas postadas no Twitter em um determinado período, aqui chamada simplesmente de conjunto de notícias falsas, deseja-se entender:

- No recorte temporal observado, quais as origens da divulgação deste conjunto de notícias falsas postadas no Twitter?

    - Para as notícias selecionadas no passo anterior, quais perfis de usuário iniciam a disseminação das respectivas notícias e, deste grupo de perfis, quais são os mais referenciados na rede?
- Como é o processo de disseminação deste conjunto de notícias falsas postadas no Twitter?
    - Quais as características da rede formada pelos perfis de usuário que disseminam este grupo de notícias falsas?
    - Com que velocidade a disseminação dessas notícias falsas acontece?
- É possível caracterizar as redes deste conjunto de notícias falsas?
    - Qual sua topologia?
    - Como se comparam as topologias das redes formadas pela disseminação dessas notícias falsas com as topologias previstas em estudos anteriores que abordam redes sociais?

Tais objetivos serão buscados por meio da análise de conjuntos de dados formados a partir da coleta de postagens da rede social Twitter. Essas coletas serão feitas a partir de buscas utilizando ferramenta específica, abordada no Capítulo 3, com filtros formados por expressões textuais que representarão as notícias falsas estudadas.

Para o cálculo de algumas das métricas necessárias para esse estudo, os resultados da coleta serão representados como um *grafo*, representação matemática muito utilizada para representar relações entre atores (indivíduos ou instituições). A literatura matemática relacionada ao estudo dos grafos evoluiu bastante nos últimos anos e permitiu o desenvolvimento de uma nova área de Análise de Redes Sociais - ARS.

1.2. Estrutura do trabalho

Além do primeiro capítulo, referente a essa introdução, onde é oferecida uma visão geral sobre as questões endereçadas por esse trabalho de mestrado e sobre a proposta de abordagem escolhida para tratar tais questões, o Capítulo 2 fornece uma fundamentação sobre a Teoria dos Grafos e sobre sua utilização no estudo de redes sociais, com indicadores específicos criados para facilitar o entendimento de conjuntos de dados que representem redes de relações.

Detalhes sobre a metodologia adotada neste trabalho serão fornecidos no Capítulo 3, no qual serão explicados os passos seguidos para a preparação prévia dos dados e a estratégia de análise, com as métricas escolhidas para esclarecer as questões propostas.

No Capítulo 4 será feita uma análise detalhada de cada um dos universos de dados gerados pelas coletas, examinando-se os resultados obtidos com os cálculos dos indicadores pertinentes. Ao final, será feita uma análise global de todas as redes representantes das coletas realizadas, explorando as questões levantadas como objetivo desse estudo.

Finalmente, no Capítulo 5, as conclusões finais obtidas com base na análise dos dados serão abordadas e documentadas.

2. Revisão Teórica e Bibliográfica

Nesse capítulo é feita uma revisão do conhecimento teórico necessário para a compreensão da metodologia utilizada neste trabalho e uma revisão bibliográfica com um levantamento de estudos acadêmicos que abordam temas associados ao proposto nessa dissertação. Na primeira seção é feita uma introdução à Teoria dos Grafos, que fornece fundamentação matemática para o estudo das redes. Na segunda seção é apresentada a Análise de Redes Sociais juntamente com suas principais métricas associadas dentro do presente trabalho. Na terceira seção são apresentados os modelos de redes previstos em estudos anteriores. Por fim, apresenta-se uma revisão bibliográfica sobre alguns trabalhos relacionados ao tema dessa dissertação.

2.1. Teoria dos Grafos nas Redes Sociais

Nessa seção serão apresentados alguns elementos da Teoria dos Grafos, utilizada como fundamentação teórica neste trabalho. As definições e métricas abordadas serão eventualmente seguidas por comentários sobre sua utilização aplicada no estudo das redes sociais.

Um *grafo* G=(V,E) é um conjunto finito não-vazio V e um conjunto E de pares não ordenados de elementos distintos de V, em que |V|=n e |E|=m. Quando necessário, se utiliza o termo *grafo não direcionado* para designar um grafo. Os *vértices* são os elementos de V e as *arestas* os elementos de E. Cada aresta e $\in$ E é denotada pelo par de vértices e = (v,w) que a forma. Os vértices v e w são os extremos da aresta *e*, sendo denominados *adjacentes*. Quando os pares de E são ordenados o grafo G é denominado *direcionado.* Nesse caso, cada aresta (v,w) possui uma única direção de v para w, e diz-se também que a aresta (v,w) sai de v e entra em w. Um *multigrafo* é um grafo que

possui múltiplas arestas entre o mesmo par de vértices e, além disso, admite a existência de *laços*, isto é, arestas do tipo e = (v,v).

Em um grafo não direcionado, o *grau* de um vértice v ∈ V corresponde ao número de vértices adjacentes a v. Vértices de grau zero recebem a denominação de *isolados*. Em um grafo direcionado o *grau de entrada* do vértice v é o número de arestas que entram em v e o *grau de saída* é o número de arestas que saem de v. Um *caminho* do vértice v_1 ao v_k é uma sequência finita de *k* vértices $v_{1,\ldots,}v_{k-1},v_k \in V$, $k \geq 1$, tal que $(v_j, v_{j+1}) \in E$, $1 \leq j \leq |k-1|$. O valor k-1 é o *comprimento* do caminho. Um *ciclo* é um caminho $v_{1,\ldots,}v_k,v_{k+1}$ onde $v_1 = v_{k+1}$ e $k \geq 3$. Um *triângulo* é um ciclo de comprimento três.

A *distância* entre um par de vértices *v* e *w* de um grafo é equivalente ao comprimento do menor caminho entre *v* e *w*. A *excentricidade* de um vértice *v* é a maior distância entre os pares formados por esse vértice e cada um dos outros vértices do grafo. O *diâmetro* de um grafo G é igual à maior de suas excentricidades, ou seja, representa a maior distância entre todos os pares de vértices do grafo.

Na Figura 1 mostra um exemplo de cálculo do diâmetro é obtido a partir da maior excentricidade identificada, na distância entre os vértices *A* e *C* (três).

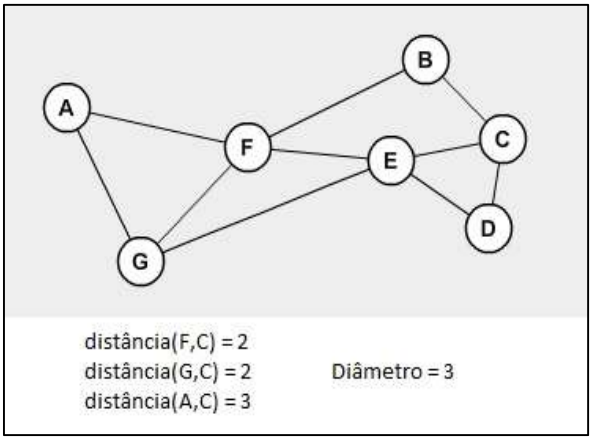

Figura 1 – Diâmetro de um grafo (fonte: autor).

Um *subgrafo* $G_2=(V_2,E_2)$ de um grafo $G_1=(V_1,E_1)$ é um grafo tal que $V_2 \subseteq V_1$ e $E_2 \subseteq E_1$. Um grafo é denominado *conexo* quando existe caminho entre cada par de vértices, caso contrário o grafo é dito *desconexo*. Uma *componente conexa* de um grafo G é um subgrafo conexo maximal de G. O adjetivo *maximal* utilizado aqui significa que, sendo S' um subgrafo conexo maximal, não pode existir outro subgrafo conexo tal que S'' ⊃ S'.

Na Figura 2 é mostrado um exemplo de um grafo com duas componentes conexas.

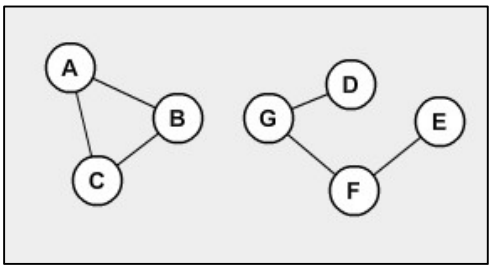

Figura 2 – Grafo com duas componentes conexas (fonte: autor).

Um *grafo completo* é um grafo tal que todos os pares de vértices são adjacentes. Logo, se G=(V,E) é um grafo completo (não orientado), então m é igual a n(n-1)/2.

A *densidade* de um grafo não orientado é igual 2m/n(n-1). Quando o grafo é direcionado, a densidade é definida por m/n(n-1).

2.2. Análise de Redes Sociais

Análise de Redes Sociais (ARS) é um campo de pesquisa originado da sociologia que faz um estudo metódico das redes sociais utilizando como fundamentação matemática a Teoria dos Grafos, criando alguns conceitos que procuram melhor definir a importância relativa de um ator dentro da rede social estudada (WASSERMAN e FAUST, 1994). Esses novos conceitos foram incorporados por modernas ferramentas de análise, como, por exemplo, o NodeXL (SOCIALMEDIA, 2014), utilizada para esse estudo, e serão mais bem explicados adiante.

Dentre os novos conceitos definidos para a ARS, a *centralidade* de um vértice *v* é uma medida que procura quantificar a importância relativa de um vértice dentro do grafo analisado (FREEMAN, 1979). Existem quatro variações principais dessa medida, diferindo entre si de acordo com o enfoque sob o qual se avalia a centralidade do vértice: *centralidade de grau, centralidade de intermediação, centralidade de proximidade e centralidade de autovetor*.

A *centralidade de grau* é a variação conceitualmente mais simples e equivale ao *grau* de um vértice *v*, definição já introduzida na primeira parte desse Capítulo 2. Seu cálculo é considerado trivial, dependendo apenas da contagem do número de ligações que esse vértice possui com outros vértices vizinhos diretos. É considerada, portanto, uma métrica *local*, no sentido de não ser necessário varrer toda a rede para calcular seu

valor. Nos grafos direcionados, pode ser calculada separadamente como *grau de entrada*, que representa a quantidade de arestas que terminam em um vértice *v*. De forma análoga, o *grau de saída* é a quantidade de arestas que saem de um vértice *v*.

Quando se pretende avaliar a influência de um vértice na rede, é possível que a centralidade de grau não espelhe corretamente a importância do vértice em questão, como quando se deseja identificar os pontos críticos de conexão entre sub-redes, como exemplificado na Figura 3.

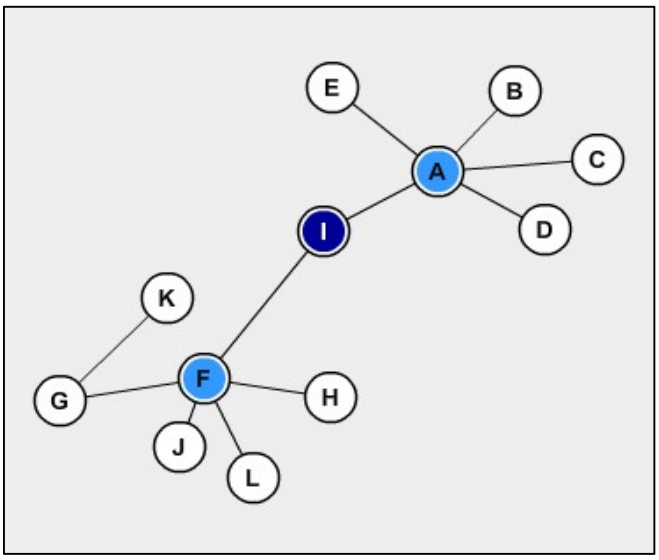

Figura 3 – A maior centralidade de grau dos vértices A e F não espelha a importante influência do vértice I para a interconexão da rede (fonte: autor).

Para atender essa necessidade, outras variantes da métrica de centralidade foram propostas para levar em consideração a rede como um todo no cálculo da importância relativa de cada vértice para a rede. Cada uma dessas métricas possui fórmula específica com o objetivo de definir um enfoque sob o qual a importância de cada vértice será estipulada.

A *centralidade de intermediação* é a quantidade de vezes em que um vértice *v* faz parte do caminho mais curto ligando os outros pares de vértices do grafo conectado, ou seja, a importância desse vértice como "ponte" de comunicação no grafo. Sua fórmula de cálculo mais aceita foi proposta por BRANDES (2001) e definida pela equação:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Equação 1 – fórmula para o cálculo da centralidade de intermediação (BRANDES, 2001).

Onde *v* é um vértice qualquer do grafo e *s* e *t* são pares de vértices diferentes formados a partir do conjunto V de vértices do grafo, com exceção do próprio vértice *v*. σ_{st} é a quantidade de caminhos curtos entre *s* e *t* e $\sigma_{st}(v)$ é a quantidade desses caminhos curtos que passam pelo vértice *v*. A centralidade de intermediação $C_B(v)$ é calculada como a soma de todas as razões entre $\sigma_{st}(v)$ e σ_{st}, para todos os vértices do grafo.

Centralidade de proximidade, em grafos conectados, é uma métrica de vértice calculada a partir da soma das distâncias entre esse vértice e todos os outros vértices do grafo. Portanto, quanto mais centralizado estiver o vértice, mais próximo estará de todos os outros vértices do grafo.

A centralidade de proximidade foi definida por BAVELAS (1950) como o inverso da soma das distâncias entre um vértice *x* e todos os outros vértices do grafo (Equação 2.a).

Quando se trata da centralidade de proximidade, é comum se referir à sua forma normalizada, que representa a média das distâncias no lugar da soma. É geralmente dada pela fórmula anterior multiplicada por *N-1*, onde *N* é o número total de vértices do grafo. Para grafos com maior número de vértices, o *-1* é eliminado da fórmula por ter impacto irrelevante no resultado (Equação 2.b).

a) $$C(x) = \frac{1}{\sum_y d(y,x)}.$$ b) $$C(x) = \frac{N}{\sum_y d(y,x)}.$$

Equação 2 – fórmulas para o cálculo da centralidade de proximidade, em suas versões não-normalizada (a) e normalizada (b). BAVELAS (1950).

Esse ajuste permite a avaliação de vértices de grafos de diferentes tamanhos.

A *componente (conexa) máxima* representa a componente conexa com maior número de vértices identificada no grafo analisado. Na Figura 4 é mostrado um exemplo de grafo com a componente máxima agrupando os vértices de A a G. Os vértices de M e N desse mesmo exemplo são considerados isolados. Os vértices de H a L formam outra componente conexa do grafo.

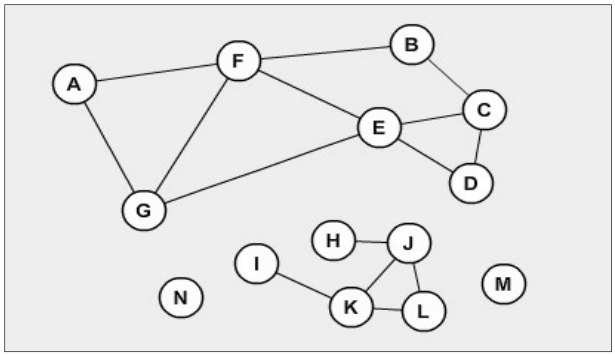

Figura 4 – Componente máxima (vértices A a G) e vértices isolados (vértices M e N) (fonte: autor).

O *coeficiente de agrupamento (clustering coefficient)* é outra métrica comumente utilizada para a análise de redes sociais. Seu objetivo é quantificar a tendência dos vértices da rede de se agruparem. Existem duas versões principais dessa métrica:

O *coeficiente de agrupamento local* se propõe a determinar o quão próximo os vértices vizinhos de um vértice *v* estão de formarem um grafo completo. Seu cálculo foi proposto por WATTS e STROGATZ (1998) e consiste na proporção de ligações entre os vértices da sua vizinhança, dividido pelo número de ligações que poderiam existir entre estes, como mostrado no exemplo da Figura 5:

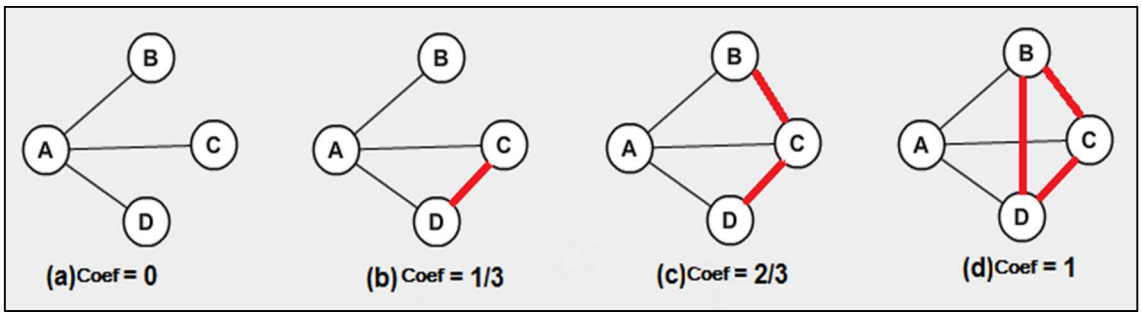

Figura 5 – Exemplos de cálculo de coeficiente de agrupamento local (Coef) para o vértice A das redes hipotéticas (a), (b), (c) e (d) (fonte: autor).

O cálculo do coeficiente de agrupamento local para grafos direcionados é calculado pela fórmula:

$$C_{oef} = \frac{|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i\,(k_i - 1)}$$

Equação 3 – fórmula para o cálculo do coeficiente de agrupamento local para grafos direcionados (WATTS e STROGATZ, 1998).

Onde C_{oef} é o coeficiente de agrupamento local de um vértice *i*, e_{jk} representa a quantidade total de arestas conectando vértices *j* e *k* pertencentes ao conjunto de vértices vizinhos de *i* (*Ni*), por sua vez subconjunto do conjunto total de arestas do grafo (*E*); e a expressão $k_i\ (k_i - 1)$ representa a quantidade máxima de arestas – em ambos os sentidos – interligando os vértices pertencentes a *Ni.*

Para os grafos não direcionados a fórmula é ajustada considerando que o número máximo de arestas ligando os vértices do conjunto *Ni* passa a ser $(k_i\ (k_i - 1)) / 2$:

$$C_{oef} = \frac{2\,|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i\,(k_i - 1)}$$

Equação 4 – fórmula para o cálculo do coeficiente de agrupamento local para grafos não direcionados (WATTS e STROGATZ, 1998).

O *coeficiente de agrupamento global* busca usar metodologia semelhante ao cálculo do coeficiente de agrupamento local, mas com enfoque na rede como um todo. A métrica pode ser calculada como a média de todos os coeficientes de agrupamentos locais dos vértices do grafo (WATTS e STROGATZ, 1998) ou através de outra fórmula proposta por LUCE e PERRY (1949), que a define como o coeficiente entre o número de trios de vértices conectados sobre o número de trios conectados da rede.

C*entralidade de autovetor* é uma medida que quantifica a importância de um vértice *v* dentro da rede, atribuindo a *v* uma pontuação baseada na quantidade de arestas direcionadas que chegam a *v* e nas pontuações dos vértices de origem dessas arestas. Existem várias fórmulas e algoritmos correspondentes para o cálculo dessa centralidade, em geral baseadas na seguinte fórmula base:

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t$$

Equação 5 – fórmula para o cálculo da centralidade de autovetor (fonte: autor).

Onde x_v é a centralidade de um vértice *v* pertencente ao conjunto V de vértices do grafo G, λ é uma constante e M(*v*) é o conjunto de vértices vizinhos a *v* e $a_{v,t}$ é a matriz de adjacências, onde $a_{v,t} = 1$ se os vértices *v* e *t* estão conectados e $a_{v,t} = 0$ se não estiverem.

Uma variante famosa da centralidade de autovetor é o *Pagerank,* que em sua origem foi proposto para resolver o problema de quantificar a importância de um site da internet.

A fama da métrica Pagerank foi gerada, principalmente, pela implementação proposta no algoritmo homônimo utilizado no site de buscas mais usado na internet – o Google (www.google.com). Esse algoritmo funciona de forma recursiva, calculando a importância de um site primário somando as importâncias dos sites secundários que fazem referência a ele, através dos links entre suas páginas de conteúdo. A importância dos sites secundários, por sua vez, é calculada seguindo o mesmo critério, até que sejam encontrados sites secundários que não recebam referência de outros, recebendo nesse caso um valor mínimo de importância. O objetivo é permitir que a própria rede formada pelas referências entre os sites atribua um valor maior de importância para os sites com conteúdos mais relevantes.

Essa descrição simplificada não aborda todas as variáveis consideradas no algoritmo atual do Google, desenvolvido por BRIN e PAGE (1998). Hoje, de acordo com seus próprios fundadores, o Google utiliza, além do Pagerank, critérios como a localização geográfica do usuário, histórico de interesses e outros critérios confidenciais para ajudar na ordernação da lista de resultados de uma busca. Entretanto, a empresa sempre deixou bastante clara a maior importância do Pagerank para seu critério de ordenação de resultados (GOOGLE, 2011).

Pode-se aplicar o mesmo conceito do Pagerank na análise das redes sociais, baseando-se na ideia de que a importância de um vértice em um grafo direcionado deve ser definida pela soma das importâncias dos vértices que fazem referência a esse vértice original. Existem algumas versões de fórmulas utilizadas pera seu cálculo, sendo uma das versões mais simplificadas a apresentada por CHO e ROY (2004):

$$Pgrank(v_i) = d + (1-d)(Pgrank(v_1)/l_1 + \ldots + Pgrank(v_n)/l_n)$$

Equação 6 – fórmula para o cálculo da métrica Pagerank (CHO e ROY, 2004).

Onde $Pgrank(v_i)$ equivale ao cálculo de Pagerank para um vértice v_i que é referenciado pelos vértices $v_1,\ldots v_n$. Sendo lj o número de arestas partindo do vértice v_j. Sendo d a probabilidade de o próximo vértice ser selecionado aleatoriamente e, portanto, $(1 - d)$ sendo a probabilidade de o próximo vértice visitado ser um dos l_i referenciados por v_i.

2.3. Modelos de Redes

Quatro tipos fundamentais de redes são aceitos na comunidade acadêmica como o padrão de categorização: Redes Sociais, Redes Tecnológicas, Redes de Informação e Redes Biológicas (NEWMAN, 2010). Além de buscar uma classificação das redes conforme os tipos de objetos e relações que elas representam, vários autores também procuraram classificá-las de acordo com suas características estruturais ou sua topologia. STROGATZ (2011) resume com bastante simplicidade e exatidão os motivos que estimulam o estudo da anatomia das redes.

> "Por quê a caracterização da anatomia das redes é tão importante? Porque estrutura sempre afeta função. Por exemplo, a topologia das redes sociais afeta a disseminação de informações e doenças, e a topologia de redes elétricas afeta a robusteza e estabilidade da transmissão de energia." (STROGATZ, 2011, p. 268. Tradução do autor).

Os exemplos dados por Strogatz são referentes a estruturas de redes do "mundo real", pois representam relações estabelecidas de forma não aleatória, com base em determinadas regras (COHEN e HAVLIN, 2010).

Outro modelo de rede é obtido quando, a partir de um conjunto de n vértices, forma-se uma rede inserindo-se de modo iterativo e randomicamente arestas ligando pares desses vértices. Essas redes possuem características topológicas próprias, que não se assemelham aos modelos de rede do mundo real, recebendo a denominação de *Redes Aleatórias* (*Random Graphs*). As redes aleatórias são estudadas em uma área de pesquisa específica introduzida por ERDŐS e RÉNYI (1959).

A diferenciação entre os modelos de redes aleatórias e redes do mundo real pode ser identificada na Figura 6, onde é possível observar como a inserção aleatória de arestas, em um mesmo grupo de vértices, pode gerar topologia com características bastante distintas das redes que, de forma distinta, distribuam essas arestas com uma probabilidade maior para os vértices que possuam maior grau.

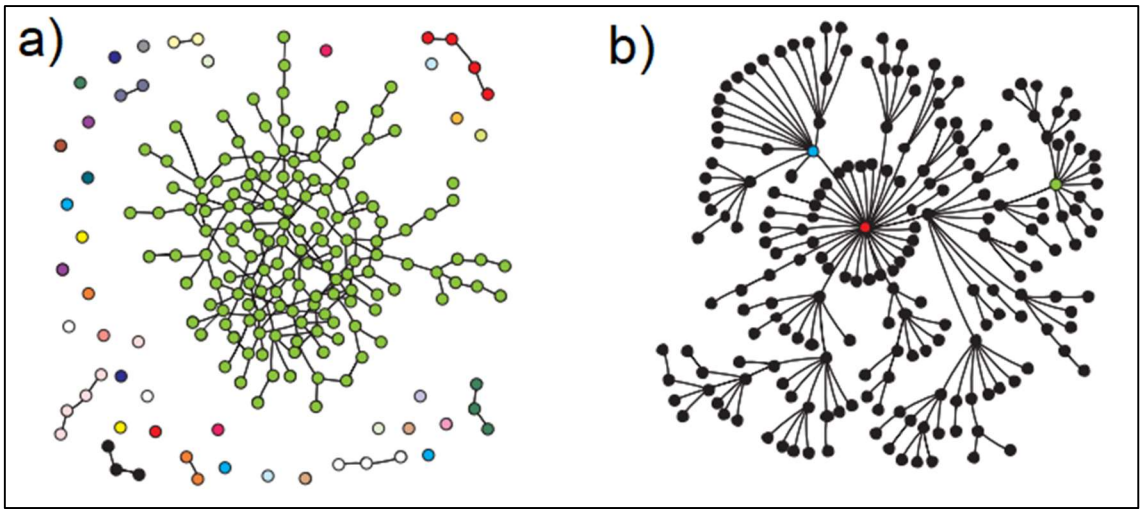

Figura 6 – Mesmo grupo de vértices conectados com algoritmo que (a) insere arestas randomicamente até que uma quantidade *n* seja atingida, formando um modelo de rede randômica, e (b) insere novos vértices sequencialmente conectando-os com os vértices já existentes na rede dando maior probabilidade de escolha para os de maior grau, formando um modelo de rede do mundo real (CALLAWAY et. al., 2000).

A topologia bastante diferenciada pode ser explicada pelos algoritmos utilizados na montagem das duas redes da Figura 6. Na rede (a), dado um conjunto de vértices sem arestas iniciais, um algoritmo insere arestas gradualmente conectando dois vértices escolhidos aleatoriamente, até que uma quantidade *n* de arestas seja atingida. Na rede (b) outro algoritmo parte de uma rede vazia e insere novos vértices sequencialmente, conectando-os com os vértices já existentes na rede estabelecendo uma probabilidade de escolha positivamente correlacionada com o grau dos vértices (CALLAWAY et. al., 2000). Esse último algoritmo simula situações de redes "reais", como a das redes sociais, em que existe uma maior chance de um novo usuário se conectar a um usuário preexistente com alto número de relações já estabelecidas.

A métrica de grau, apesar de ser calculada para cada vértice individual da rede analisada, é passível de ser utilizada como instrumento para analisar a rede como um todo. Uma das técnicas é utilizar os graus dos vértices para fazer uma análise topológica da rede, através do *histograma de distribuição de graus* (NEWMAN, 2003; COHEN e HAVLIN, 2010), que apresenta em seu eixo X o número de graus (*d*) e no eixo Y a probabilidade de que, dado um vértice qualquer $v \in V$, esse vértice tenha esse grau (*p(d)*) (Figura 7).

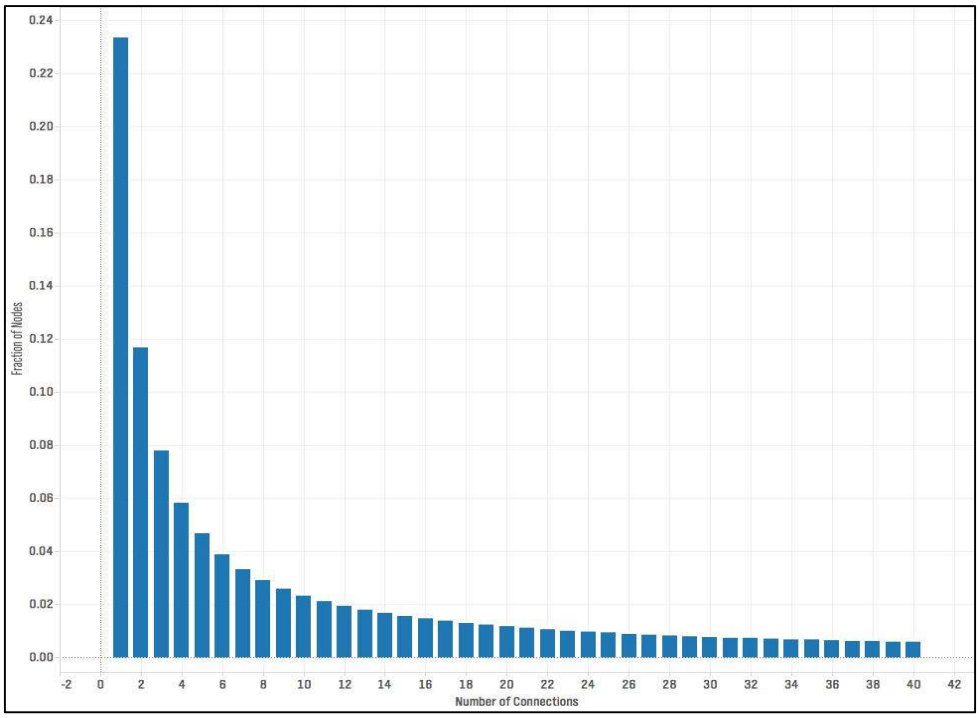

Figura 7 – Exemplo de histograma de distribuiçãode graus (LAMBERSON, 2015).

O exemplo mostrado na Figura 7 é bastante representativo da distribuição de graus nas redes sociais, em que uma grande maioria de vértices (usuários) tem um valor baixo de grau (número de conexões baixo), enquanto uma pequena quantidade de vértices apresenta um alto grau, indicando maior influência na rede. As redes que se enquadram nesse comportamento de distribuição de graus recebem, na literatura, a denominação de *redes de livre escala* (*scale-free*).

No caso de redes com arestas direcionadas, como as tratadas nessa dissertação, o histograma de distribuição de graus pode ter duas séries para mostrar as probabilidades relacionadas aos graus de entrada e de saída separadamente.

O padrão de formação mostrado na Figura 7(b) explica a distribuição de graus das redes de livre escala. Essas redes obedecem a uma lei de potência para definir a probabilidade de um determinado vértice ter uma quantidade k de arestas:

$$P(k) \sim k^{-y}$$

Equação 7 – padrão de formação das probabilidades na distribuição de grau de redes de livre escala.

Onde $k > 0$ e $y > 0$. Sendo y chamado expoente de livre escala e $P(k)$ a probabilidade de ocorrência de vértices de grau k na rede. A probabilidade, portanto, tende a decair exponencialmente conforme o aumento do grau.

Uma característica marcante das redes de livre escala, derivada desse padrão de distribuição de graus, é sua tolerância às falhas. Se um vértice aleatório é desconectado da rede, a maior probabilidade é de que esse vértice seja de grau baixo, trazendo pouco impacto para a interconectividade dos vértices restantes. Entretanto, as redes de livre escala ainda são suscetíveis a uma ação coordenada que objetivasse a retirada dos vértices de maior grau da rede, gerando um grande impacto na interconexão entre os vértices.

2.4. Revisão Bibliográfica

A Análise de Redes Sociais, conforme exposto no Capítulo 1, tem ganhado nos últimos tempos um forte estímulo para o desenvolvimento de estudos, principalmente pela disponibilidade crescente de recursos tecnológicos que permitem a montagem de representações de redes complexas e o cálculo de métricas de alto custo computacional.

Considerando esse grande e crescente universo de estudos ligados ao tema proposto nessa dissertação, essa revisão bibliográfica terá escopo definido por um conjunto de trabalhos diretamente associados à análise de redes sociais de microblogs e que tenham sido importantes para a definição dos objetivos e/ou formatação da metodologia utilizada nessa dissertação.

Sobre a potencialização de uma informação falsa através da rapidez com que ela pode ser disseminada, KEYES (2004) em seu livro *The Post-Truth Era: Dishonesty and Deception in Contemporary Life* comentou que um complicador importante que potencializa o impacto de uma informação falsa é a rapidez do processo de divulgação, que dificulta ou até inviabiliza sua correção através de esclarecimento posterior. Muitas vezes o número de pessoas atingidas é tão grande que o efeito formador de opinião já está solidificado no pensamento desses indivíduos, principalmente quando neles existem condições psicológicas prévias que os deixam mais suscetíveis a aceitar o conteúdo falso como verdadeiro. Alguns exemplos dessas condições poderiam ser observados na divulgação de notícias falsas que atribuem a algum grupo social a culpa por uma mazela qualquer, mostrando um preconceito latente contra esse grupo ou tema; ou pelo medo latente em algum grupo social de alguma situação futura que gere uma aceitação de conteúdo inverídico que venha a corroborar esse medo; ou ainda nos momentos de intenso debate político polarizado, em que a razão acaba sendo anestesiada pelo interesse em subjugar o indivíduo de pensamento contrário.

FANTINI (2016) comentou sobre outra característica que colabora para a disseminação de notícias falsas no ambiente virtual: os sistemas de direcionamento de conteúdo das redes sociais são programados de forma a criar “bolhas” de isolamento de grupos. Tal isolamento acontece quando grupos, que possuem determinadas características de pensamento identificadas pelos sistemas de direcionamento de conteúdo das redes sociais com base em comentários postados anteriormente, passam a receber com mais frequência mensagens concordantes a esses pensamentos, sem garantia de veracidade do conteúdo dessas mensagens.

Alguns trabalhos acadêmicos trataram aspectos do tema da disseminação de conteúdo falso nas redes sociais, como o de ZAGO (2012), que faz um estudo sobre os *trolls* – postagens emitidas em redes sociais de conteúdo falso ou fortemente emocional – que objetivam desestabilizar um grupo alvo de usuários. MARCHI (2012) aborda o impacto de falsas notícias, analisando como os adolescentes estão iniciando um processo de recusa da objetividade da mídia tradicional, considerando conteúdos não baseados em fatos para formar sua opinião.

O primeiro estudo que merece destaque foi proposto por KWAK (2010) e procurou avaliar o potencial do Twitter como canal de divulgação de notícias. Sua grande base de dados analisada, englobando 41,7 milhões de perfis de usuários dessa rede social, 1,47 bilhões de relações criadas a partir desses perfis, 4.262 tópicos pesquisados e 106 milhões de postagens, permitiu interessantes análises a partir de redes formadas com esses dados. Um dos enfoques propostos foi analisar o grafo direcionado formado pelos vértices representados pelos perfis de usuários e arestas formadas pela associação de seguidor/seguido entre esses usuários. A análise dessa rede trouxe características de sua topologia como a de um valor de diâmetro baixo, indicando uma rede com poucos níveis de profundidade, e de uma baixa reciprocidade entre os vértices, trazendo a informação de que a maioria das conexões é de seguidores associados a poucos perfis de usuários de grande popularidade. Nesse mesmo trabalho foi feita uma análise da correlação entre o número de seguidores de um usuário e sua capacidade de gerar referências, medida através da métrica *Pagerank*, que acabou confirmando a similaridade entre as duas medidas. Apesar dessa primeira conclusão, não foi observada a mesma correlação quando se comparou o número de seguidores com o número de repasses (grau de entrada do vértice). Ainda no mesmo estudo, foi feita uma análise que demonstrou que 85% dos *trending topics* do Twitter, que representam os termos mais citados nas postagens do período analisado, foram associados ou a notícias temporárias

de grande circulação ou a notícias antigas que mantiveram sua validade por muito tempo. Verificou-se também que um primeiro repasse de uma postagem acelera muito sua divulgação, com os próximos repasses sendo muito rápidos e atingindo uma média observada de mil usuários destinatários, independente do número de seguidores do usuário responsável pela postagem original.

Outro estudo que influenciou essa dissertação foi o de CERVI e MASSUCHIN (2010), que fez uma importante análise do Twitter como ferramenta de divulgação política, avaliando a utilização dessa rede social pelos dois candidatos principais ao governo do estado do Paraná em 2010, Beto Richa (PSDB) e Osmar Dias (PDT), que juntos somaram 98,07% dos votos válidos. Esse trabalho destacou a crescente importância do Twitter como ferramenta de divulgação política no Brasil e de seu potencial para influenciar eleições, conclusão que foi confirmada pelos importantes fatos políticos que aconteceram mais recentemente e foram comentados no Capítulo 1: o plebiscito para a saída do Reino Unido da União Européia (*Brexit*) (EL PAÍS, 2016), a eleição de Donald Trump nos EUA (FAUS, 2016; LORO, 2017; GLOBO, 2017) e o referendo para aprovação do plano de paz entre a Colômbia e o grupo guerrilheiro *Farc* (EXAME, 2016). A metodologia utilizada neste trabalho foi, principalmente, fazer análises quantitativas sobre as postagens dos dois candidatos, com o objetivo de identificar padrões na então nascente utilização política dessa rede social. Algumas das conclusões observadas foram a maior utilização das postagens para comunicação de notícias da campanha e de planos de governo dos candidatos; respostas de postagens mais frequentemente usadas como um aval dos eleitores aos candidatos (como elogios e estímulos) do que debates políticos; e intensificação do uso do Twitter pelos candidatos no período da campanha em comparação aos períodos anteriores, demonstrando que o uso da rede social fez parte de uma estratégia deliberada das campanhas dos dois candidatos.

O estudo de caso realizado por FERREIRA (2013) foi outro trabalho que trouxe conhecimentos importantes para essa dissertação. Seu objetivo foi o de analisar a disseminação da notícia sobre a realização de um evento específico no Twitter, o "Experiencie: Novas estratégias nas mídias sociais", realizado em outubro de 2011, para avaliar a utilização dessa rede social como ferramenta de divulgação. A contribuição específica desse estudo para a presente dissertação foi na avaliação feita sobre as ferramentas disponíveis de coleta e análise de dados do Twitter.

O trabalho de PRIMO (2011) foi importante para trazer uma reflexão sobre as mudanças do jornalismo provocadas pela crescente adoção das redes sociais como fonte de notícias. O autor faz uma revisão de vários outros estudos anteriores sobre o tema e encontra idéias comuns, como: que as redes sociais se estabeleceram definitivamente como instrumentos de divulgação de informações; que houve importante quebra do paradigma que no passado definia a notícia como obrigatoriamente originada por um jornalista, sendo suscetível a sua interpretação; que houve uma eliminação dos mecanismos de filtragem das notícias, que antes eram pré-selecionadas de acordo com os critérios dos detentores dos mecanismos de divulgação (grupos de comunicação); e que esse modelo atual de novo jornalismo iguala os potenciais de divulgação de grupos tradicionais de comunicação ao de simples pessoas físicas.

Outro artigo (ZHIWEI et al., 2014), também endereçou o problema da divulgação de notícias falsas nas redes sociais de microblogs. Esse texto teve um enfoque direcionado a uma proposta inovadora de algoritmo que permitiria avaliar o nível de credibilidade de um evento (fato, notícia) postado em rede social. O algoritmo realizava o processamento de um modelo hierárquico em que a raiz seria formada pelo próprio evento a ser analisado e seus subníveis seriam formados por subeventos montados a partir de outras postagens relacionadas ao primeiro evento original. A pontuação de credibilidade do evento original seria calculada a partir do processamento textual da veracidade de todos os subníveis do grafo, sendo o algoritmo proposto uma variante de solução para o problema de otimização de grafos. O algoritmo proposto nesse artigo foi testado em dois casos reais que mostraram uma melhoria mensurável na verificação de credibilidade em comparação com outros métodos propostos até então.

Uma outra ferramenta com proposta de ajudar na identificação de notícias falsas no Twitter, analisada durante a concepção desse estudo, foi a TRAILS, apresentada em artigo acadêmico (FINN et. al., 2014). A estratégia dessa ferramenta é fazer uma investigação sobre a origem da divulgação da notícia no Twitter e apresentar uma análise da evolução da audiência, através da associação de termos textuais das postagens com a notícia investigada. Essas informações apresentadas por essa solução tecnológica podem ser utilizadas para avaliar a veracidade das notícias de interesse.

SITUNGKIR (2011) também foi utilizado como referência no desenvolvimento dessa dissertação, com um artigo em que avaliava a disseminação de uma notícia falsa através de um enfoque epidemiológico. Nesse trabalho, foi feita uma análise detalhada da evolução da audiência formada por usuários do Twitter de uma postagem com

notícia reconhecidamente falsa, destacando como essa audiência se comportava antes e depois do momento em que a notícia era oficialmente esclarecida.

Um estudo que fez uma análise detalhada sobre o modelo da disseminação de notícias falsas nas redes sociais e as possíveis estratégias para redução dos danos foi o de GUANHUA, GUANLING e EIDENBENZ (2011). Nesse trabalho, os autores investigaram um caso específico de divulgação de notícia falsa e confirmaram sua aderência ao modelo de redes de livre escala. Usando o mesmo caso de exemplo, também analisaram diferentes métodos de diminuição de velocidade e probabilidade de disseminação. As duas estratégias propostas foram a "orientada a usuário", em que cada ponto potencial de disseminação seria "inoculado" através de uma postagem de esclarecimento sobre a notícia; e a "orientada a servidor", em que os próprios servidores da rede social em questão teriam a missão de identificar a disseminação da notícia falsa e se encarregariam de bloqueá-la. Considerando que os testes foram feitos sobre um caso único, os resultados apresentados neste estudo tiveram que ser analisados com ressalvas, mas o artigo trouxe importantes enfoques que foram considerados nessa dissertação.

A seguir, um quadro com o resumo dos estudos analisados.

Autor(es)	Título	Ano	Problema/Abordagem
KEYES	The Post-Truth Era: Dishonesty and Deception in Contemporary Life	2004	Analisou os impactos da disseminação de informações falsas na formação do pensamento da sociedade.
FANTINI	Pós-verdade ou o triunfo da religião?	2016	Analisou os sistemas de direcionamento de conteúdo de redes sociais, formadores de "bolhas" de grupos isolados.
ZAGO	Trolls e Jornalismo no Twitter	2012	Estudou postagens com conteúdo falso ou fortemente emocional, emitidas em redes sociais, que tinham como objetivo um grupo alvo.
MARCHI	With Facebook, Blogs, and Fake News, Teens Reject Journalistic "Objectivity"	2012	Estudou o impacto das falsas notícias nos adolescentes gerado pela desconfiança nos meios tradicionais de imprensa.
KWAK	What is Twitter, a social network or a news media?	2010	Avaliou o Twitter como ferramenta de divulgação de notícias, através do uso de grafos representando as associações entre os usuários desta rede social.

CERVI e MASSUCHIN	O uso do twitter nas eleições de 2010: O microblog nas campanhas dos principais candidatos ao governo do Paraná	2010	Avaliou o Twitter como ferramenta de divulgação política na eleição para o governo do Paraná em 2010.
FERREIRA	A influência das mídias sociais na divulgação de um evento: um estudo de caso sobre o evento Experiencie	2013	Estudou um caso específico de divulgação de evento via Twitter, para analisar a capacidade de disseminação desta rede social.
PRIMO	Transformações no jornalismo em rede: sobre pessoas comuns, jornalistas e organizações; blogs, Twitter, Facebook e Flipboard	2011	Analisou as recentes mudanças do jornalismo provocadas pelo uso mais frequente das redes sociais como canais de divulgação de notícias.
ZHIWEI et al	News Credibility Evaluation on Microblog with a Hierarchical Propagation Model	2014	Propôs um algoritmo para quantificar o nível de credibilidade de uma informação postada em uma rede social.

Tabela 1 – Quadro resumo com os estudos relacionados aos temas abordados nesta dissertação.

3. Metodologia

Nesse capítulo será apresentada a metodologia proposta, primeiramente abordando a estratégia para a seleção das notícias falsas que formarão o grupo de notícias a ser analisado e, em seguida, detalhando as ferramentas utilizadas e a preparação dos dados para a análise.

3.1. Seleção de notícias falsas

Para a seleção das notícias falsas que compõem esse trabalho, optou-se por usar como principais fontes de notícias dois sites reconhecidos pelo trabalho de esclarecimento de conteúdos falsos amplamente divulgados na internet, o E-Farsas (www.e-farsas.com) e o Boatos.org (www.boatos.org). A partir da citação e esclarecimento de notícias falsas nesses sites, foram feitas buscas de postagens no Twitter que trouxessem em seu conteúdo textual referência a essas notícias. Quando a busca logrou êxito em localizar postagens, a notícia falsa correspondente foi adicionada ao grupo de notícias a analisar. As coletas foram realizadas no período compreendido entre os meses de março e junho de 2017.

3.2. Coleta dos dados

Para a coleta de dados relacionados a postagens emitidas na rede social de microblogs Twitter utilizou-se a ferramenta *NodeXL* (SOCIALMEDIA, 2014). Esse software é um módulo de expansão da planilha Microsoft Excel (MICROSOFT, 2016) que tem como funcionalidade principal a análise de tabelas representativas de grafos (SMITH et. al., 2009). Essas tabelas podem ser informadas manualmente ou, conforme recursos fornecidos pela própria ferramenta, coletadas diretamente de algumas redes

sociais, como Twitter (www.twitter.com), Facebook (www.facebook.com), Flickr (www.flickr.com) ou Youtube (www.youtube.com).

O recurso de coleta de postagens do NodeXL acessa internamente a API (*Application Program Interface* – Interface de Programação de Aplicativos) de busca do Twitter para localizar as postagens associadas a determinados filtros, como palavras-chave, intervalo de data de postagem ou origem geográfica (TWITTER INC., 2017). Essa API é disponibilizada pela empresa Twitter sob determinadas condições de uso que impõem limitações que precisam ser destacadas para a correta validação desse estudo. A primeira dessas limitações é que, devido ao imenso número de postagens emitidas diariamente no Twitter, o universo pesquisável é limitado a um período de até sete dias anteriores à data da realização da coleta. Não é possível, portanto, realizar uma busca por postagens emitidas em um determinado momento histórico acontecido meses atrás. Essa janela cronológica será fortemente explorada nessa pesquisa através do uso de termos de busca relacionados a conteúdo inverídico recentemente divulgado.

Ao mesmo tempo em que limitações são impostas pela API do Twitter, algumas vantagens importantes surgem de sua utilização e corroboram a escolha desse mecanismo para a obtenção dos dados utilizados neste estudo, como a possibilidade de integração com outras ferramentas tecnológicas (como a NodeXL) e a garantia de acesso ao universo de todas as postagens, já que é uma ferramenta disponibilizada pela própria empresa responsável pela rede de microblogs Twitter. Além dessas, outra principal vantagem advinda da utilização da API do Twitter, identificada a partir de vários testes e checagens dos dados coletados no período desse estudo, é que a busca de postagens vai além da simples análise de seu conteúdo textual, procurando também no conteúdo das páginas externas referenciadas no texto das postagens. Como exemplo, uma busca pela expressão “Tapioca pode matar”, que representa uma das notícias falsas pesquisadas nessa dissertação, traz não somente as postagens que em seu conteúdo possuem essas palavras, mas também as postagens que façam referências a sites externos que tenham em seu conteúdo as palavras usadas na busca. Essa importante característica permite o adequado aumento da abrangência do universo de pesquisa, já que a maioria dos usuários do Twitter emitem postagens apenas com links para conteúdo externo, seja pela praticidade ou pela imposição que o Twitter estabelece de 140 caracteres no texto. Não considerar essas postagens que em seu conteúdo possuam apenas links externos seria uma grave ausência no universo de busca, comprometendo seriamente os resultados.

Para a formação de cada conjunto de postagens será feita uma busca no NodeXL com uma expressão textual que represente alguma das notícias comprovadamente falsas que tenham circulado em período recente. A partir da escolha de algumas dessas notícias, é formada uma expressão textual que deve espelhar ao máximo seu título. Para facilitar a localização das postagens relacionadas, procura-se inserir na expressão apenas os termos principais da notícia falsa, evitando-se o uso de artigos e/ou palavras de ligação. Esses termos foram testados para garantir que as postagens retornadas realmente estivessem associadas à notícia falsa estudada.

Para a correta validação desse estudo, é importante destacar que a coleta realizada buscando postagens associadas à notícia falsa pode trazer conteúdo que, ao contrário de afirmar a notícia falsa estudada, traz um esclarecimento sobre essa notícia, portanto negando sua veracidade. Uma hipótese que chegou a ser avaliada para fazer parte da metodologia foi de separar a rede total em duas redes distintas. A primeira seria a rede formada por todas as postagens que em seu conteúdo não possuam expressões que indiquem uma tentativa de negação ou esclarecimento acerca da notícia falsa: “falso”, “falsa”, “falsos”, “falsas”, “mentira”, “inverídico”, “enganação” e “não é verdade”, recebendo o nome de *rede afirmativa*. A segunda seria a *rede negativa*, resultado da subtração da rede total e da rede afirmativa, ou seja, contendo apenas as postagens que possuem em seu conteúdo as expressões que indicam negação ou esclarecimento acerca da notícia falsa pesquisada. Entretanto, testes desse algoritmo de separação da rede total aplicado às postagens mostraram que, muitas vezes, o uso de palavras específicas para caracterizar a postagem como parte da rede afirmativa ou negativa é insuficiente, pois nossa língua é bastante diversificada e permite inúmeras formas de criar frases que negam uma afirmação. Uma situação que exemplifica esse problema é a frase: “Tapioca pode matar”, que indiscutivelmente traz uma mensagem afirmativa sobre a suposta mortalidade do alimento tapioca; já outras frases muito semelhantes: “Tapioca pode matar?” ou “Tapioca pode matar, será?” ou ainda “Tapioca pode matar, rsrsrs” são de difícil enquadramento nas redes afirmativa ou negativa, pois não é possível determinar a intenção do emissor. A separação da rede total usando esses critérios acabaria também por exigir a criação de uma terceira rede *indeterminada*, além das redes afirmativa e negativa, que teria a função de agrupar um grande subconjunto das postagens, prejudicando a análise das redes restantes. Além disso, considera-se que mesmo uma postagem que procura esclarecer uma notícia falsa participa indiretamente de sua disseminação, mesmo que com objetivos de esclarecimento. Pelos motivos

expostos, nessa dissertação optou-se por realizar as análises apenas sobre a rede total, que abrange todas as postagens coletadas associadas a uma notícia falsa, sem filtros posteriores.

Os resultados da coleta realizada pela ferramenta NodeXL são distribuídos em planilhas que representam: 1) as postagens que atendem ao filtro definido, com seu conteúdo e informações associadas e 2) a lista de usuários emissores dessas postagens, com seus atributos associados (HANSEN et. al., 2010; KIM et. al., 2014).

Com os resultados da coleta, são propostas as seguintes análises sobre os conjuntos de dados, de acordo com a questão endereçada. Considerando-se um conjunto de notícias falsas de interesse coletadas em um dado período:

- No recorte temporal observado, quais as origens da divulgação das notícias falsas desse conjunto no Twitter?
    - Identificação do emissor da primeira postagem da coleta, de acordo com sua data e hora de emissão.
    - Identificação do emissor que recebeu mais referências, a partir das métricas de grau de entrada e centralidades de intermediação, proximidade, autovetor e *Pagerank* dos vértices representantes dos perfis de usuário.
- Como é o processo de disseminação de notícias falsas no Twitter para o conjunto em questão?
    - Calcular as métricas básicas do grafo representativo da coleta.
        - Diâmetro e diâmetro médio.
        - Componente máxima.
        - Densidade.
        - Distribuição de grau.
        - Coeficiente de clusterização.
    - Avaliar, através da data de emissão das postagens e do número de seguidores dos perfis de usuário emissores correspondentes (chamado neste estudo de *audiência*), a velocidade de disseminação da notícia falsa entre os usuários do Twitter.

Sobre a avaliação da velocidade de transmissão de uma dada notícia falsa com base no número de seguidores dos usuários emissores das postagens que disseminam essa notícia, é importante destacar o risco de duplicidade de contagem desses

seguidores. No caso de uma nova postagem de um mesmo usuário, é relativamente simples evitar somar novamente seus seguidores, bastando configurar a fórmula de soma para evitar usuários que já foram emissores na rede analisada. Entretanto, em uma situação hipotética em que um mesmo seguidor X está associado a dois usuários A e B distintos, a contagem desse seguidor seria duplicada. A eliminação desse tipo de duplicidade exigiria acesso à lista completa de usuários seguidores de cada usuário emissor, informação à qual não foi possível ter acesso neste trabalho. Mesmo com essa ressalva, esse estudo considera que a soma de usuários seguidores ainda é bastante relevante, pois dá um indicativo da velocidade de crescimento da audiência de uma dada notícia falsa em função do tempo. Em contrapartida, é preciso lembrar que existem variáveis externas de difícil mensuração que poderiam equilibrar esse risco de duplicidade de seguidores, como o fato de alguns desses seguidores contabilizados como apenas um usuário representarem grupos de indivíduos, como organizações ou famílias.

3.3. Preparação dos dados

Os resultados das coletas, realizadas utilizando-se a ferramenta NodeXL, foram armazenados em pastas de trabalho do Microsoft Excel (arquivos de extensão XLSX) e, dentro desses arquivos, os resultados foram distribuídos em planilhas. As duas planilhas principais e suas colunas de dados mais importantes, conforme detalhado em (SMITH et. al., 2009), são apresentadas a seguir.

3.3.1. Planilha “Arestas”

A primeira planilha do arquivo de coleta – *Edges* – traz informações sobre as postagens encontradas na busca, cada uma representada em uma linha da tabela. As postagens podem ser de autoria própria (nos casos de postagem avulsa ou resposta a outra postagem) ou de reencaminhamento, quando o usuário apenas repassa o conteúdo da postagem original. As colunas de dados utilizadas neste estudo são:

- Vértice 1: perfil do usuário do Twitter que emitiu a postagem. Nesse estudo, tratado como emissor. No grafo, será representado por um vértice de onde poderá partir uma aresta direcionada. Isso não acontecerá caso a postagem tenha sido avulsa de autoria própria.

- *Vértice 2*: perfil do usuário do Twitter referenciado na postagem. Nesse estudo, tratado como receptor. Nos casos em que a postagem é de resposta ou repasse, esse usuário será simbolizado no grafo como o vértice destino da aresta.
- Postagem: conteúdo textual completo da postagem.
- *Data da Postagem*: data e hora da emissão da postagem.
- Tipo: tipo da postagem (*Tweet* - avulso, *Mentions* - repasse ou *Replies to* - resposta).
- URLs na postagem: endereços da internet referenciados no texto da postagem. Informação usada para a identificação das URLs mais citadas e os canais de comunicação (domínios da internet) mais referenciados pelas postagens.

3.3.2. Planilha "Vértices"

A planilha "Vértices" traz informações sobre todos os usuários da coleta que foram listados na planilha de postagens - Arestas. A tabela apresentada nessa planilha tem as seguintes colunas de dados principais:

- Vértice: nome do perfil do usuário do Twitter.
- Postagens: quantidade de postagens.
- Seguidores: quantidade de seguidores (usuários que se registram para receber automaticamente novas postagens emitidas do usuário em questão).
- Métricas (*grau de entrada, grau de saída, centralidade de intermediação, centralidade de proximidade, centralidade de autovetor, Pagerank)* - informações calculadas com base nos indicadores mais comuns de análise de redes sociais, explicados na revisão teórica no Capítulo 2).

3.3.3. Postagens órfãs e duplicadas

Uma importante decisão tomada neste trabalho foi quanto à manutenção de todas as ocorrências de postagens encontradas. Isso implica em deixar como parte do conjunto de dados as postagens:

- Órfãs – quando uma postagem é emitida por um usuário, é gerada uma aresta com saída e retorno ao vértice que o representa - laço. Como as postagens órfãs são válidas, mesmo nos casos em que não sejam referenciadas posteriormente por nenhuma outra postagem, e possuem audiência mínima formada pelos seguidores dos seus usuários emissores, optou-se por mantê-las nos conjuntos de dados.
- Duplicadas – em que uma mesma referência entre dois usuários através de uma postagem é repetida em momentos distintos. Nesse caso considerou-se que excluí-las poderia eliminar do universo de análise ocorrências válidas que podem ter sido estimuladas por novas interações legítimas entre os usuários, como uma sequência em que um usuário A repassa uma postagem de um usuário B e, logo em seguida, o mesmo usuário A resolve também responder à postagem original de B.

As quantidades correspondentes a essas situações de postagens órfãs e de duplicidade de referências serão destacadas nas análises para permitir uma avaliação quanto à sua influência sobre a estrutura da rede analisada.

3.4. Desenho dos grafos e métricas analisadas

Para dar uma primeira visão da rede formada a partir da coleta das postagens, é apresentado um desenho dessa rede baseado no algoritmo de FRUCHTERMAN e REINGOLD (1991), que procura simular as forças de atração e repulsão observadas na natureza no cálculo da distribuição espacial dos vértices. Nesse algoritmo, a atração é calculada com base nos vértices adjacentes e a repulsão é determinada pelo conjunto total de vértices do grafo. Para facilitar a identificação dos vértices presentes na rede que possuam mais seguidores, o tamanho do vértice foi configurado para ter uma correlação positiva com a quantidade de seguidores do usuário representado pelo vértice, seguindo a fórmula:

$$Tam = 100 * \frac{\text{seguidores}(v)}{\sum \text{seguidores}(v \in V)}$$

Equação 8 – fórmula para o cálculo do tamanho do vértice.

Onde *Tam* é o tamanho de um vértice *v* representativo de um usuário, variando de 0 a 100, de acordo com a razão entre o número de seguidores do usuário representado por v e a soma de seguidores de todos os vértices da rede.

A seguir, relacionam-se as métricas de grafos apresentadas no Capítulo 2, com o sentido que elas possuem no contexto das redes geradas a partir das coletas das notícias falsas estudadas neste trabalho. Algumas das métricas foram calculadas com uma versão não direcionada do grafo e em outras foi considerado o direcionamento das arestas. Em geral, o uso da versão não direcionada do grafo foi feito nas métricas em que se optou por desconsiderar em que sentido a conexão se estabeleceu entre os vértices. Em cada métrica será informada a escolha feita na metodologia desse estudo.

- Informações básicas do grafo: quantidade de vértices, arestas, postagens órfãs. Dão uma primeira ideia do tamanho da rede analisada e de como são as relações entre seus vértices. O número de arestas foi totalizado considerando o direcionamento das arestas, para dar uma melhor ideia do número de conexões entre os vértices.

- Componentes e componente máxima: a informação do total de vértices na maior componente do grafo ajuda a dar uma melhor visão da rede. Foram totalizadas usando a versão não direcionada do grafo.

- Diâmetro: define o tamanho da rede coletada sob o enfoque da distância entre os usuários referenciados nas postagens. Foi calculado usando a versão não direcionada do grafo.

- Diâmetro médio: essa métrica é importante para evitar avaliações incorretas que o valor do diâmetro pode trazer, por uma distância atípica entre os pares de vértices mais distantes do grafo. Seu valor será destacado quando mostrar grande diferença em relação ao diâmetro. Foi calculado usando a versão não direcionada do grafo.

- Densidade: métrica utilizada para avaliar o nível de interconectividade do grafo. Um valor alto dessa métrica – considerado assim um valor decimal próximo ou igual a um – indica uma rede de padrão “teia”, em que todas (ou quase todas) as

arestas possíveis estariam presentes. Foi calculada considerando o direcionamento das arestas.

- Grau de entrada: no universo da coleta, esta métrica representa a quantidade de postagens (arestas) emitidas em resposta ao usuário (vértice) ou repasse de uma de suas postagens por outro usuário. Por definição, foi calculada considerando o direcionamento das arestas.

- Grau de saída: no universo da coleta, esta métrica representa a quantidade de postagens (arestas) emitidas pelo usuário (vértice). Essas postagens podem representar respostas ou repasses de postagens de outros usuários que, neste caso, serão o vértice destino da aresta direcionada. Por definição, foi calculada considerando o direcionamento das arestas.

- Histograma de distribuição de graus: para uma melhor visualização das características de interconexão dos vértices da rede, é exibido um histograma de distribuição de graus, com séries distintas para a métrica de grau de entrada e de saída.

- Centralidade de intermediação: no universo da coleta, identifica usuários (vértices) relevantes que atuam como intermediários entre grupos diferentes de usuários. Foi calculada considerando o direcionamento das arestas.

- Centralidade de proximidade: um menor valor significa maior velocidade na transmissão de informações originadas pelo usuário (vértice) e, portanto, maior influência dele no universo da coleta. Foi calculada usando a versão não direcionada do grafo.

- Centralidade de autovetor: usuários (vértices) com valor mais elevado para essa métrica são caracterizados como de maior influência. Foi calculada considerando o direcionamento das arestas.

- PageRank: essa métrica mostrou uma maior eficiência em classificar os vértices das redes analisadas neste trabalho, evitando a grande quantidade de valores nulos calculados pela métrica de centralidade de autovetor. Por esse motivo,

uma tabela com a classificação dos cinco primeiros usuários por ordem descendente de PageRank é apresentada, com o objetivo de mostrar os usuários mais influentes da rede segundo essa métrica. Foi calculado considerando o direcionamento das arestas.

- Primeiro emissor e vértice com maior número de referências: com bases nas datas de postagens e no grau de entrada dos vértices serão identificados o primeiro emissor da notícia falsa no período da coleta e o usuário que recebeu o maior número de referências (respostas ou repasses de postagens de sua autoria).

- Audiência por data: gráfico mostrando a evolução do número de usuários que teriam acesso à notícia falsa pelo Twitter – que recebe neste estudo a denominação de *audiência* – por data. A audiência será baseada na soma do número de seguidores de cada usuário emissor das postagens presentes na rede.

No próximo capítulo são detalhadas as análises de cada uma das redes formadas a partir da coleta das dez notícias falsas selecionadas. Ao final será feita uma análise conjunta de todas as redes, destacando as características nas quais tenha sido possível observar um padrão.

4. Análise dos Dados

A seguir, detalham-se as análises sobre cada coleta, termo que está sendo utilizado neste trabalho para representar os dados gerados a partir da extração de postagens da rede de microblogs Twitter em uma data específica e com um termo de busca que representa uma das notícias falsas pesquisadas.

4.1. Notícia: Vacina da febre amarela tem veneno mortal

Essa notícia falsa circulou com mais intensidade no final da segunda semana de março e afirmava que o governo, após vacinar mais de 40 milhões de pessoas contra a febre amarela, teria descoberto a presença de uma substância tóxica com alto potencial de mortalidade na vacina. As principais métricas da rede total, formada por todas as postagens obtidas na coleta, são apresentadas na Tabela 2.

Métrica	Valor	Métrica	Valor
Vértices	206	Componentes	130
Arestas	226	Componente máxima	39
Arestas duplicadas	24	Diâmetro	2
Coef. clusterização global	0,0294435	Diâmetro médio	1,696162
Postagens órfãs	136	Densidade	0,001989

Tabela 2 – Métricas básicas da rede total da coleta “vacina febre amarela veneno mortal”.

A coleta de postagens dessa notícia foi feita no dia 26/03/2017 com a expressão de busca “vacina febre amarela veneno mortal” e retornou um total de 226 postagens e 206 usuários, sendo a primeira postagem emitida no dia 20/03/2017 às 21h35min e a última no dia 26/03/2017 às 01h49min. O desenho da rede total pode ser visto na Figura 8.

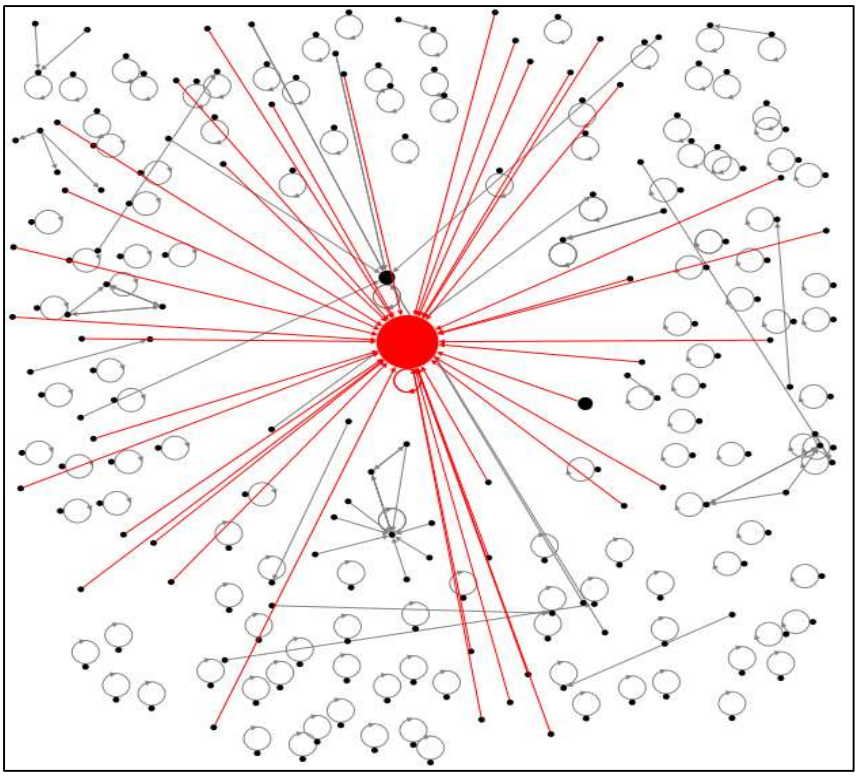

Figura 8 – Desenho da rede total da coleta “vacina febre amarela veneno mortal” (fonte: autor).

Analisando-se o desenho da Figura 8, é possível destacar a presença de 136 postagens órfãs formando laços. Como explicado na metodologia, a ocorrência de postagens avulsas sem referências de repasse ou resposta de outros usuários é comum em algumas coletas. O número de arestas duplicadas encontrado foi de 24, aproximadamente 10% do total de arestas, caracterizando que a maioria das interações entre os usuários foi única no período analisado. O desenho também mostra uma rede bastante centralizada em um vértice específico, destacado com suas arestas adjacentes no centro da figura. Esse vértice representa o usuário “g1”, que nessa rede possui o maior número de seguidores. Esse formato centralizado é reforçado pelo valor baixo de diâmetro – dois – indicando um máximo de duas arestas de distância entre os vértices conectados do grafo. Entretanto, o valor de 39 para a componente máxima indica que os vértices conectados estão distribuídos em componentes distintas, informação comprovada pelo elevado número de componentes do grafo (130).

A baixa densidade observada no grafo da Figura 8 (0,001989) foi dentro do esperado, considerando a expectativa que as interações do Twitter sejam de muitas respostas e repasses a postagens de usuários centrais de maior audiência. Condizente com a baixa densidade, o coeficiente de agrupamento global também trouxe um valor

baixo (0,0294435) em relação ao seu valor máximo teórico (um). Essa expectativa de rede relativamente esparsa foi comprovada no histograma de distribuição de graus exibido na Figura 9. Nele é possível observar a maior probabilidade de localização de vértices com baixos graus de entrada, com curva de tendência semelhante à encontrada em redes de livre escala, exceto o aumento verificado entre a probabilidade do grau um para o grau zero. Considerando apenas a métrica de grau de saída, vê-se uma quase totalidade de concentração no grau um.

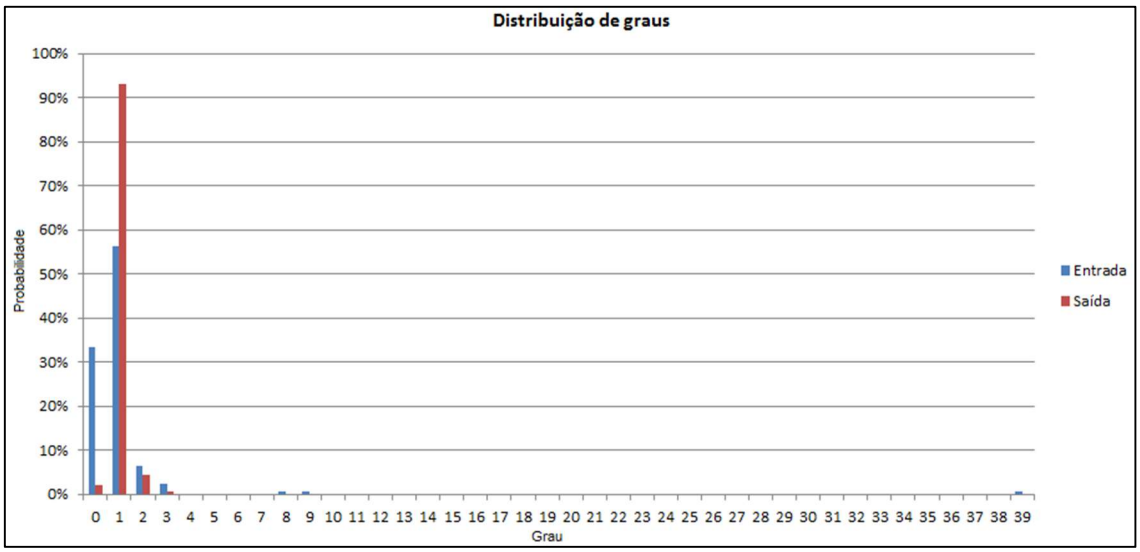

Figura 9 – Distribuição de graus de entrada e saída da coleta “vacina febre amarela veneno mortal” (fonte: autor).

O primeiro emissor de postagem relacionada a notícia falsa “vacina febre amarela veneno mortal” foi o usuário “boatosorg”, que representa o site “www.boatos.org” – um dos sites que fazem o trabalho de identificação e esclarecimento de notícias falsas utilizadas como fonte de material para esse estudo. Essa postagem não teve referências posteriores. O emissor com maior grau de entrada (39) foi o usuário “g1”, que representa o portal de notícias G1, do grupo Globo de comunicação.

A Figura 10 exibe o gráfico que mostra a evolução da audiência em função do tempo. É possível observar que até o dia 24/03/2017 o padrão era de postagens emitidas por usuários com um número relativamente baixo de seguidores. A soma da audiência era de 802.195 até as 09h20min dessa data, quando uma postagem efetuada pelo usuário “g1”, com grande quantidade de seguidores (9.021.320), elevou a soma da audiência para um patamar de 9.823.515. Este valor permaneceu relativamente estável até o

momento da última postagem, atingindo ao final um total de 10.507.182 leitores. Vê-se, portanto, que a quase totalidade da audiência foi proporcionada por apenas um usuário.

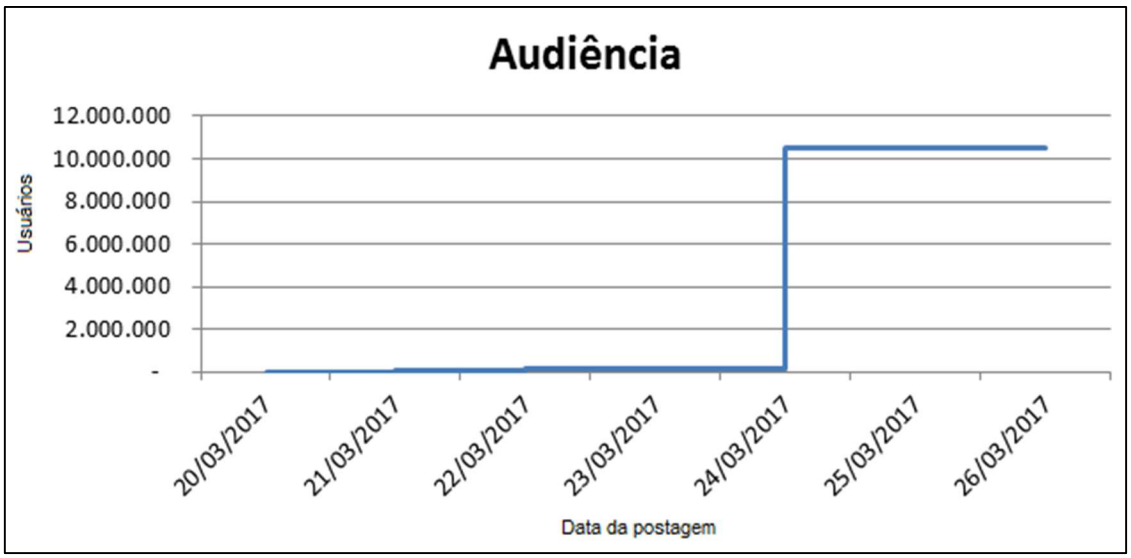

Figura 10 – Gráfico de evolução da audiência da coleta “vacina febre amarela veneno mortal” (fonte: autor).

A Tabela 3 apresenta a classificação de usuários da rede por ordem descendente da métrica PageRank dos vértices associados.

Usuário	Grau de Entrada	Grau de Saída	Central. Intermed.	Central. Proximid.	Central. Autov.	PageRank	Coef. Clust. Local	Seguidores
g1	39	1	1406,000	0,026	0,150	18,215	0,000	9021320
Efarsas	9	2	54,000	0,125	0,000	3,999	0,036	15459
bemestar	8	1	42,000	0,143	0,000	3,986	0,000	649322
moreno_n19	0	3	6,000	0,333	0,000	1,919	0,000	36
_soaresc	3	1	2,000	0,500	0,000	1,723	0,000	1389

Tabela 3 – Usuários mais influentes por ordem de PageRank da coleta “vacina febre amarela veneno mortal”.

4.2. Notícia: Morte do cantor Arlindo Cruz

O famoso cantor Arlindo Cruz, internado em 17/03/2017 por ter sofrido um acidente vascular cerebral, foi vítima de notícia falsa disseminada no final de março que divulgava seu falecimento. As principais métricas da rede total, formada por todas as postagens obtidas na coleta, são apresentadas na Tabela 4.

Métrica	Valor	Métrica	Valor
Vértices	53	Componentes	16
Arestas	54	Componente máxima	26
Arestas duplicadas	2	Diâmetro	2
Coef. clusterização global	0	Diâmetro médio	1,766537
Postagens órfãs	17	Densidade	0,01342525

Tabela 4 – Métricas básicas da rede total da coleta “Arlindo Cruz morre”.

A coleta de postagens dessa notícia foi feita no dia 23/03/2017 com a expressão de busca “Arlindo Cruz morre” e retornou um total de 54 postagens e 53 usuários, sendo a primeira postagem emitida no dia 17/03/2017 às 21h29min e a última no dia 23/03/2017 às 01h30min.

O período da coleta parece seguir o mesmo padrão observado na coleta sobre a notícia da vacina da febre amarela, com seu início demarcado pelo limite imposto pela API de busca do Twitter: até sete dias anteriores à data em que a consulta é realizada. Isso reforça a ideia de que essas postagens não formam um conjunto sobre tudo o que foi falado sobre o tema, mas sim um corte desse período específico.

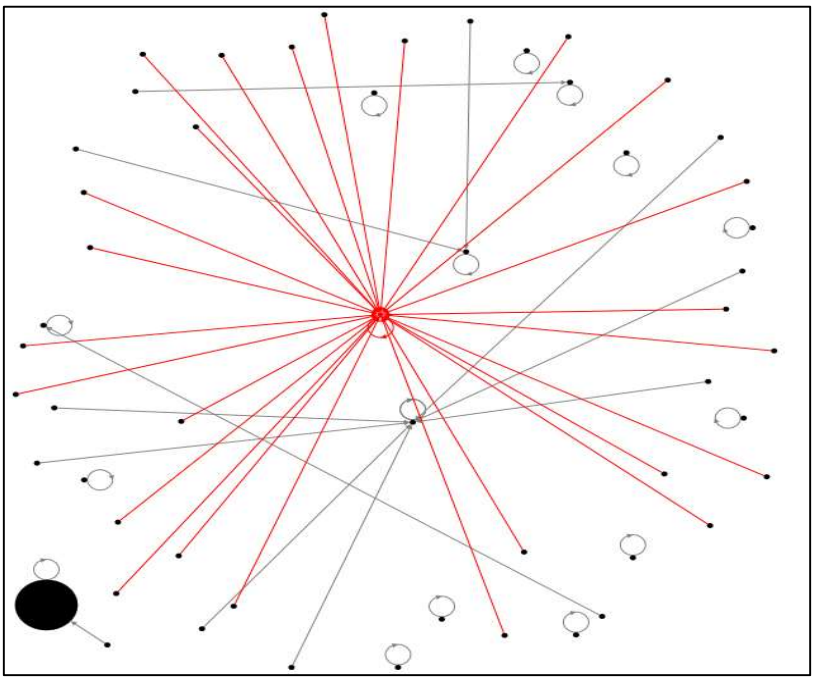

Figura 11 – Desenho da rede total da coleta “Arlindo Cruz morre” (fonte: autor).

A Figura 11 apresenta o desenho da rede total relativa a notícia “Arlindo Cruz morre”. Mais uma vez se destacam visualmente as postagens órfãs, totalizando 17. O número de arestas duplicadas encontrado foi de apenas duas, mantendo o padrão de interação única entre usuários observado na coleta anterior, relativa à notícia falsa sobre

a vacina da febre amarela. Seguiu-se também o mesmo formato centralizado em um vértice específico, destacado com suas arestas adjacentes ao centro da Figura 11. Entretanto, diferentemente do observado na coleta anterior, o vértice que representa o usuário com maior número de seguidores não foi o mesmo que se posicionou como vértice central, mostrando que a quantidade de ligações – critério principal considerado para a centralização do vértice no algoritmo de FRUCHTERMAN e REINGOLD (1991) – não é obrigatoriamente correlata com a quantidade de seguidores do Twitter.

Essa visão centralizada da rede foi reforçada pelo valor baixo de diâmetro – dois – indicando um máximo de duas arestas de distância entre os vértices conectados do grafo. A componente máxima, com 26 vértices, indica que a quase metade dos vértices se encontra conectada na maior componente.

Continuou-se observando baixa densidade (0,01342525), condizente com o mostrado no desenho do grafo, em que a maioria das arestas segue o padrão de ligar vértices de baixo grau de entrada com vértices de alto grau de entrada. A métrica de coeficiente de agrupamento local em zero demonstra ausência de triangulações entre vértices. Comprovando esse padrão, no histograma de distribuição de graus da Figura 12 é possível observar uma correlação negativa entre o grau de entrada e sua probabilidade de localização em uma situação de busca aleatória na rede, indicando uma curva de tendência de distribuição de probabilidades muito semelhante à encontrada em redes de livre escala. Considerando apenas a métrica de grau de saída, vê-se que todos os vértices dessa rede possuem o valor um para essa métrica.

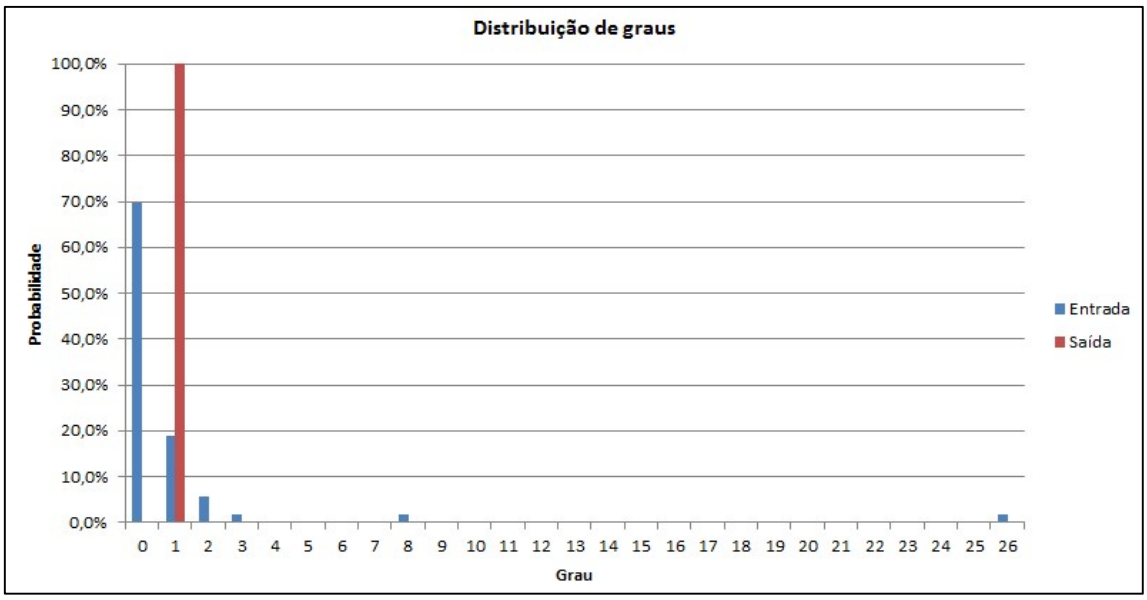

Figura 12 – Distribuição de graus de entrada e saída da coleta “Arlindo Cruz morre” (fonte: autor).

O primeiro emissor de postagem relacionada a notícia falsa “Arlindo Cruz morre” foi o usuário “bandsportstv”, que representa a emissora de TV Band Sports, do grupo Band de comunicação. Essa postagem teve referências posteriores, apesar de seus vértices e arestas associados não terem formado a maior componente do grafo. O emissor com maior grau de entrada – 26 – foi o usuário “melissadeusa”, com descrição no Twitter indicando se tratar de pessoa física.

A evolução da audiência em função do tempo é mostrada na Figura 13. Novamente se observa o padrão de salto súbito na audiência ocorrido no dia 17/03/2017. A soma da audiência era de 1.030 na primeira postagem, feita nessa mesma data às 21h29min, e aumentou consideravelmente quando uma postagem efetuada pelo usuário “g1”, com grande quantidade de seguidores (9.021.320), elevou a soma da audiência para um patamar de 9.066.899 que permaneceu estável até o momento da última postagem, atingindo ao final um total de 9.161.284 leitores. Manteve-se, portanto, a característica de elevação abrupta de audiência com posterior manutenção de seu patamar.

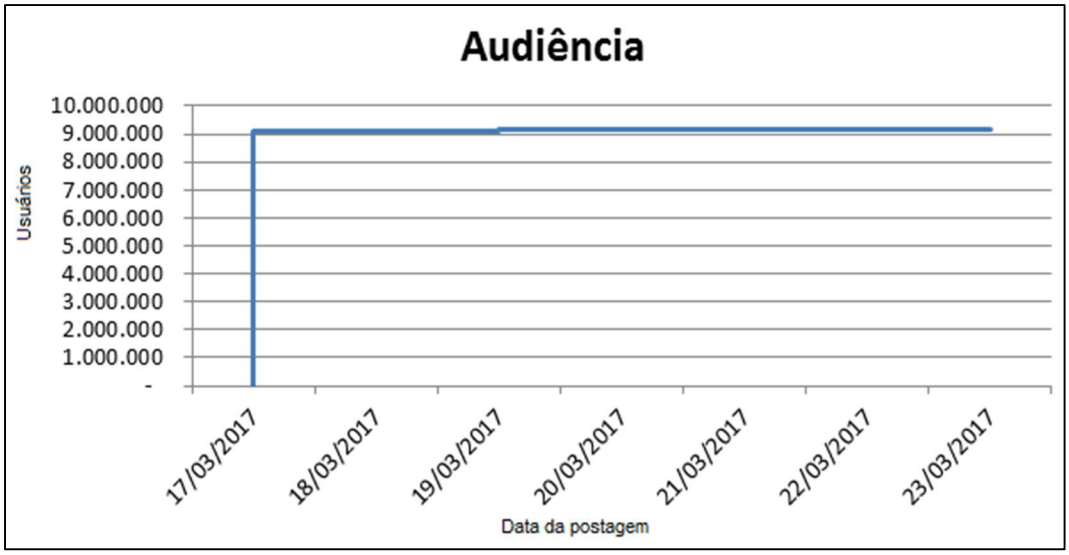

Figura 13 – Gráfico de evolução da audiência da coleta “Arlindo Cruz morre” (fonte: autor).

A Tabela 5 mostra a classificação de usuários da rede por ordem descendente da métrica PageRank dos vértices associados. Dessa vez, observa-se como o número de seguidores diretos não deve ser utilizado como única métrica para definir a capacidade de disseminação de um usuário. Além de ter o maior índice PageRank, o usuário “Melissadeusa” teve o maior número de referências (grau de entrada) e marcante

presença na ligação entre os vértices (centralidade de intermediação), mesmo com baixa quantidade de seguidores (1.637) em comparação com outros usuários dessa rede.

Usuário	Grau de Entrada	Grau de Saída	Central. Intermed.	Central. Proximid.	Central. Autov.	PageRank	Coef. Clust. Local	Seguidores
Melissadeusa	26	1	600,000	0,040	0,181	12,243	0,000	1637
bandsportstv	8	1	42,000	0,143	0,000	3,986	0,000	1030
ieojs	3	1	2,000	0,500	0,000	1,723	0,000	1048
g1	2	1	0,000	1,000	0,000	1,298	0,000	9065869
ryanbricklayer	2	1	0,000	1,000	0,000	1,298	0,000	15

Tabela 5 – Usuários mais influentes por ordem de PageRank da coleta “Arlindo Cruz morre”.

4.3. Notícia: Taxista que xingou Jesus morre em acidente

A notícia falsa “Taxista que xingou Jesus morre em acidente” circulou na segunda quinzena de março e foi pesquisada no dia 31/03/2017. Trazia a informação de que um taxista teria morrido em um acidente após cometer o sacrilégio religioso de xingar Jesus de "terrorista vagabundo e mentiroso". O taxista e seu comentário polêmico foram verdadeiros, mas o acidente automobilístico fatal noticiado foi com outro taxista e sua esposa, que não possuíam nenhuma ligação com o taxista que teria xingado Jesus. As principais métricas da rede total, formada por todas as postagens obtidas na coleta, são apresentadas na Tabela 6.

Métrica	Valor	Métrica	Valor
Vértices	40	Componentes	35
Arestas	58	Componente máxima	5
Arestas duplicadas	25	Diâmetro	2
Coef. clusterização global	0	Diâmetro médio	0,548387
Postagens órfãs	52	Densidade	0,0038461

Tabela 6 – Métricas básicas da rede total da coleta “Taxista que xingou Jesus morre em acidente”.

A expressão de busca utilizada para a pesquisa dessa notícia foi “Taxista xingou Jesus morre” e retornou um total de 58 postagens e 40 usuários, sendo a primeira postagem emitida no dia 26/03/2017 às 15h52min e a última no dia 31/03/2017 às 15h34min.

O período da coleta abrangendo seis dias parece seguir o mesmo padrão observado nas coletas anteriores, com início e fim demarcados pelo limite cronológico imposto pela API de busca do Twitter. Tal fato leva a crer que essas postagens não foram as únicas sobre o tema, podendo existir outras postagens, além do prazo de sete dias, que tenham divulgado essa notícia.

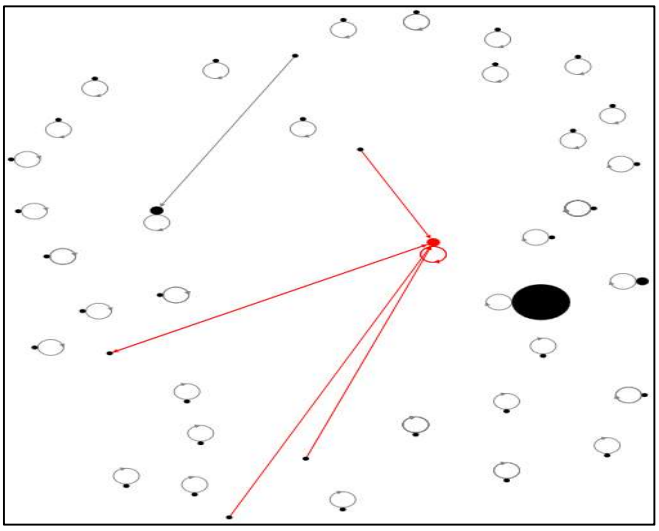

Figura 14 – Desenho da rede total da coleta “Taxista que xingou Jesus morre em acidente” (fonte: autor).

O desenho da rede total relacionada à notícia “Taxista que xingou Jesus morre em acidente” é apresentado na Figura 14. Pela terceira vez, verifica-se o padrão de muitas postagens órfãs. Entretanto, diferentemente das duas postagens anteriores, a quantidade de postagens órfãs, 42, é quase igual à quantidade de arestas, 58. Foi encontrado também um número considerável de arestas duplicadas (25). Pode ser verificado, pelo próprio desenho e pelas métricas calculadas até aqui, que a rede tem poucas conexões entre os vértices. As outras métricas poderão ratificar esse indicativo.

O vértice com mais alto grau de entrada, e que por esse motivo foi posicionado mais ao centro do desenho, foi destacado com suas arestas adjacentes na Figura 14. Observou-se o segundo caso em que a quantidade de ligações dos vértices não foi positivamente correlacionada com a quantidade de seguidores do Twitter. Em outras palavras, essa é mais uma rede que mostra que a maior quantidade de referências a postagens não leva automaticamente a uma maior audiência.

O valor baixo de diâmetro – dois – indicando um máximo de duas arestas de distância entre os vértices conectados do grafo é exatamente o mesmo encontrado nas duas coletas anteriores, o que chama a atenção quando se considera que essa rede tem

um tamanho bem menor que as duas analisadas anteriormente. A análise de outras coletas poderá confirmar se esse valor de diâmetro é uma constante para todas as redes coletadas do Twitter.

A componente máxima conectou cinco vértices de um total de quarenta, mostrando que a maioria dos vértices não estava presente na componente com mais vértices.

A baixa densidade (0,0038461) foi outra característica que se manteve, em comparação às duas coletas anteriores. A métrica de coeficiente de agrupamento local em zero demonstra ausência completa de triangulações entre vértices. O histograma de distribuição de graus da Figura 15 ratifica o padrão em que a maioria das arestas que ligam vértices distintos o faça ligando vértices de baixo grau de entrada com vértices de alto grau de entrada. Vê-se novamente a correlação negativa entre o grau de entrada e sua probabilidade de localização em busca aleatória na rede, indicando uma curva de tendência de distribuição de probabilidades muito semelhante à encontrada em redes de livre escala. Analisando apenas a série relativa à métrica de grau de saída, observa-se novamente uma quase completa concentração no grau um.

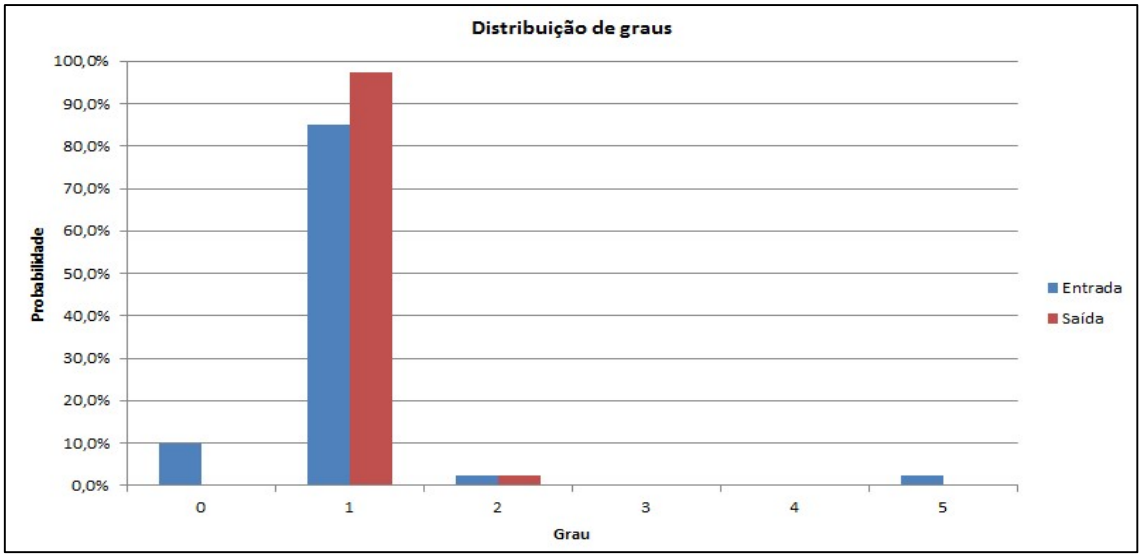

Figura 15 – Distribuição de graus de entrada e saída da coleta “Taxista que xingou Jesus morre em acidente” (fonte: autor).

O primeiro emissor de postagem relacionada a notícia falsa “Taxista que xingou Jesus morre em acidente” foi o usuário “efarsas”, que representa um dos sites que fazem o trabalho de esclarecimento de notícias falsas, usado como uma das fontes de notícias para esse estudo. Essa primeira postagem foi a que acabou formando, a partir de si, a componente máxima que, conforme comentado anteriormente, agrupou apenas cinco

vértices. O emissor com maior grau de entrada foi também o “efarsas”, com valor de cinco para essa métrica.

A evolução da audiência em função do tempo é mostrada na Figura 16. Mais uma vez se observa o padrão de salto súbito na audiência, com um acréscimo destacado no dia 27/03/2017. A soma da audiência era de 26.104 até às 22h34min dessa data, quando uma postagem efetuada pelo usuário “sou_portoalegre”, com grande quantidade de seguidores (291.066), elevou a soma da audiência para um patamar de 317.170, onde permaneceu estável até o dia seguinte, quando duas postagens de usuários com seguidores na faixa de 15 mil elevaram mais uma vez a audiência de forma marcante. Ao final desse universo de postagens, a audiência atingida chegou a 352.405 usuários. Pode-se afirmar, entretanto, que mesmo considerando essas elevações de menor amplitude do dia 28/03/2017, manteve-se a característica de elevação abrupta de audiência com posterior relativa estabilidade de seu patamar.

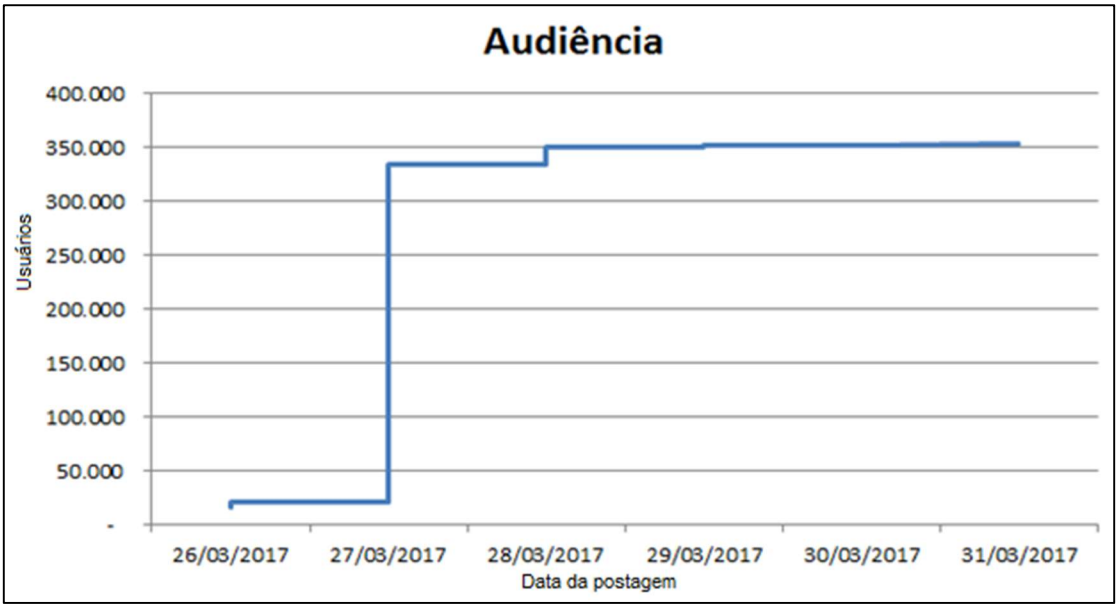

Figura 16 – Gráfico de evolução da audiência da coleta “Taxista que xingou Jesus morre em acidente” (fonte: autor).

A Tabela 7 apresenta a classificação de usuários da rede por ordem descendente da métrica PageRank dos vértices associados. O usuário com maior valor para essa métrica (“efarsas”) foi também o usuário com maior grau de entrada e centralidade de intermediação, além de estar muito próximo da maior quantidade de seguidores vista nessa rede.

Usuário	Grau de Entrada	Grau de Saída	Central. Intermed.	Central. Proximid.	Central. Autov.	PageRank	Coef. Clust. Local	Seguidores
efarsas	5	2	12,000	0,250	0,390	2,619	0,000	15842
agora_rn	2	1	0,000	1,000	0,000	1,298	0,000	15939
ajduksportal	1	1	0,000	0,000	0,000	1,000	0,000	747
amaduxx	1	1	0,000	0,000	0,000	1,000	0,000	453
ateaorgbr	1	1	0,000	0,000	0,000	1,000	0,000	14309

Tabela 7 – Usuários mais influentes por ordem de PageRank da coleta “Taxista que xingou Jesus morre em acidente”.

4.4. Notícia: Tapioca pode matar

A circulação da notícia “Tapioca pode matar” aconteceu na última semana de março. Seu conteúdo trazia a informação de que a tapioca – prato típico do nordeste do Brasil – seria potencialmente mortal em uma situação de consumo normal, devido à existência de ácido cianídrico na fécula da mandioca – ingrediente principal da tapioca. O esclarecimento veio em seguida por autoridades sanitárias, informando que, apesar de ser verdade que existe um tipo de mandioca que possui essa substância (chamada “mandioca brava”), o processamento da fécula acaba por remover o ácido cianídrico totalmente. As principais métricas da rede total relacionada à notícia “Tapioca pode matar”, formada por todas as postagens obtidas na coleta, são apresentadas na Tabela 8.

Métrica	Valor	Métrica	Valor
Vértices	15	Componentes	7
Arestas	19	Componente máxima	6
Arestas duplicadas	2	Diâmetro	3
Coef. clusterização global	0	Diâmetro médio	1,298246
Postagens órfãs	8	Densidade	0,05238

Tabela 8 – Métricas básicas da rede total da coleta “Tapioca pode matar”.

A expressão de busca utilizada para a pesquisa, que foi feita no dia 02/04/2017, foi “Tapioca pode matar” e retornou uma quantidade pequena de 19 postagens e 15 usuários, sendo a primeira postagem emitida no dia 28/03/2017 às 00h18min e a última no dia 31/03/2017 às 22h42min.

Apesar da baixa quantidade de postagens associadas a essa notícia, optou-se por mantê-la no conjunto de coletas analisadas nessa dissertação, por entender-se que é útil comparar as características identificadas em redes de tamanhos distintos.

Diferentemente das três coletas anteriores analisadas, o período de quatro dias entre a primeira e última postagem sugere que se obteve um universo mais completo das postagens sobre esse assunto, ainda que não se possa descartar completamente a possibilidade de que outras postagens em datas anteriores ao dia da coleta subtraído de sete dias (26/03/2017) tenham sido ignoradas pelo limite de sete dias imposto pelo Twitter.

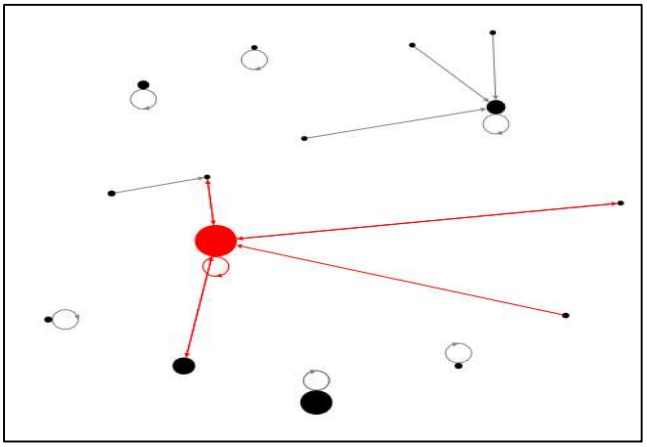

Figura 17 – Desenho da rede total da coleta “Tapioca pode matar” (fonte: autor).

A Figura 17 apresenta o desenho da rede total relacionada à notícia “Tapioca pode matar”. Manteve-se a já observada destacada quantidade de postagens órfãs, equivalente a oito postagens. Essa rede apresenta apenas duas arestas duplicadas, caracterizando até então uma rede pequena, esparsa e de poucas conexões.

O vértice com mais alto grau de entrada, posicionado em destaque próximo ao centro do desenho, foi destacado com suas arestas adjacentes na Figura 17. Para essa rede foi possível verificar visualmente que o vértice de maior número de referências foi o mesmo que apresentou a maior quantidade de seguidores do Twitter. Entretanto, diferentemente do observado em coletas anteriores, observou-se nessa rede a presença de outros vértices com quantidade considerável de seguidores, mas que possuem um número de ligações próximo de zero.

Apesar de nessa pequena rede ser mantida a característica de valor baixo de diâmetro, seu valor três é mais alto que o encontrado nas três coletas anteriores. A análise de outras coletas mais adiante neste trabalho poderá confirmar se esse valor de diâmetro voltará a se aproximar de dois, conforme visto nas redes anteriores.

A componente máxima possui seis vértices de um total de quinze vértices do grafo. A densidade calculada de 0,05238 informa que a quantidade de arestas chegou a

mais de 5% das arestas teoricamente possíveis, um valor que não pode ser considerado alto em comparação com o limite teórico, mas que é marcadamente maior que o encontrado nas três coletas anteriores. A métrica de coeficiente de agrupamento local em zero demonstra inexistência de triangulações entre vértices.

O histograma de distribuição de graus da notícia “Tapioca pode matar” é exibido na Figura 18. Observa-se uma maior probabilidade de localização de vértices de grau de entrada mais baixo, entre zero e um. Com relação ao grau de saída, é alta a probabilidade no grau um.

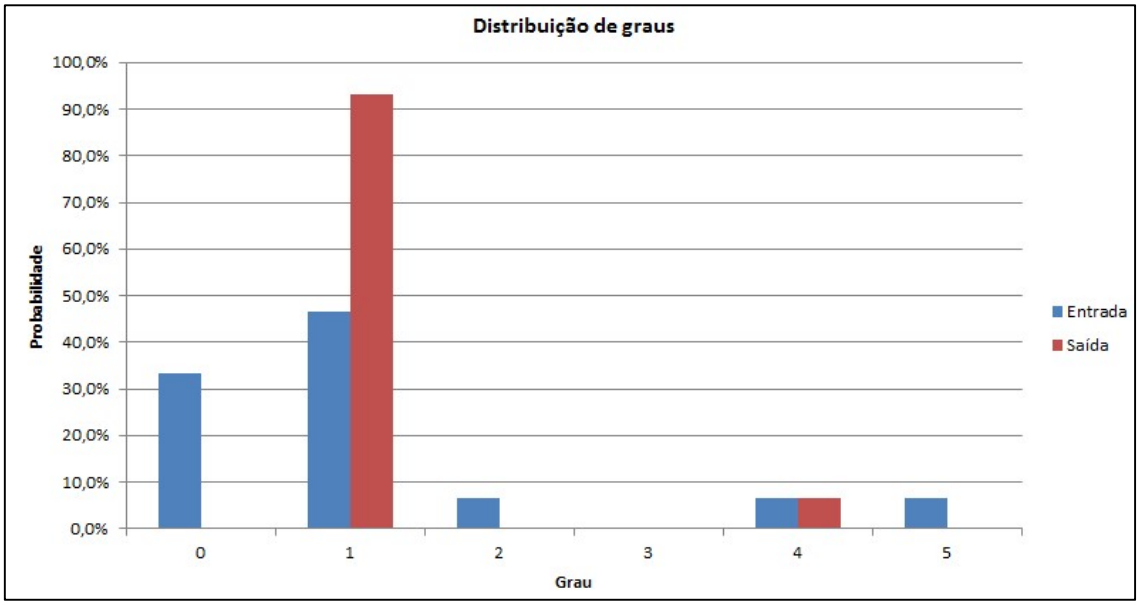

Figura 18 – Distribuição de graus de entrada e saída da coleta “Tapioca pode matar” (fonte: autor).

O primeiro emissor de postagem relacionada a notícia falsa “Tapioca pode matar” foi o usuário “danixmoraes_”, que aparentemente se trata de pessoa física. Essa primeira postagem acabou se tornando órfã dentro da rede pesquisada, sem referências posteriores a ela. O emissor com maior grau de entrada foi o “efarsas” com valor de cinco para essa métrica.

A evolução da audiência em função do tempo é mostrada na Figura 19. Nela se pode ver mais uma vez o padrão de salto súbito na audiência, com o aumento mais pronunciado no dia 30/03/2017, seguido de um aumento de menor magnitude no dia seguinte, 31/03/2017. Esse primeiro aumento súbito ocorrido no dia 30/03/2017 deveu-se principalmente a uma postagem às 15h39min que elevou a soma da audiência de 3.639 para 19.481, que teve o usuário “efarsas” – possuidor de grande número de seguidores (15.842) – como emissor. Na mesma data, às 15h49min, outra postagem,

dessa vez do usuário “everiting” – com 9.223 seguidores – elevou mais uma vez a soma da audiência para um total de 29.981, que se seguiu de novas elevações de menor magnitude no dia seguinte, 31/03/2017, atingindo um total final de audiência de 37.195 leitores. Pode-se afirmar, considerando apenas as elevações de maior amplitude vistas no dia 30/03/2017, que se manteve a característica de elevação abrupta de audiência com posterior estabilidade de seu patamar.

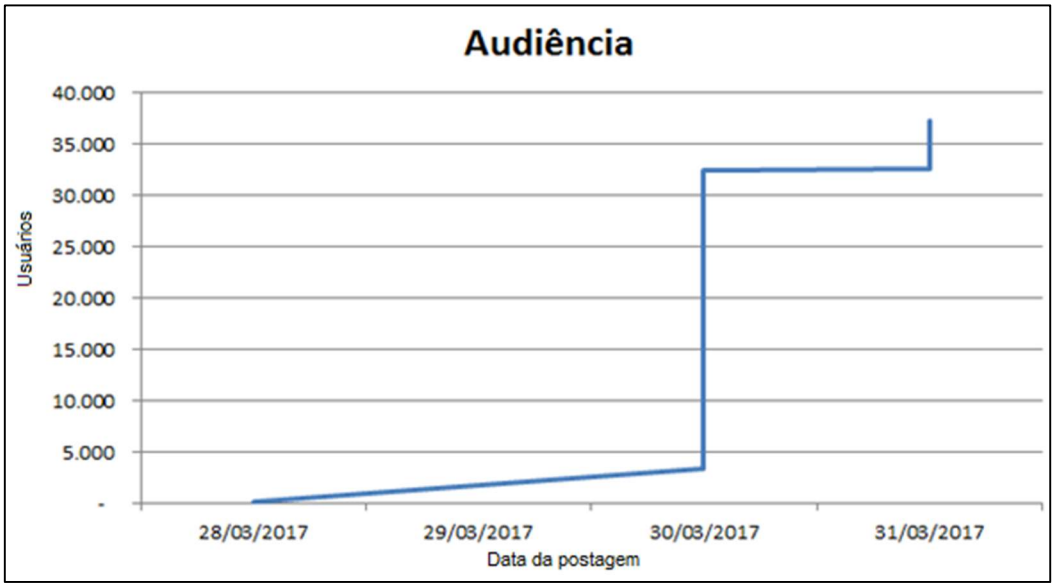

Figura 19 – Gráfico de evolução da audiência da coleta “Tapioca pode matar” (fonte: autor).

A Tabela 9 mostra a classificação de usuários da rede por ordem descendente da métrica PageRank dos vértices associados. Assim como na coleta sobre a notícia da morte do cantor Arlindo Cruz, o usuário “efarsas” foi o que mostrou maior valor para essa métrica, juntamente com maior grau de entrada, centralidade de intermediação e maior quantidade de seguidores entre os principais usuários dessa rede.

Usuário	Grau de Entrada	Grau de Saída	Central. Intermed.	Central. Proximid.	Central. Autov.	PageRank	Coef. Clust. Local	Seguidores
efarsas	5	4	18,000	0,167	0,360	2,530	0,000	15842
boatosorg	4	1	6,000	0,333	0,000	2,168	0,000	3226
marite_coelho	2	1	8,000	0,125	0,162	1,108	0,000	117
danixmoraes_	1	1	0,000	0,000	0,000	1,000	0,000	120
everiting	1	1	0,000	0,000	0,000	1,000	0,000	9223

Tabela 9 – Usuários mais influentes por ordem de PageRank da coleta “Tapioca pode matar”.

4.5. Notícia: Pink Floyd anuncia turnê no Brasil

Uma notícia divulgada no dia 1º de abril em um site especializado em informações sobre o gênero musical rock (Baú do Rock), comunicava a vinda da mundialmente famosa banda Pink Floyd ao Brasil para cumprir uma turnê de shows em várias cidades. Apesar da data sugestiva da divulgação – dia da mentira – ainda assim uma parte expressiva de usuários da rede repassou essa informação, como pode ser constatado pelas principais métricas da rede total relacionada à essa notícia falsa, formada por todas as postagens obtidas na coleta, apresentadas na Tabela 10.

Métrica	Valor	Métrica	Valor
Vértices	321	Componentes	274
Arestas	334	Componente máxima	22
Arestas duplicadas	25	Diâmetro	2
Coef. clusterização global	0	Diâmetro médio	1,260302
Postagens órfãs	285	Densidade	0,00046729

Tabela 10 – Métricas básicas da rede total da coleta “Pink Floyd anuncia turnê Brasil”.

A expressão de busca utilizada para a pesquisa, que foi feita no dia 08/04/2017, foi “Pink Floyd anuncia turnê Brasil” e retornou 334 postagens de 321 usuários, sendo a primeira postagem emitida no dia 01/04/2017 às 12h14min e a última no dia 08/04/2017 às 04h18min.

Aqui é possível notar uma repetição do padrão de janela cronológica definida pelo intervalo de datas compreendido entre a data da divulgação da notícia – 1º de abril – e a própria data de realização da coleta – 08/04/2017. Isso demonstra ser bastante provável que outras postagens ainda fossem encontradas em datas posteriores.

A Figura 20 apresenta o desenho da rede total relacionada a notícia “Pink Floyd anuncia turnê Brasil”.

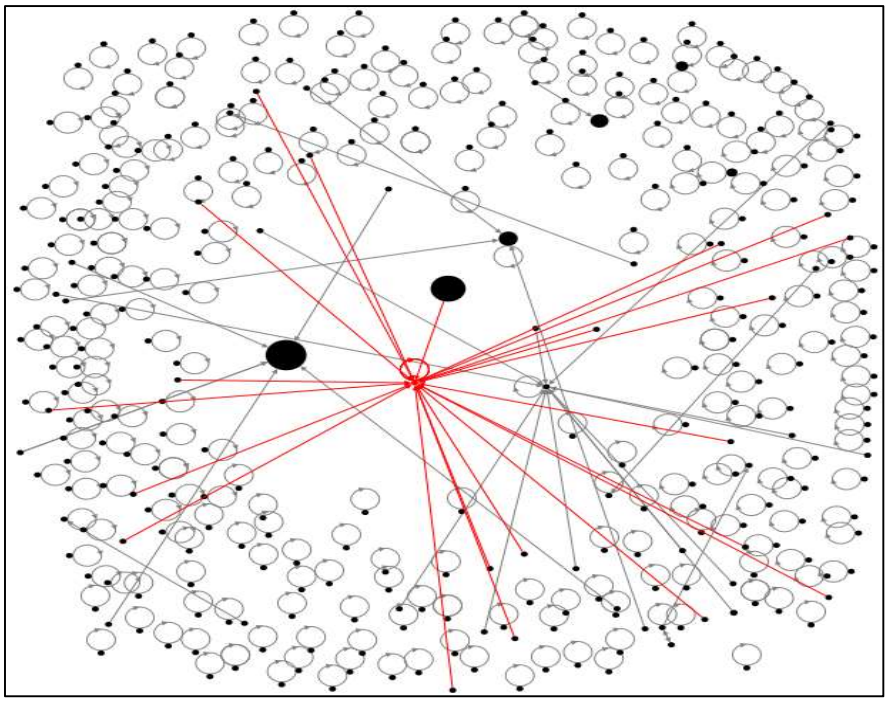

Figura 20 – Desenho da rede total da coleta “Pink Floyd anuncia turnê Brasil” (fonte: autor).

O primeiro destaque no desenho é a grande quantidade de postagens órfãs, como já visto nas quatro coletas anteriores. Apesar da mesma justificativa para o surgimento dessas postagens avulsas – postagens sem referência de outros usuários – ainda assim a quantidade observada chama atenção, somando mais de 88% das postagens totais nessa coleta. As arestas duplicadas, demonstrando situações de repetição de referências entre dois usuários, foram relativamente raras – 25 ao todo. Agregando essas informações à métrica da componente máxima – que agrupa uma quantidade pequena de vértices em comparação ao total (22) – e à densidade calculada de 0,00046729, tem-se formada uma visão de uma rede muito esparsa e de poucas conexões entre os vértices, mesmo comparada às redes das quatro notícias falsas analisadas anteriormente, que já tinham apresentado características desse tipo. A métrica de coeficiente de agrupamento local em zero demonstra ausência de triangulações entre vértices.

O vértice com mais alto grau de entrada, posicionado em destaque próximo ao centro do desenho, foi destacado com suas arestas adjacentes na Figura 20. Essa rede entra para o grupo formado pelas coletas das notícias falsas “Taxista que xingou Jesus morre em acidente”, “Tapioca pode matar” e “Arlindo cruz morre”, que não mostram correlação positiva entre o número de seguidores e a influência medida pelo grau de

entrada. O recebedor do maior número de referência pelos outros usuários apresenta uma quantidade relativamente baixa de seguidores: 4.075.

O valor de diâmetro voltou ao padrão de dois, observado em todos os casos anteriores, com exceção apenas da coleta “Tapioca pode matar”. Isso informa que as ligações entre os usuários se dão de forma direta ou com, no máximo, um usuário intermediário.

O histograma de distribuição de graus da Figura 21 mostra um padrão semelhante ao de todas as coletas anteriores, com relação às altas probabilidades de localização de vértices de graus baixos e, analogamente, baixas probabilidades de localização de vértices de grau mais alto. Sob a ótica do grau de entrada, a grande maioria dos vértices (87,3%) se concentrou no grau um e sob a ótica do grau de saída, a probabilidade associada ao grau um foi ainda mais expressiva (99,1%).

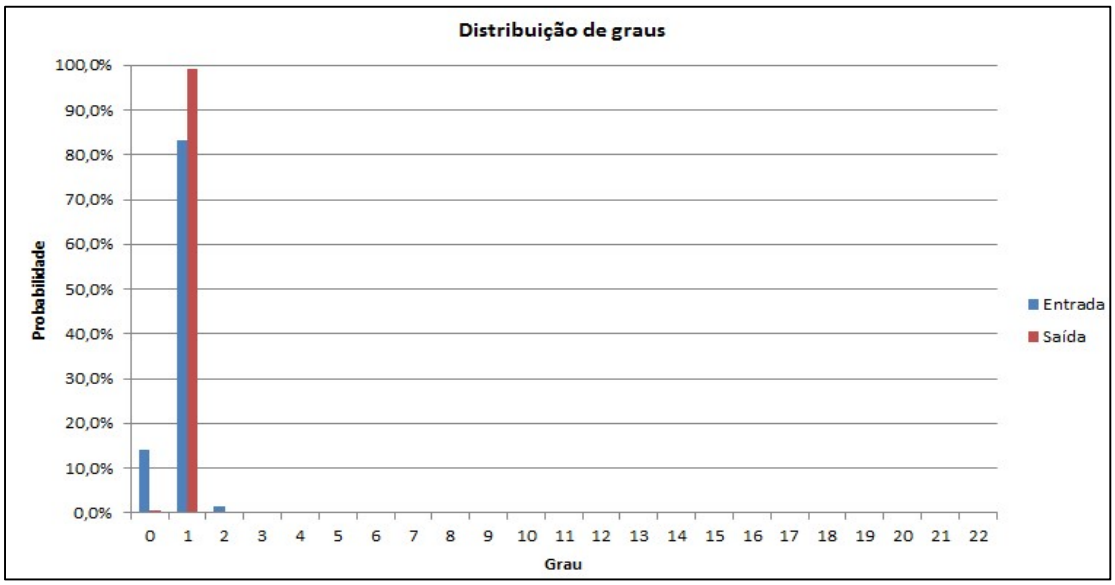

Figura 21 – Distribuição de graus de entrada e saída da coleta “Pink Floyd anuncia turnê Brasil” (fonte: autor).

O primeiro emissor de postagem relacionada a notícia falsa “Pink Floyd anuncia turnê Brasil” foi o usuário “daciofernandes”, que aparentemente se trata de pessoa física. Essa primeira postagem acabou se tornando órfã dentro da rede pesquisada, sem referências posteriores a ela. O emissor com maior grau de entrada foi o “ivanmagoo48” com valor de 22 para essa métrica, uma quantidade de referências alta em comparação à observada em outras coletas.

A evolução da audiência em função do tempo é mostrada na Figura 22. O padrão de salto súbito na audiência mais uma vez se repetiu, com o aumento mais pronunciado

no dia 02/04/2017, quando a audiência subiu de um patamar de 100 mil usuários para se aproximar da faixa de 350 mil usuários. Esse primeiro aumento súbito ocorrido no dia 02/04/2017 deveu-se principalmente a uma postagem emitida às 06h50min que elevou a soma da audiência de 121.540 para 259.494, que teve como usuário emissor "zehdeabreu", representante do famoso ator brasileiro José de Abreu, possuidor de grande número de seguidores no Twitter (137.954). Outras elevações de magnitude menor podem ser observadas no gráfico, mas foram geradas por uma grande quantidade de postagens de usuários com menor número de seguidores associados. A quantidade final de usuários atingidos (audiência) chegou a 343.756.

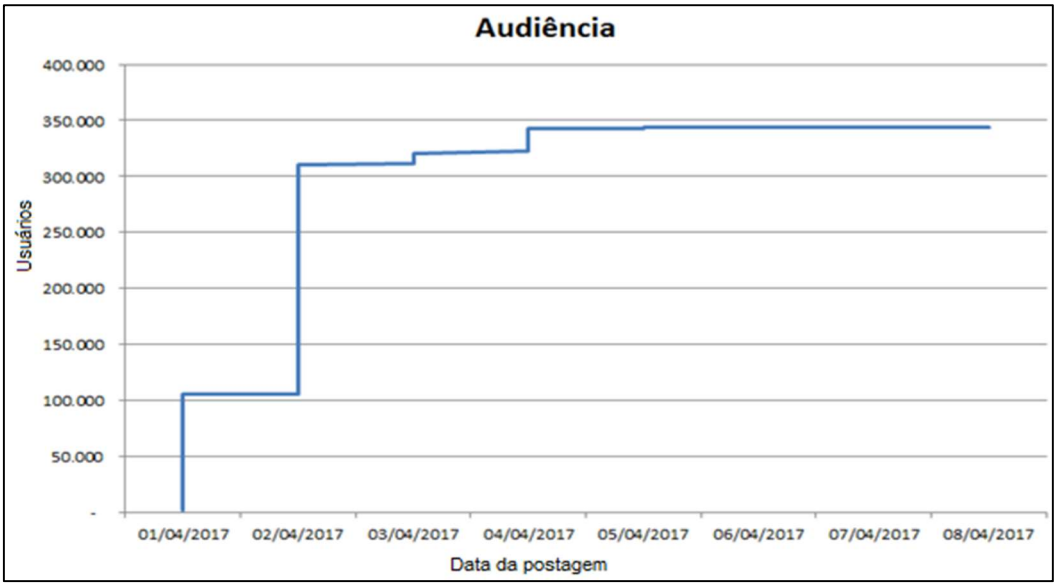

Figura 22 – Gráfico de evolução da audiência da coleta "Pink Floyd anuncia turnê Brasil" (fonte: autor).

A Tabela 11 apresenta a classificação de usuários da rede por ordem descendente da métrica PageRank dos vértices associados. Apresentou-se mais um caso mostrando não correlação entre essa métrica e o número de seguidores, mas mantendo o padrão de correlação positiva entre o grau de entrada e a centralidade de intermediação observado em todas as outras coletas.

Usuário	Grau de Entrada	Grau de Saída	Central. Intermed.	Central. Proximid.	Central. Autov.	PageRank	Coef. Clust. Local	Seguidores
ivanmagoo48	22	1	420,000	0,048	0,196	10,407	0,000	4075
curtamaissp	13	1	132,000	0,083	0,000	6,276	0,000	9
jnflesch	5	0	20,000	0,200	0,000	2,838	0,000	187178
giovannkfkd	4	1	6,000	0,333	0,000	2,168	0,000	41968
bruno1609vaz	2	3	2,000	0,500	0,000	1,458	0,000	706

Tabela 11 – Usuários mais influentes por ordem de PageRank da coleta "Pink Floyd anuncia turnê Brasil".

4.6. Notícia: Gritaram “Prendam Lula!” na encenação da peça teatral “A Paixão de Cristo”

Notícia falsa divulgada na segunda semana de abril de 2017 informava que, durante o famoso espetáculo teatral “A Paixão de Cristo”, encenado na cidade de Nova Jerusalém anualmente no período da Páscoa, espectadores teriam interrompido a apresentação com gritos de “Prendam o Lula!” no momento da peça em que se encenava a conhecida cena de pergunta à população a quem, entre Jesus e Barrabás, eles queriam salvar. A farsa pôde ser descoberta através da divulgação posterior do vídeo da apresentação, no qual não se viu nenhuma interrupção semelhante à mencionada na notícia. As métricas da rede total relacionada à essa notícia falsa, formada por todas as postagens obtidas na coleta, são apresentadas na Tabela 12.

Métrica	Valor	Métrica	Valor
Vértices	126	Componentes	4
Arestas	126	Componente máxima	122
Arestas duplicadas	0	Diâmetro	2
Coef. clusterização global	0	Diâmetro médio	1,966689
Postagens órfãs	4	Densidade	0,007746

Tabela 12 – Métricas básicas da rede total da coleta “Gritaram Prendam Lula Paixão de Cristo”.

A expressão de busca utilizada para a pesquisa da notícia foi “Gritaram Prendam Lula Paixão de Cristo” e sua coleta foi feita no dia 17/04/2017. O conjunto resultado foi formado por 126 postagens de 126 usuários, sendo a primeira postagem emitida no dia 15/04/2017 às 22h20min e a última no dia 17/04/2017 às 00h47min.

Essa janela cronológica definida pelo intervalo de datas entre a divulgação da notícia – 15/04/2017 – e a própria data de realização da coleta – 17/04/2017 – confirma mais uma vez que os dados são um recorte de postagens emitidas até o momento da coleta, acreditando-se na possível existência de outras postagens relacionadas à notícia em datas posteriores. O desenho da rede total pode ser visto na Figura 23.

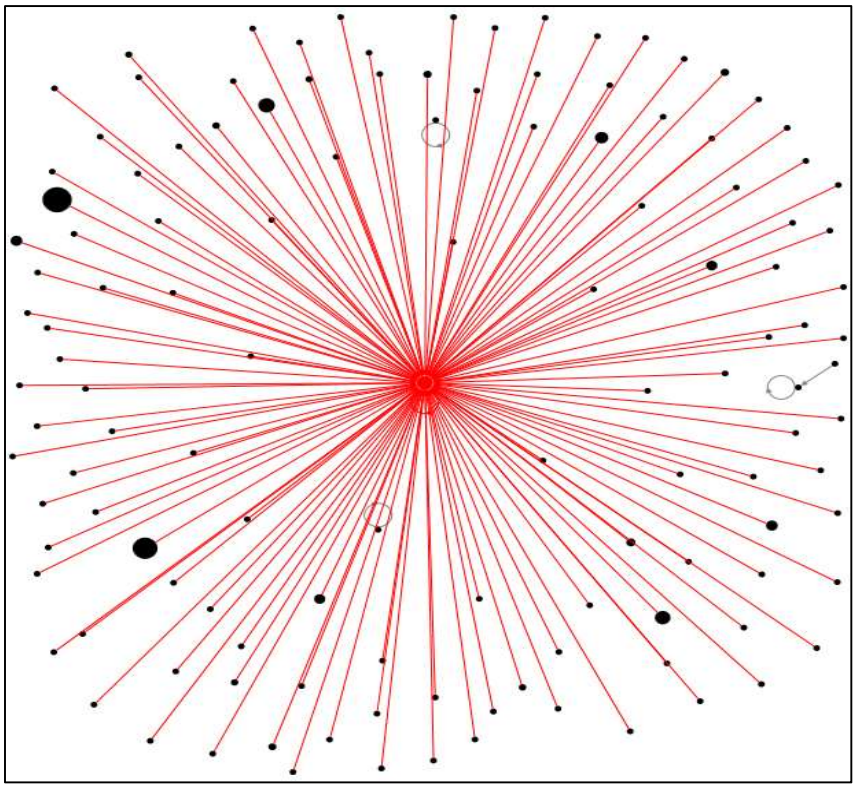

Figura 23 – Desenho da rede total da coleta “Gritaram Prendam Lula Paixão de Cristo” (fonte: autor).

Comparando ao padrão apresentado pelo desenho das coletas anteriores, o primeiro destaque é a pequena quantidade de postagens órfãs. Isso significa que a grande maioria das postagens não foi de autoria própria, mas sim repasses ou respostas a outras postagens emitidas anteriormente (ou a uma única postagem, como sugerido no desenho). Dessa vez não foram encontradas arestas duplicadas, mostrando inexistência de repetições de referências entre dois usuários.

A componente máxima identificada agrupa a quase totalidade dos vértices da rede (122 de um total de 126). A densidade calculada de 0,007746 mantém a característica observada em outras coletas de rede esparsa A métrica de coeficiente de agrupamento local em zero demonstra inexistência de triangulações entre vértices.

As primeiras informações justificam o desenho observado de uma rede no padrão estrela, com a maioria dos vértices ligados a um vértice central. Com o tamanho dos vértices diretamente proporcional ao número de seguidores, é possível observar que o vértice com mais alto grau de entrada, em destaque próximo ao centro do desenho da Figura 23 com suas arestas adjacentes, não representa o usuário de maior número de seguidores.

O valor de diâmetro manteve-se em dois, igualando-se às quatro redes analisadas formadas pelas coletas das notícias falsas "vacina febre amarela veneno mortal", "Arlindo cruz morre", "Taxista que xingou Jesus morre em acidente" e "Pink Floyd anuncia turnê Brasil", mostrando um padrão relativamente estável dessa métrica. Apesar de haver uma expectativa inicial de que os valores de diâmetro das redes representativas das coletas seriam baixos devido a uma tendência esperada de topologia em estrela, surpreende que o diâmetro se apresente com valor fixo em dois em quase todas as redes analisadas até agora.

O histograma de distribuição de graus da Figura 24 mostra a continuação do comportamento observado em todas as coletas anteriores, no que se refere à maior probabilidade de localização de vértices de graus baixos e baixas probabilidades de localização de vértices de grau mais alto. Para facilitar a visualização do gráfico, foram cortados do eixo X os graus de 11 a 121, que tiveram probabilidade zero nessa rede.

Foi possível verificar na série de grau de entrada, entretanto, que foi maior a concentração de probabilidades de vértices de grau zero nessa rede, diferentemente da maioria das redes de outras coletas que apresentavam a maior probabilidade nos vértices de grau um. Isso se encaixa na topologia estrela já comentada, que nessa rede se mostrou bastante evidente desde o início de sua análise. Sob a ótica do grau de saída o padrão foi semelhante ao observado em outras coletas, com probabilidade de 100% de localização de vértices de grau um.

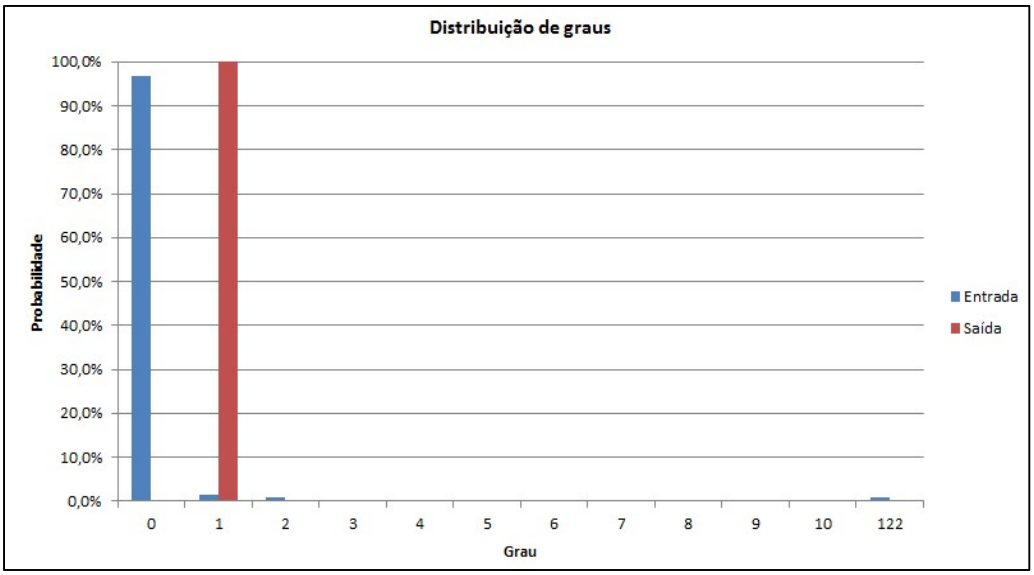

Figura 24 – Distribuição de graus de entrada e saída da coleta “Gritaram Prendam Lula Paixão de Cristo” (fonte: autor).

Essa notícia falsa teve seu primeiro emissor numa postagem emitida no dia 15/04/2017 às 22h20min. O perfil de usuário “teresinhalopes” – aparentemente um perfil de pessoa física –se apresentou como o vértice com o número mais alto de referências a sua postagem original, atingindo grau de entrada de 122.

A evolução da audiência em função do tempo é mostrada na Figura 25. Continuou-se identificando saltos súbitos na audiência através de postagens de usuários com alto número de seguidores em comparação com o restante dos usuários presentes na rede. Analisando o gráfico em maior detalhe, o maior incremento de audiência foi observado logo no primeiro dia – 02/04/2017 – explicado principalmente por duas postagens emitidas às 22h31min e às 22h52min, de usuários emissores que possuíam uma quantidade de seguidores de 46.276 e 64.458, respectivamente. Foi possível observar outro incremento destacado de audiência no dia 16/04/2017, mas dessa vez explicado por uma grande quantidade de postagens de usuários com quantidades relativamente menores de seguidores. O total de audiência atingido na última postagem foi de 330.794 usuários.

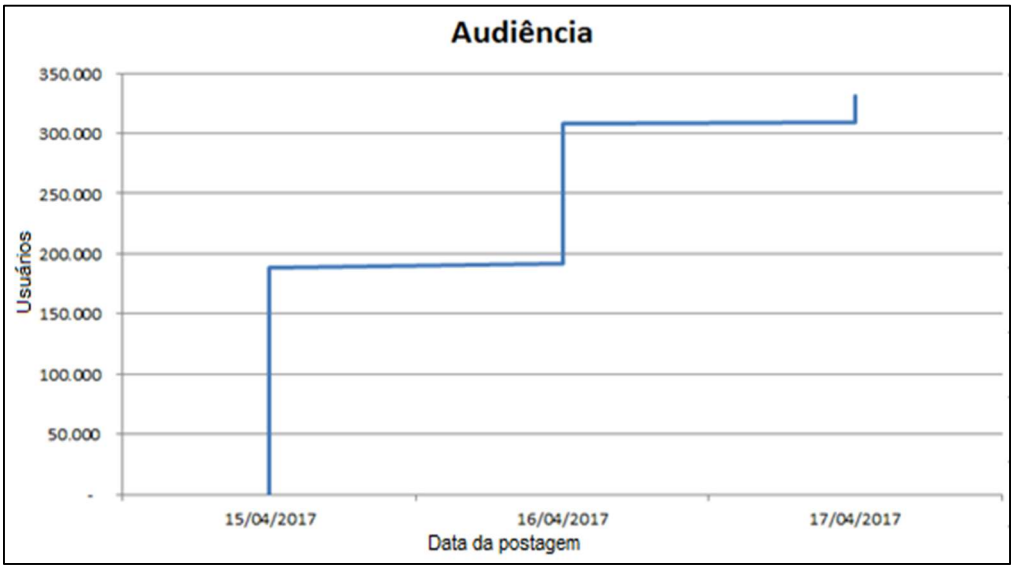

Figura 25 – Gráfico de evolução da audiência da coleta “Gritaram Prendam Lula Paixão de Cristo” (fonte: autor).

A Tabela 13 mostra a classificação de usuários da rede por ordem descendente da métrica PageRank dos vértices associados. A liderança nessa métrica do usuário “teresinhalopes” foi acompanhada da liderança também nas métricas de grau de entrada e centralidade de intermediação, como em todas as outras coletas, mas neste caso também liderando no número de seguidores.

Usuário	Grau de Entrada	Grau de Saída	Central. Intermed.	Central. Proximid.	Central. Autov.	PageRank	Coef. Clust. Local	Seguidores
teresinhalopes	122	1	14520,000	0,008	0,087	56,347	0,000	20918
boatosorg	2	1	0,000	1,000	0,000	1,298	0,000	3343
apaccinifot	1	1	0,000	0,000	0,000	1,000	0,000	447
wanderander	1	1	0,000	0,000	0,000	1,000	0,000	51
maximusjunior	0	1	0,000	1,000	0,000	0,702	0,000	930

Tabela 13 – Usuários mais influentes por ordem de PageRank da coleta “Gritaram Prendam Lula Paixão de Cristo”.

4.7. Notícia: Esposa do juiz Sérgio Moro está ligada a esquema de corrupção na APAE

Essa notícia falsa circulou em abril de 2017 e informava que a esposa do juiz federal Sérgio Moro – um dos principais representantes do poder judiciário no desenrolar da operação lava-jato – estaria envolvida num esquema de corrupção na APAE (Associação de Pais e Amigos dos Excepcionais) do estado do Paraná. A própria notícia falsa definia como fonte da informação o jornalista Luís Nassif, que responsavelmente esclareceu que o mal-entendido foi provocado por uma interpretação incorreta de uma de suas palestras, realizada em 2015, em que afirmava a existência de um esquema de corrupção na APAE e citava que a esposa de Sérgio Moro trabalhava no setor jurídico da organização, mas que não havia nenhum processo aberto que estivesse investigando a participação dela nesse esquema. As métricas da rede total relacionada à essa notícia, formada por todas as postagens obtidas na coleta, são apresentadas na Tabela 14.

Métrica	Valor	Métrica	Valor
Vértices	203	Componentes	23
Arestas	344	Componente máxima	162
Arestas duplicadas	60	Diâmetro	5
Coef. clusterização global	0,31597881	Diâmetro médio	3,12054
Postagens órfãs	56	Densidade	0,0066819

Tabela 14 – Métricas básicas da rede total da coleta “Mulher Moro esquema corrupção APAE”.

A expressão de busca utilizada para a pesquisa dessa notícia foi “Mulher Moro esquema corrupção APAE”, e sua coleta foi feita no dia 01/05/2017, formando um

conjunto resultado de 344 postagens de 203 usuários, sendo a primeira postagem emitida no dia 28/04/2017 às 02h10min e a última no dia 01/05/2017 às 13h01min.

Novamente, foi observado o padrão quase absoluto (com exceção da coleta "Tapioca pode matar") de intervalo de datas compreendido entre a data da divulgação da notícia – 28/04/2017 – e a própria data de realização da coleta – 01/05/2017 – trazendo a informação de que o universo de postagens encontradas poderia ser estendido com novas coletas posteriores. O desenho da rede total pode ser visto na Figura 26.

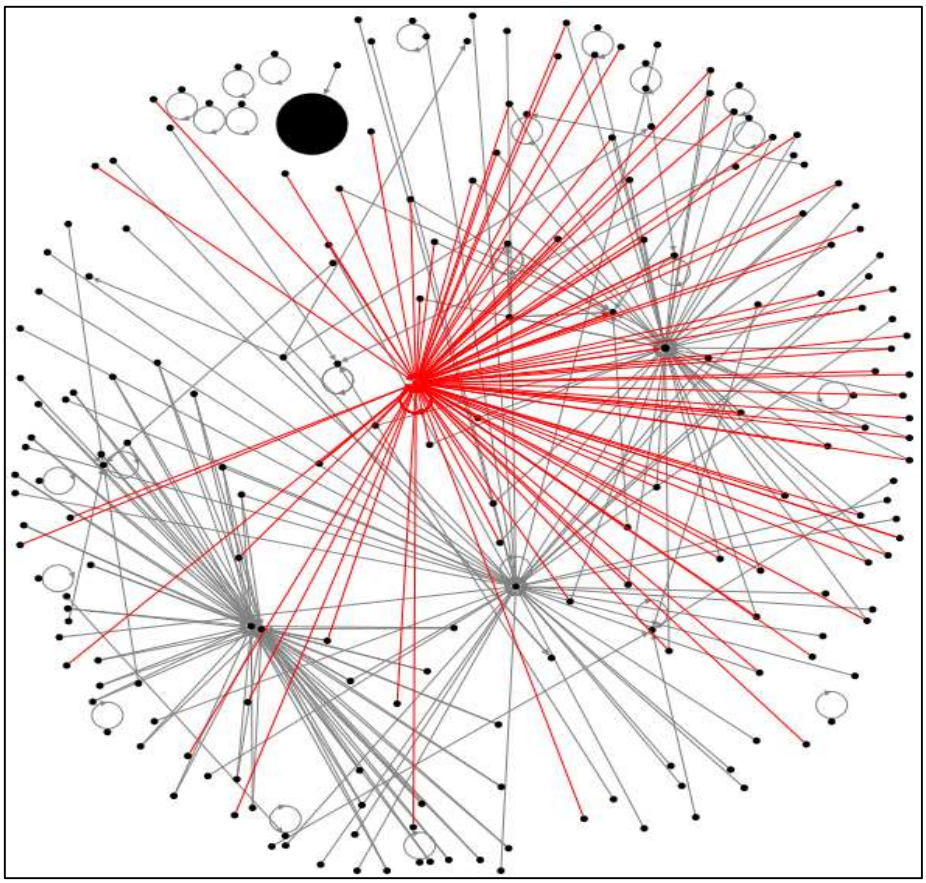

Figura 26 – Desenho da rede total da coleta "Mulher Moro esquema corrupção APAE" (fonte: autor).

Fazendo uma primeira análise da rede e comparando ao observado nas coletas anteriores, verifica-se uma quantidade baixa de postagens órfãs, caracterizando essa rede, sob esse aspecto específico, como similar às redes coletadas relacionadas às notícias falsas "Gritaram Prendam Lula Paixão de Cristo" e "Tapioca pode matar". A relativamente alta quantidade encontrada de arestas duplicadas (60), representando 17,44% do total de arestas, destaca uma característica que não teve similaridade com as outras coletas.

A componente máxima encontrada agrupou quase todos os vértices da rede (162). A densidade calculada foi de 0,0066819, mantendo o padrão observado em todas as coletas de rede esparsa, mostrando um número baixo de ligações entre os vértices em relação ao potencial máximo da rede, característica de todas as redes analisadas até aqui. Por outro lado, o coeficiente de agrupamento global de 0,31597881 indica um número relativamente alto em comparação com o observado em todas as outras coletas, podendo ser explicado pela existência de alguns vértices centrais com alto grau de entrada que acabam por ser referenciados em conjunto por vértices de menor grau, gerando um maior número de triangulações entre os vértices da rede.

As métricas analisadas até aqui são condizentes com o desenho observado de uma rede no padrão estrela, com a maioria dos vértices ligados a vértices centrais, com a ressalva de que nessa rede foi possível observar quatro vértices centrais que merecem destaque.

Seguindo com a estratégia de dimensionar os vértices de acordo com seu número de seguidores, verifica-se visualmente a não correlação dessa variável com a métrica de grau de entrada. O vértice de maior grau de entrada é mostrado em destaque próximo ao centro do desenho da Figura 26 com suas arestas adjacentes e o vértice com maior número de seguidores está localizado no alto da figura, com grau de entrada um.

O diâmetro foi calculado com valor cinco, apresentando um valor atípico em relação ao padrão observado em todas as outras coletas, que tinham o valor dessa métrica sempre próximo de dois.

A análise do histograma de distribuição de graus, apresentado na Figura 27, demonstra o mesmo padrão de probabilidades observada nas coletas anteriores e compatível com a distribuição de graus esperada para redes de livre escala, com grande probabilidade de localização de vértices com graus baixos e baixas probabilidades de localização de vértices de grau mais alto. Para facilitar a visualização do gráfico, foram cortados do eixo X os graus de probabilidade zero entre 10 e 30 e entre 50 e 82.

Analisando especificamente a série relativa ao grau de saída, pôde-se observar a mesma concentração nos graus baixos (inferiores a três) observada em todas as coletas, mas dessa vez com uma distribuição mais dividida entre os graus um e dois, apresentando nessa característica específica similaridade com a coleta “Tapioca pode matar”.

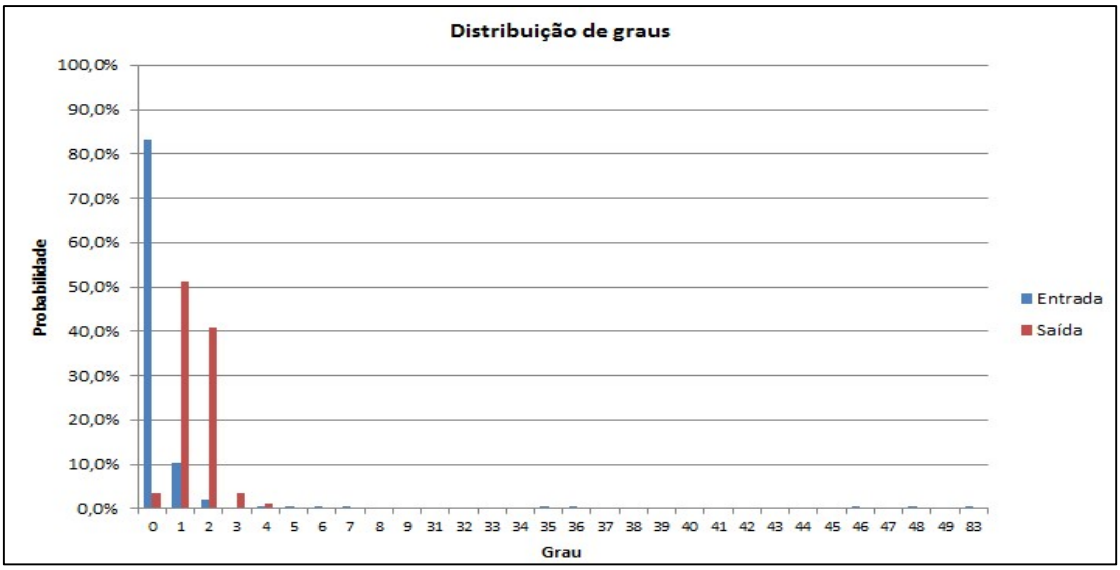

Figura 27 – Distribuição de graus de entrada e saída da coleta “Mulher Moro esquema corrupção APAE” (fonte: autor).

O primeiro emissor dessa notícia falsa emitiu postagem no dia 28/04/2017 às 02h10min. O usuário “eduardonino” – aparentemente associado a pessoa física – concentrou o maior número de referências diretas, atingindo grau de entrada 83. Entretanto, diferentemente do observado em todas as coletas anteriores, essa rede apresentou outros quatro vértices (“pelademocracia1”, “maluz_br”, “jornalggn” e “sandraselmadere”) que merecem destaque por valores também consideráveis de grau de entrada (46, 48, 36 e 35, respectivamente).

A Figura 28 mostra a evolução da audiência em função do tempo, onde é possível observar um aumento mais cadenciado do total de seguidores atingidos pela notícia falsa. Esse padrão de crescimento se diferencia de todas as outras coletas, com exceção da relacionada à notícia “Gritaram Prendam Lula Paixão de Cristo” que também apresentou crescimento escalonado em dias distintos. Observa-se um crescimento destacado da audiência logo no primeiro dia (28/04/2017) provocado principalmente por duas postagens emitidas às 16h52min e às 17h, de usuários emissores que possuíam uma quantidade de seguidores de 18.942 e 11.097, respectivamente. Outro incremento destacado de audiência é observado no dia 29/04/2017, explicado principalmente por duas postagens emitidas às 01h05min e 03h08min, com seguidores associados aos seus usuários emissores de 16.595 e 57.394, respectivamente. Um último salto destacado se dá no dia 01/05/2017, por várias postagens de usuários com quantidade relativamente pequena de seguidores, mas com duas postagens que merecem destaque emitidas às 02h53min e 11h51min, com

quantidade de seguidores dos seus usuários emissores de, respectivamente, 15.174 e 26.488. A audiência final atingida no momento da última postagem foi de 436.099 usuários.

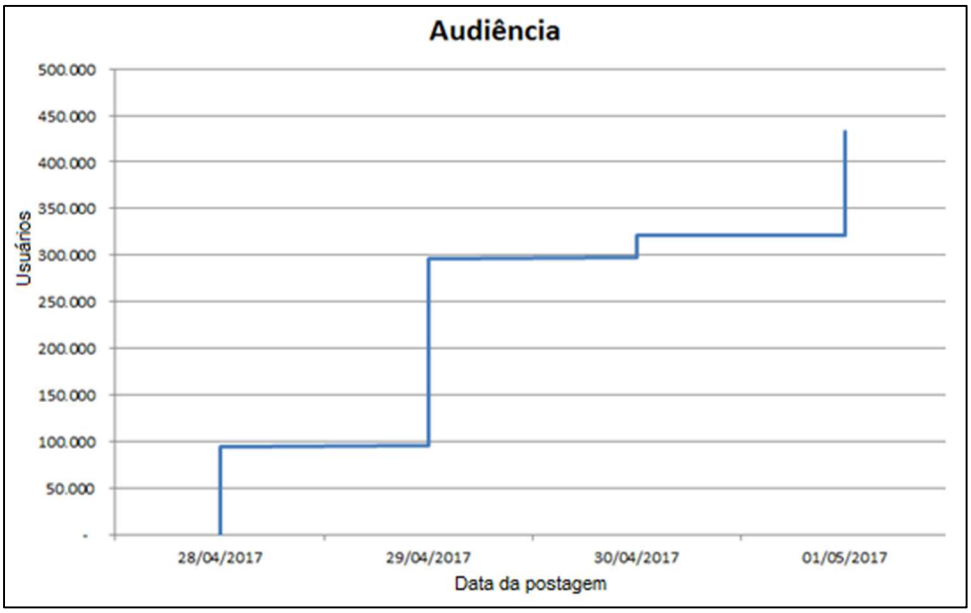

Figura 28 – Gráfico de evolução da audiência da coleta “Mulher Moro esquema corrupção APAE” (fonte: autor).

A Tabela 15 apresenta a classificação de usuários da rede por ordem descendente da métrica PageRank dos vértices associados. O usuário “eduardonino” foi o usuário mais influente sob os enfoques desse indicador e da centralidade de intermediação, mas assim como observado nas coletas “Pink Floyd anuncia turnê Brasil”, “Arlindo Cruz morre” e, em menor intensidade, na coleta relacionada à notícia “Taxista que xingou Jesus morre em acidente”, o número de seguidores não mostrou correlação positiva com o grau de influência medido por essas métricas de centralidade globais.

Usuário	Grau de Entrada	Grau de Saída	Central. Intermed.	Central. Proximid.	Central. Autov.	PageRank	Coef. Clust. Local	Seguidores
eduardonino	83	3	18264,167	0,003	0,083	24,104	0,008	8066
pelademocracia1	46	1	11464,700	0,003	0,003	18,522	0,000	1217
maluz_br	48	1	1127,000	0,002	0,056	12,732	0,021	57394
jornalggn	36	0	4594,067	0,002	0,002	8,833	0,028	23515
sandraselmadere	35	1	4594,067	0,002	0,002	8,833	0,028	5578

Tabela 15 – Usuários mais influentes por ordem de PageRank da coleta “Mulher Moro esquema corrupção APAE”.

4.8. Notícia: Delegado que investigava a morte de Teori Zavascki foi assassinado

Essa notícia falsa circulou no final de maio e informava sobre o assassinato a tiros do delegado da polícia federal responsável pelo inquérito que investigava a morte do ministro do Supremo Tribunal Federal (STF) Teori Zavascki, que respondia pelos pedidos relacionados à famosa operação Lava-jato. Algum tempo depois, a própria polícia federal divulgou nota de esclarecimento informando que o delegado assassinado foi o que abriu o processo em Angra dos Reis, mas que todo o decorrer da investigação está sendo tratado em Brasília/DF, sob a responsabilidade de outro delegado também da polícia federal. Informações posteriores também esclareceram que as circunstâncias em que o delegado foi assassinado foram de tiroteio associado a uma ação policial da qual ele fazia parte, na cidade de Florianópolis. A divulgação dessa notícia falsa foi muito abrangente, possivelmente pelo grande interesse com que a população vem acompanhando os desdobramentos da operação Lava-jato, que tem levado à condenação de um grande número de integrantes da elite política e empresarial do Brasil. As principais métricas da rede total, formada por todas as postagens obtidas na coleta, são apresentadas na Tabela 16.

Métrica	Valor	Métrica	Valor
Vértices	6489	Componentes	1084
Arestas	7759	Componente máxima	5271
Arestas duplicadas	314	Diâmetro	12
Coef. clusterização global	0,01240548	Diâmetro médio	3,987033
Postagens órfãs	1332	Densidade	0,00015163

Tabela 16 – Métricas básicas da rede total da coleta “delegado inquérito Teori Zavascki assassinado”.

A coleta de postagens dessa notícia foi feita no dia 03/06/2017 com a expressão de busca “delegado inquérito Teori Zavascki assassinado” e retornou um total de 7.759 postagens e 6.489 usuários, sendo a primeira postagem emitida no dia 31/05/2017 às 12h35min e a última no dia 03/06/2017 às 17h19min.

Como coleta que trouxe a maior quantidade de postagens até esse momento, a rede montada a partir desses dados permitiu observar que características são distintas ou similares às coletas anteriores, com tamanho relativamente menor. O desenho da rede total pode ser visto na Figura 29.

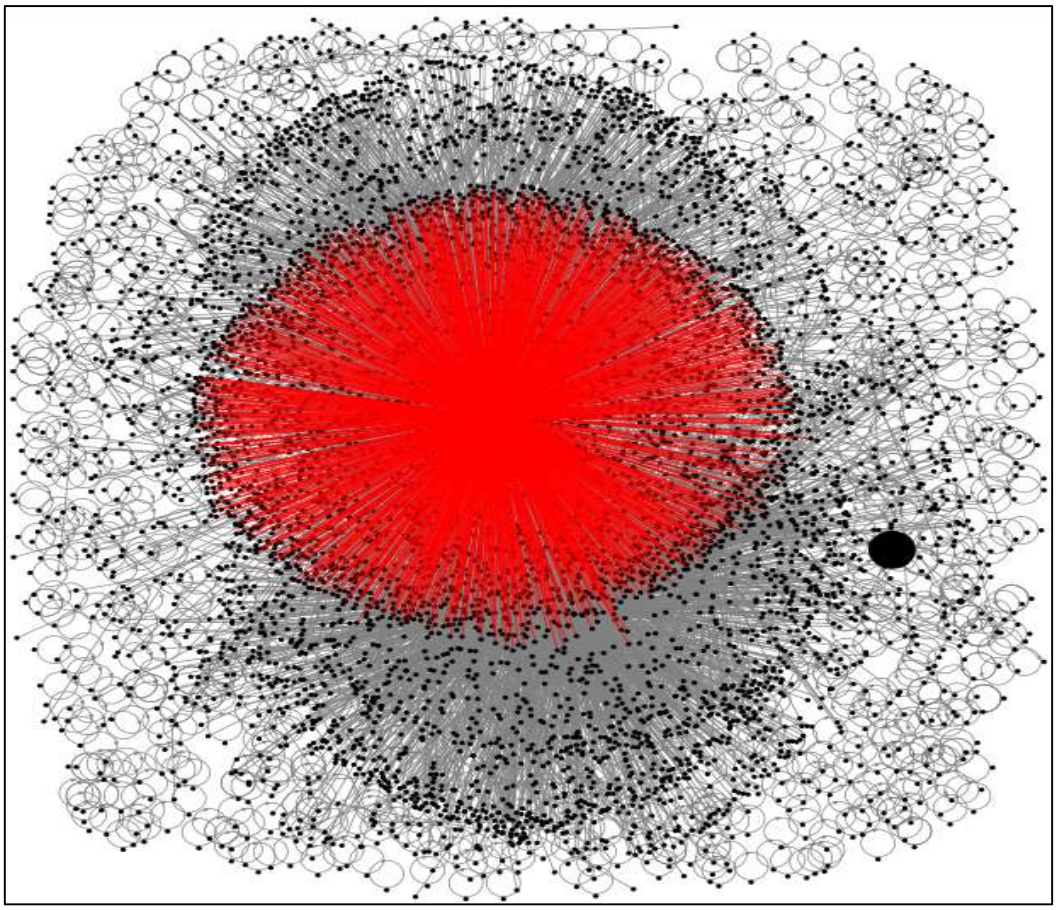

Figura 29 – Desenho da rede total da coleta “delegado inquérito Teori Zavascki assassinado” (fonte: autor).

Analisando-se o desenho da Figura 29, o primeiro destaque é a grande quantidade de vértices e arestas visíveis nessa coleta, que trouxe um desafio muito grande para o algoritmo de desenho de redes proposto por FRUCHTERMAN e REINGOLD (1991), que não logrou êxito em evitar a frequente superposição visual das arestas. Entretanto, ainda é possível identificar no desenho o grande número de postagens órfãs posicionadas pelo algoritmo de desenho na periferia do grafo. O número de arestas duplicadas encontrado foi de 314, aproximadamente 4% do total de arestas, o que leva à conclusão que a maioria das interações entre os usuários foi única no período analisado. Como estabelecido na metodologia, os vértices têm seu tamanho proporcional ao número de seguidores dos usuários correspondentes, sendo possível identificar que o usuário com maior número de seguidores – posicionado no centro inferior do desenho – não foi o mesmo vértice cujo usuário teve o maior número de referências dos outros usuários (vértice com maior grau de entrada). Esse último vértice e suas arestas associadas estão destacados ao centro da Figura 29. Esse vértice corresponde ao usuário “comunacritico”.

O valor alto de diâmetro – 12 – contrastou fortemente com os valores baixos dessa métrica encontrados em todas as outras coletas, que se situaram sempre próximos de dois. Entretanto, o valor do diâmetro médio calculado em 3,987033 apresentou uma grande diferença em relação ao diâmetro, não sendo possível descartar uma situação pontual e atípica que pode ter provocado esse valor alto no cálculo original do diâmetro. O valor de 5.271 para a componente máxima indica que um alto percentual dos vértices (> 81,2%) estão conectados em componente única.

A baixa densidade observada (0,00015163) foi dentro da expectativa estabelecida no início das coletas – e, de fato, observada em todas as analisadas até aqui – de que as coletas do Twitter tenham muitas respostas e repasses a postagens de usuários centrais de maior audiência. A métrica de coeficiente de agrupamento local calculada em 0,01240548 demonstrou baixo número de triangulações entre vértices. A expectativa de rede centralizada em poucos usuários também foi comprovada no histograma de distribuição de graus exibido na Figura 30, que para melhor visualização teve removidos do eixo X os graus de entrada e saída com probabilidade zero de localização no grafo.

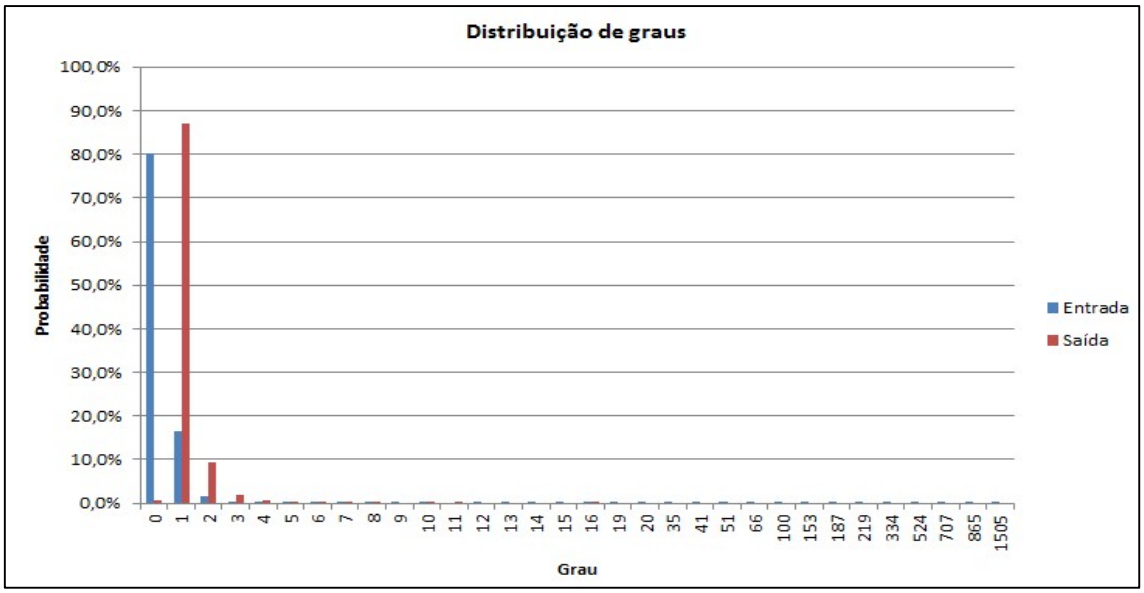

Figura 30 – Distribuição de graus de entrada e saída da coleta “delegado inquérito Teori Zavascki assassinado” (fonte: autor).

O primeiro emissor de postagem relacionada a notícia falsa “delegado inquérito Teori Zavascki assassinado” foi o usuário “zerohora”, que representa o site associado ao conhecido jornal “Zero Hora”. Essa postagem foi emitida no dia 31/05/2017 às 12h35min e não teve referências posteriores feitas por outras postagens. A última

postagem foi emitida no dia 03/06/2017 às 17h19min pelo usuário “edilsonfogaca”, indicando que o tempo final da coleta foi muito próximo da própria realização da busca, sugerindo possibilidade de que outras postagens sobre o mesmo assunto possam ser localizadas em tempo futuro. O emissor com maior grau de entrada – 1.505 – foi o usuário “comunacritico”, que representa uma pessoa física com posições políticas contrárias ao sistema capitalista.

A Figura 31 exibe o gráfico que mostra a evolução da audiência em função do tempo para essa coleta. Logo no primeiro dia – 31/05/2017 – houve um aumento forte da audiência, provocada principalmente por quatro postagens de usuários com grande número de seguidores. A primeira dessas postagens foi também a primeira da coleta, emitida às 12h35min pelo usuário “zerohora”, que tem cadastrado um total de 1.063.307 seguidores. A segunda postagem foi emitida às 14h17min pelo usuário “catracalivre”, também com grande quantidade de seguidores: 1.617.731. A terceira postagem foi emitida às 14h40min pelo usuário “xicosa”, com 1.126.531 seguidores. A quarta postagem teve sua emissão feita às 18h17min do mesmo dia, pelo usuário “revistapiaui”, com 1.324.549 seguidores. Um fato importante que merece destaque é que os usuários emissores dessas quatro postagens tiveram grau de entrada entre zero e um. Até o momento da última postagem, a audiência atingida foi de 20.187.403 leitores.

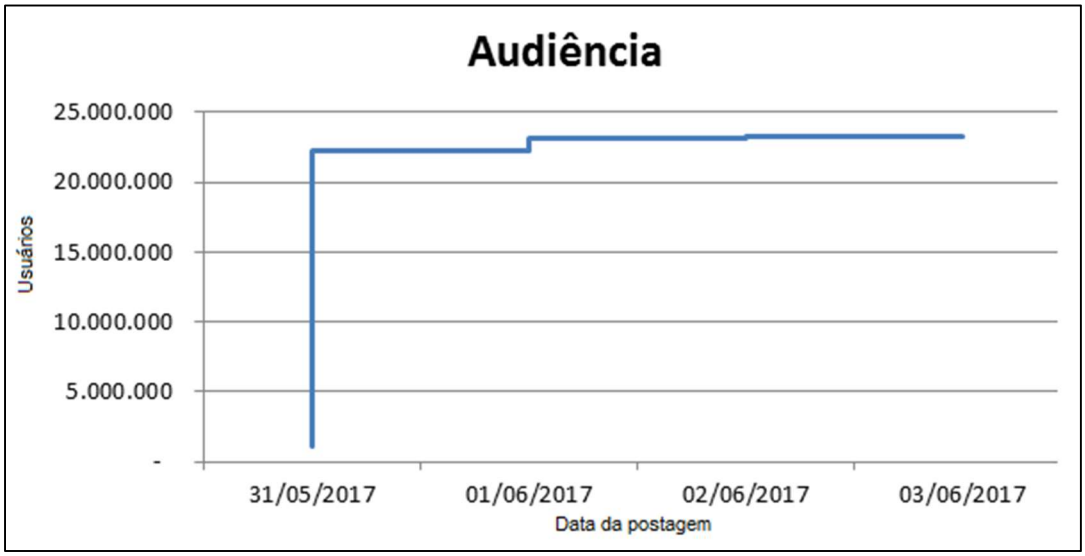

Figura 31 – Gráfico de evolução da audiência da coleta “delegado inquérito Teori Zavascki assassinado” (fonte: autor).

A Tabela 17 apresenta a classificação de usuários da rede por ordem descendente da métrica PageRank dos vértices associados. Observa-se mais uma rede na qual o número de seguidores não esteve diretamente relacionado com a influência dos

usuários/vértices centrais sob a ótica das centralidades de intermediação e PageRank, assim como visto nas coletas “Pink Floyd anuncia turnê Brasil”, “Arlindo Cruz morre”, “Taxista que xingou Jesus morre em acidente” e “Mulher Moro esquema corrupção APAE”.

Usuário	Centralidade de Intermediação	Centralidade de Proximidade	Centralidade de Autovetor	PageRank	Coeficiente de Clusterização Local	Seguidores
comunacritico	13224248,242	0,000	0,018	601,062	0,000	1232
georgmarques	8198743,671	0,000	0,002	344,223	0,000	70691
delucca	5926238,009	0,000	0,003	263,723	0,000	13593
isaacnervoso	5317511,720	0,000	0,001	208,166	0,000	28090
brasil_fotos	3675651,881	0,000	0,000	135,408	0,000	61643

Tabela 17 – Usuários mais influentes por ordem de PageRank da coleta “delegado inquérito Teori Zavascki assassinado”.

4.9. Notícia: Homossexuais terão cota em concursos públicos

A notícia falsa “Homossexuais terão cota em concursos públicos” circulou no início do ano de 2017 e foi pesquisada no dia 09/04/2017. Trazia a informação de que um projeto de lei estaria circulando no congresso nacional propondo a definição de uma cota de 15% das vagas em concursos públicos de todo o Brasil para homossexuais. A notícia ia além, indicando o deputado Marquinhos Freire (PT-BA) como autor do projeto de lei e comentando sobre a resistência oferecida por supostas entidades representativas de bissexuais que estariam indignadas por não estarem contempladas na proposta original. Tanto o projeto de lei como o próprio deputado citado nunca existiram. As principais métricas da rede total, formada por todas as postagens obtidas na coleta, são apresentadas na Tabela 18.

Métrica	Valor	Métrica	Valor
Vértices	172	Componentes	94
Arestas	186	Componente máxima	28
Arestas duplicadas	4	Diâmetro	4
Coef. clusterização global	0,06151427	Diâmetro médio	1,637705
Postagens órfãs	87	Densidade	0,0033659

Tabela 18 – Métricas básicas da rede total da coleta “Homossexuais terão cota concursos”.

A expressão de busca utilizada para a pesquisa dessa notícia foi “Homossexuais terão cota concursos” e retornou um total de 186 postagens e 172 usuários, sendo a primeira postagem emitida no dia 07/04/2017 às 22h48min e a última no dia 09/04/2017 às 22h29min.

O período da coleta, abrangendo três dias, estabelece que a janela cronológica de busca de sete dias imposta pela API do Twitter não foi determinante para limitar o universo de postagens. Mesmo assim, considerando que a data e hora da última postagem encontrada se aproximaram do momento da realização da coleta, não se pode eliminar a possibilidade de que outras postagens pudessem ser encontradas em dias posteriores.

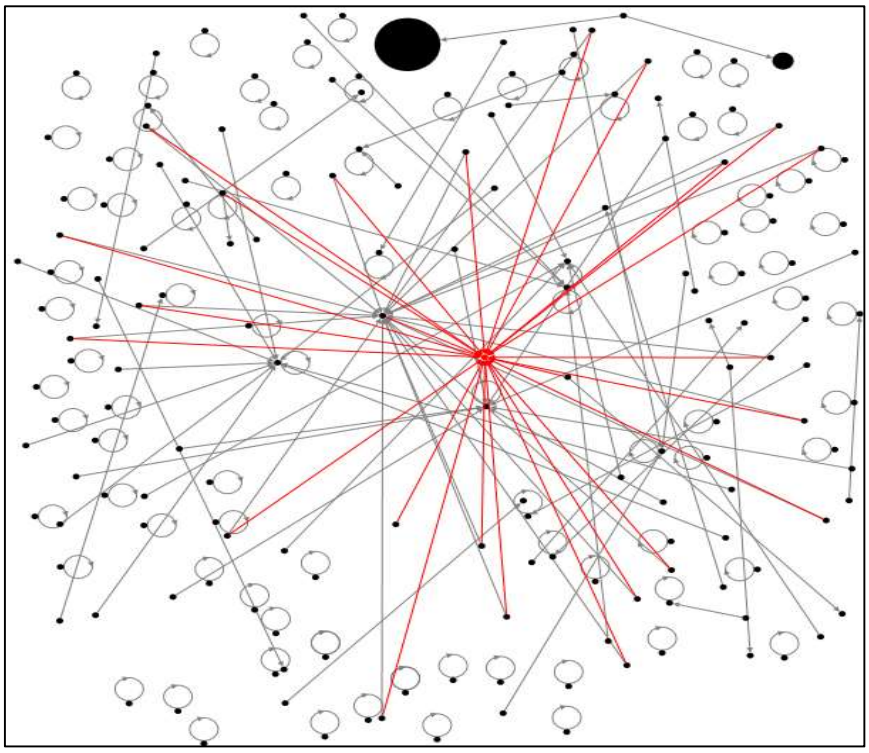

Figura 32 – Desenho da rede total da coleta “Homossexuais terão cota concursos” (fonte: autor).

O desenho da rede total relacionada à notícia “Homossexuais terão cota concursos” é apresentado na Figura 32. Mais uma vez é possível identificar o padrão visual de muitas postagens órfãs periféricas e alguns poucos vértices centrais com maior grau de entrada. Esse elevado número de postagens órfãs em relação ao total foi encontrado em todas as coletas, com exceção apenas das relacionadas às buscas das notícias “Mulher Moro esquema corrupção APAE”, “Gritaram Prendam Lula Paixão de Cristo” e “Tapioca pode matar”. Essa primeira análise do desenho encontra respaldo na

quantidade de postagens órfãs: 87, indicando que quase a metade das postagens (46,7%) não teve referências de outros usuários.

O vértice com mais alto grau de entrada, e por esse motivo posicionado mais ao centro do desenho, foi destacado com suas arestas adjacentes na Figura 32. Aqui se observa o terceiro caso em que a quantidade de ligações dos vértices não foi positivamente correlacionada com a quantidade de seguidores do Twitter.

Assim como o observado em todas as coletas anteriores, com exceção apenas da coleta relacionada à notícia “delegado inquérito Teori Zavascki assassinado”, que apresentou um valor de diâmetro igual a 12, o valor calculado de diâmetro foi baixo – quatro.

A componente máxima, com apenas 28 vértices de um total de 172, somado à elevada quantidade de componentes (94) em relação ao observado em todas as outras coletas, e à densidade (0,0033659) que se manteve baixa como em todas as coletas, permitem caracterizar essa rede como relativamente esparsa, com pouca interconexão entre seus vértices. O valor de coeficiente de agrupamento global trouxe valores baixos de triangulação de vértices (0,06151427) em comparação com o limite teórico (um).

O histograma de distribuição de graus da Figura 33 ratifica o padrão observado no desenho e indicado pelas métricas analisadas até então, de grafo em que a maioria das arestas que ligam vértices distintos o faça ligando vértices de baixo grau de entrada com vértices de alto grau de entrada. Vê-se novamente uma curva de tendência de distribuição de probabilidades muito semelhante à encontrada em redes de livre escala. Analisando apenas a série relativa à métrica de grau de saída, observa-se, novamente, uma quase completa concentração no grau um, mas com uma pequena distribuição também nos graus dois e zero que não diferenciaram essa rede de todas as outras observadas até então.

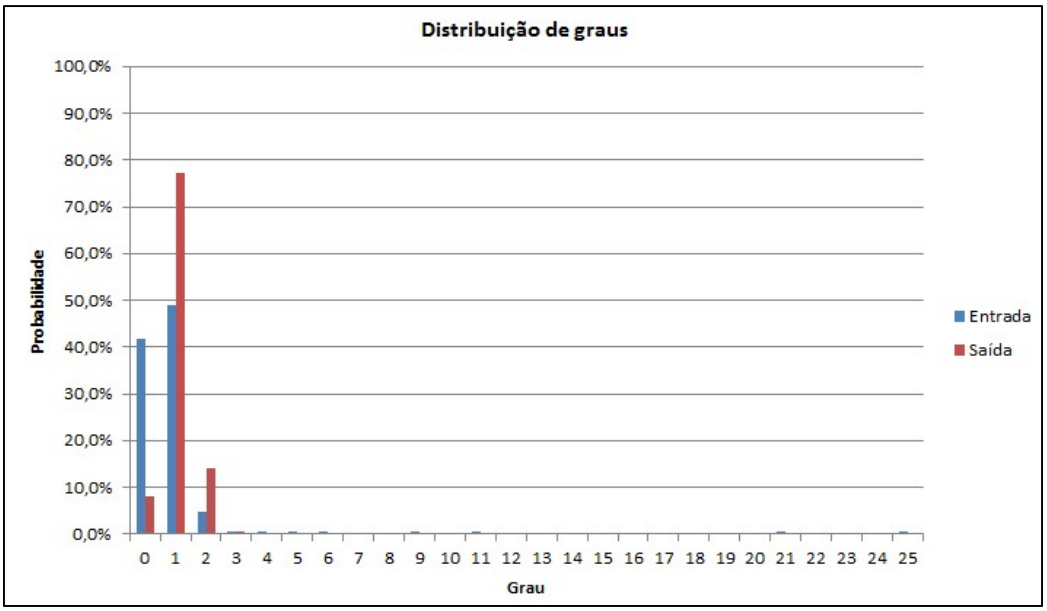

Figura 33 – Distribuição de graus de entrada e saída da coleta “Homossexuais terão cota concursos” (fonte: autor).

O primeiro emissor de postagem relacionada a notícia falsa “Homossexuais terão cota concursos” foi o usuário “ctrucolo”, que aparentemente se trata de pessoa física. Essa primeira postagem teve apenas duas referências posteriores. O emissor com maior grau de entrada foi “pxandreia”, também pessoa física, com 25 referências de outras postagens nessa coleta.

A evolução da audiência em função do tempo é mostrada na Figura 34. O padrão de saltos na audiência se assemelha mais diretamente com o observado nas coletas referentes às notícias “Mulher Moro esquema corrupção APAE” e “Gritaram Prendam Lula Paixão de Cristo”, com um acréscimo escalonado nos dias em que as postagens apareceram, sem saltos diferenciados em uma data específica. A mudança mais destacada na soma da audiência veio por postagem realizada no dia 07/04/2017 às 22h56min efetuada pelo usuário “annacelico”, com uma quantidade de seguidores (24.003) alta apenas em relação aos outros usuários dessa coleta, mas inferior às quantidades de seguidores observadas em outras redes analisadas anteriormente. Essa postagem específica elevou a soma da audiência de um valor muito baixo (202) para um patamar de 24.205, de onde passou a crescer por influência de outros usuários com quantidades menores de seguidores. Ao final do universo de postagens coletado, a audiência atingida foi de 198.411 usuários.

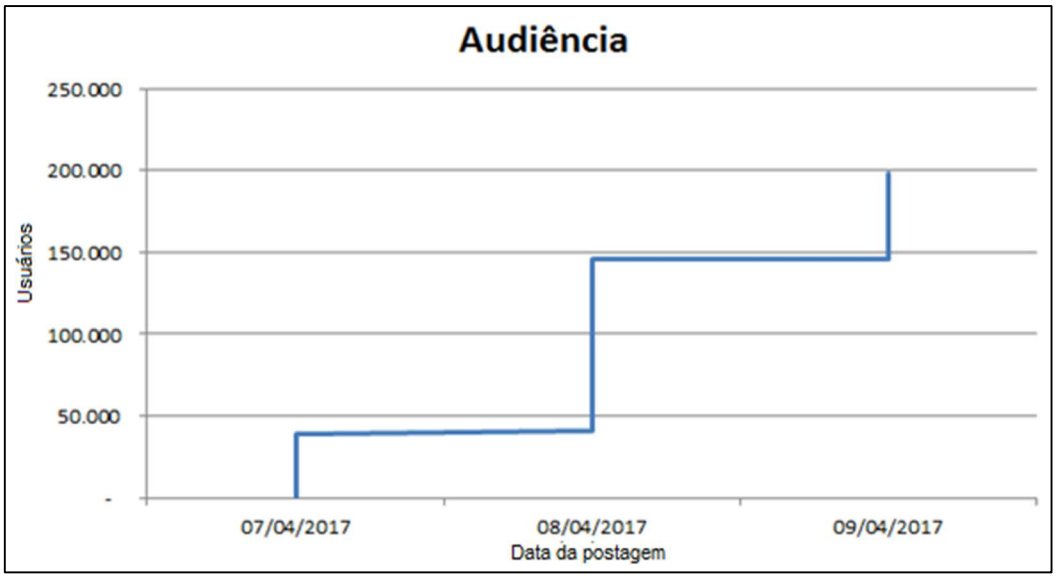

Figura 34 – Gráfico de evolução da audiência da coleta “Homossexuais terão cota concursos” (fonte: autor).

A Tabela 19 mostra a classificação de usuários da rede por ordem descendente da métrica PageRank dos vértices associados. Aqui vê-se mais uma situação em que o usuário mais influente, sob a ótica da métrica PageRank e da centralidade de intermediação, não foi o que ostentou maior número de seguidores diretos. O usuário “pxandreia”, além da liderança nas duas métricas citadas, também foi o que recebeu maior número de referências (grau de entrada), mas apresentou um total de 2.489 seguidores, baixo em relação aos usuários seguintes da classificação. Sob esse aspecto, essa rede teve similaridade com as redes formadas a partir das notícias “Pink Floyd anuncia turnê Brasil”, “Arlindo Cruz morre”, “Taxista que xingou Jesus morre em acidente”, “Mulher Moro esquema corrupção APAE” e “delegado inquérito Teori Zavascki assassinado”.

Usuário	Grau de Entrada	Grau de Saída	Central. Intermed	Central. Proximid.	Central. Autov.	PageRank	Coef. Clust. Local	Seguidores
pxandreia	25	0	446,000	0,034	0,122	6,641	0,035	2489
annacelico	21	1	210,000	0,029	0,116	5,693	0,045	24003
menezesramiro	11	1	90,000	0,100	0,000	5,359	0,000	1429
henrique_mauro1	9	1	56,000	0,125	0,000	4,443	0,000	1413
efarsas	6	1	20,000	0,200	0,000	3,073	0,000	15912

Tabela 19 – Usuários mais influentes por ordem de PageRank da coleta “Homossexuais terão cota concursos”.

4.10. Notícia: Andreas Von Richthofen encontrado na “Cracolândia”

Notícia divulgada por várias fontes de grande credibilidade jornalística no Brasil, como o jornal O Globo e o portal de notícias UOL, informou que no final de maio de 2017, Andreas Von Richthofen teria sido encontrado com distúrbios psicológicos na chamada “Cracolândia”, conhecida área de consumo de entorpecentes na cidade de São Paulo. Andreas ficou conhecido no Brasil por ser irmão de Suzane Von Richthofen, condenada pelo assassinato de seus pais em 2002, caso que gerou comoção nacional. Apesar da notícia ser verdadeira no que se refere ao fato do jovem ter sido encontrado com surto psicológico pela polícia, o local em que foi encontrado é distante 15km da região da “Cracolândia”. Isso gerou uma associação direta dos leitores dessa notícia com o consumo de drogas, que não seria a única explicação para os problemas psicológicos observados. O risco de compreensão incorreta levou a Polícia Militar a divulgar nota oficial esclarecendo o caso e também ao jornal O Globo e o portal de notícias UOL a reescreverem o conteúdo divulgado. As métricas da rede total relacionada a essa notícia falsa, formada por todas as postagens obtidas na coleta, são apresentadas na Tabela 20.

Métrica	Valor	Métrica	Valor
Vértices	597	Componentes	155
Arestas	632	Componente máxima	294
Arestas duplicadas	29	Diâmetro	8
Coef. clusterização global	0,00543192	Diâmetro médio	4,119891
Postagens órfãs	172	Densidade	0,001287

Tabela 20 – Métricas básicas da rede total da coleta “Andreas Von Richthofen encontrado Cracolândia”.

A expressão de busca utilizada para a pesquisa dessa notícia foi “Andreas Von Richthofen encontrado Cracolândia” e sua coleta foi feita no dia 03/06/2017. O conjunto resultado foi formado por 632 postagens de 597 usuários, sendo a primeira postagem emitida no dia 30/05/2017 às 22h05min e a última no dia 03/06/2017 às 18h55min.

Essa janela cronológica definida pelo intervalo de datas compreendido entre a data da divulgação da notícia – 30/05/2017 – e a própria data de realização da coleta – 03/06/2017 confirma mais um caso em que a busca obteve um recorte de postagens emitidas até o momento da realização da coleta, acreditando-se ser possível que outras

postagens ainda fossem encontradas em buscas posteriores. O desenho da rede total pode ser visto na Figura 35.

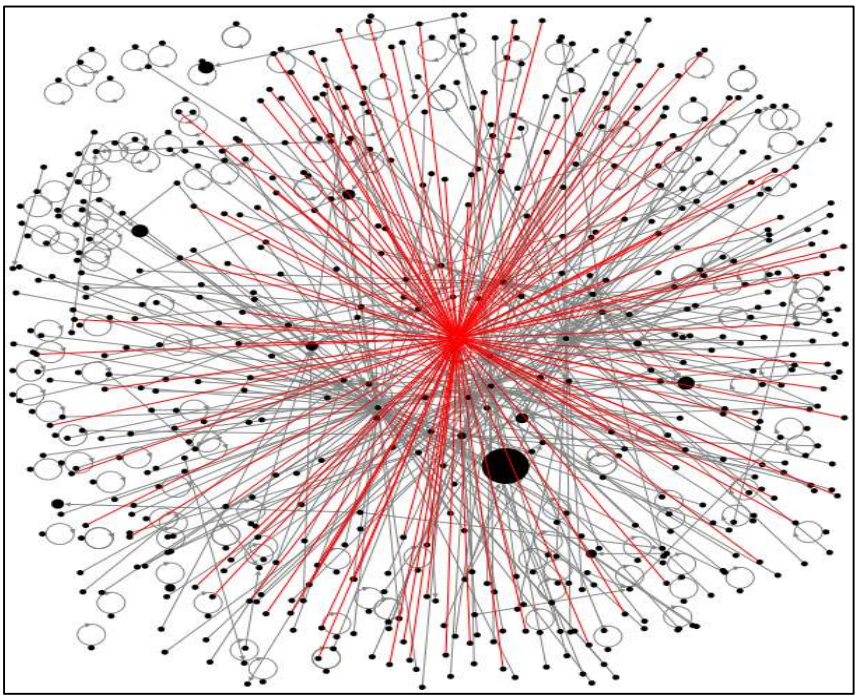

Figura 35 – Desenho da rede total da coleta “Andreas Von Richthofen encontrado Cracolândia” (fonte: autor).

Comparando ao padrão apresentado pelo desenho das coletas anteriores, é possível identificar a grande quantidade de postagens órfãs periféricas, mas não numa predominância visível que permita a conclusão de que a maioria das postagens não tenha tido referências posteriores. A quantidade de arestas duplicadas, mesmo presente, foi baixa (29) em relação ao total de arestas da rede (632), mostrando pouca repetição de referência entre pares de usuários.

A componente máxima identificada agrupa 294 vértices, valor que não chega à metade do total da rede (597). Adicionada à informação de 155 componentes conectadas no grafo e à densidade calculada de 0,001287, tem-se uma visão de rede esparsa, com múltiplas subredes presentes. Importante destacar que a densidade baixa em relação ao máximo teórico mantém-se no mesmo patamar observado em todas as outras coletas. A métrica de coeficiente de agrupamento global trouxe um valor calculado muito baixo (0,00543192), próximo de zero, mostrando a quase inexistência de triangulações entre vértices, mantendo o padrão visto em todas as outras coletas, com exceção apenas da coleta referente à notícia “Esposa do juiz Sérgio Moro está ligada a esquema de corrupção”, que trouxe um valor para essa métrica relativamente mais alto

(0,31597881), apesar de se manter distante do limite teórico que seria atingido numa rede completa.

Apesar de ser possível visualizar pelo desenho da rede uma repetição do padrão estrela, com muitos dos vértices de grau de entrada baixo ligados a vértices centrais de grau maior, a maior presença desses vértices centrais é um diferencial dessa rede. Essa primeira impressão, trazida pelo desenho, poderá ser confirmada com a análise do histograma de graus adiante. Com o tamanho dos vértices diretamente proporcional ao número de seguidores, é possível observar que o vértice com mais alto grau de entrada, em destaque próximo ao centro do desenho da Figura 35 com suas arestas adjacentes, não representa o usuário de maior número de seguidores. Padrão já observado em quase todas as coletas anteriores, com exceção das coletas “Vacina febre amarela veneno mortal” e “Tapioca pode matar”.

O valor de diâmetro foi calculado com valor oito, apresentando-se acima do observado na maioria das coletas, com exceção das associadas às notícias “delegado inquérito Teori Zavascki assassinado” e “Mulher Moro esquema corrupção APAE”, que tiveram os valores de 12 e 5 calculados para essa métrica, respectivamente. O risco de um valor de diâmetro alto influenciado por um caminho atipicamente alto ligando um par de vértices conectados qualquer da rede é descartado pelo diâmetro médio de 4,11, que também se mostra acima da média observada em quase todas as outras coletas, com exceção das duas coletas associadas às notícias “delegado inquérito Teori Zavascki assassinado” e “Mulher Moro esquema corrupção APAE”, que também tiveram essa métrica próxima do valor quatro.

O histograma de distribuição de graus da Figura 36 mostra a continuação do comportamento observado em todas as coletas anteriores, no que se refere à maior probabilidade de localização de vértices de graus baixos e baixas probabilidades de localização de vértices de grau mais alto. Especificamente em relação ao grau de entrada, a evolução se aproximou do esperado em redes de livre escala. Para facilitar a visualização do gráfico, foram cortados do eixo X alguns graus que tiveram probabilidade zero para entrada e/ou saída nessa rede.

Sob a ótica do grau de saída o padrão foi semelhante ao observado em todas as outras coletas, com alta probabilidade (>90%) de localização de vértices de grau um em busca aleatória. Esse padrão só não foi seguido totalmente à risca na rede relacionada à notícia “Mulher Moro esquema corrupção APAE”, que apesar de manter a maior

concentração de probabilidade no grau de saída um, também teve uma probabilidade relativamente alta (>40%) no grau dois.

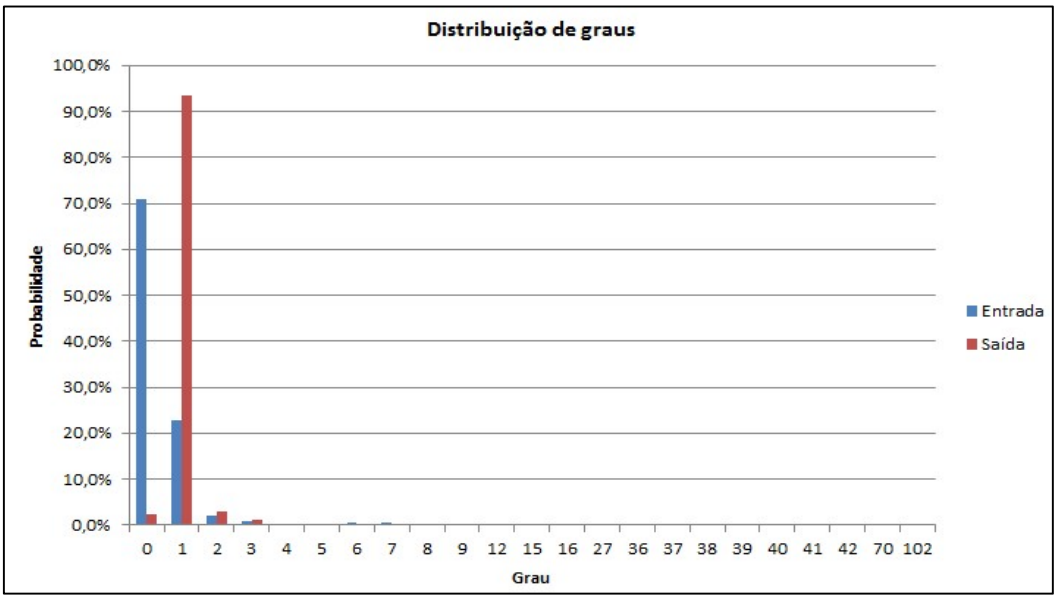

Figura 36 – Distribuição de graus de entrada e saída da coleta “Andreas Von Richthofen encontrado Cracolândia” (fonte: autor).

Essa notícia falsa teve como seu primeiro emissor o usuário “elpabrito”, com uma postagem emitida no dia 30/05/2017 às 22h05min, mas no período da coleta tendo apenas mais duas referências de outras postagens. O usuário “agorasabe” – categorizado em seu perfil no Twitter como um disseminador de “curiosidades, lugares e notícias” – foi o vértice com grau de entrada destacado em relação aos outros (102). O segundo vértice com maior valor para essa métrica alcançou valor de 70.

A evolução da audiência em função do tempo é mostrada na Figura 37. Aqui se observa um padrão que, ao mesmo tempo em que apresenta saltos escalonados nos dias que englobam a coleta, também apresenta um salto de maior dimensão no dia 31/05/2017. Essa grande mudança de patamar de audiência é explicada principalmente por duas postagens emitidas nos horários de 00h54min e 22h23min, de usuários emissores que possuíam uma quantidade de seguidores de 419.185 e 290.436, respectivamente; e por outras postagens na mesma data com quantidades também destacadas de seguidores. Importante destacar que o emissor com maior número de seguidores (446.306) emitiu postagem apenas no dia 02/06/2017, quando o número de emissores (12) foi significativamente menor que o visto no dia 31/05/2017 (450). Essa informação, analisada no contexto dessa coleta, mostra como uma grande quantidade de emissores com número menor de seguidores pode influenciar mais fortemente a

audiência atingida por uma notícia que um emissor específico que tenha elevado número de seguidores. Ao final do universo de postagens coletado, a audiência atingida foi de 2.929.462 usuários.

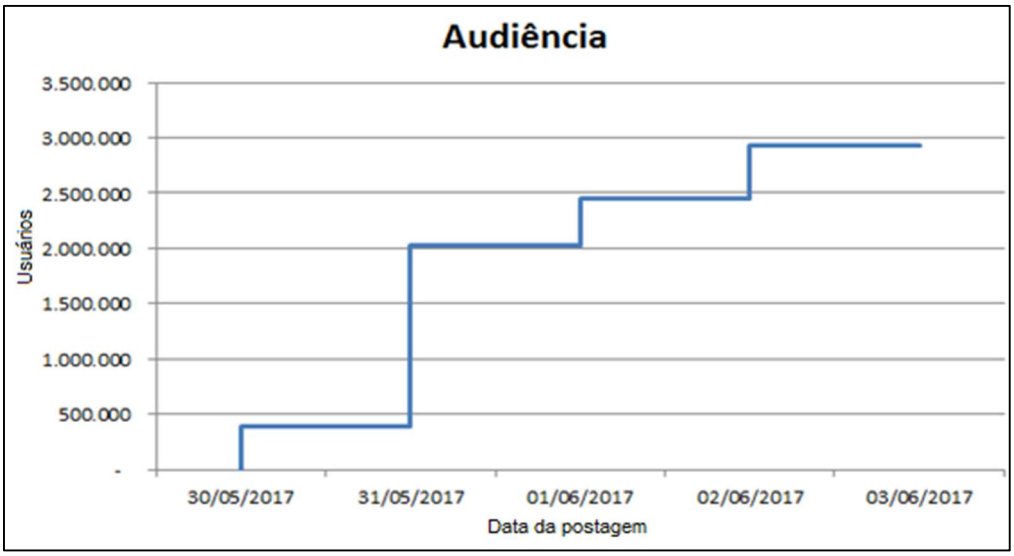

Figura 37 – Gráfico de evolução da audiência da coleta “Andreas Von Richthofen encontrado Cracolândia” (fonte: autor).

A Tabela 21 mostra a classificação de usuários da rede por ordem descendente da métrica PageRank dos vértices associados. Nessa rede foi possível observar uma concordância do usuário mais influente – “agorasabe” – sob a ótica das métricas de grau de entrada, centralidade de intermediação e PageRank, mas novamente com ausência de correlação direta visível entre essas métricas e o número de seguidores, como visto nas coletas “Pink Floyd anuncia turnê Brasil”, “Arlindo Cruz morre”, “Taxista que xingou Jesus morre em acidente”, “Mulher Moro esquema corrupção APAE”, “delegado inquérito Teori Zavascki assassinado” e “Homossexuais terão cota concursos”.

Usuário	Grau de Entrada	Grau de Saída	Central. Intermed.	Central. Proximid.	Central. Autov.	PageRank	Coef. Clust. Local	Seguidores
agorasabe	102	1	57619,733	0,001	0,091	45,530	0,000	30274
silvio_mkt	70	1	38475,033	0,001	0,003	31,169	0,000	15505
imstickingwithu	41	1	1560,000	0,025	0,000	19,133	0,000	634
marcogomes	36	1	1190,000	0,029	0,000	16,836	0,000	71022
mandafss	38	1	28899,833	0,001	0,001	16,715	0,000	4107

Tabela 21 – Usuários mais influentes por ordem de PageRank da coleta “Andreas Von Richthofen encontrado Cracolândia”.

4.11. Análise do conjunto das notícias selecionadas

A Tabela 22 sumariza as métricas das dez notícias falsas analisadas neste estudo. A seguir, para cada coluna da tabela, relacionam-se o nome da coluna com a explicação do seu conteúdo.

A coluna *Notícia,* mostra a descrição da notícia falsa analisada. A coluna *Data coleta* traz a data da realização da coleta das postagens. A coluna *Prim. post* informa a data e hora da primeira postagem da coleta. A coluna *Últim. post* traz a data e hora da última postagem da coleta. As colunas *Vért.* e *Arestas* mostram, respectivamente, a quantidade de vértices e arestas da rede. A coluna *Aresta dupla* informa a quantidade de arestas duplas encontradas na rede. A coluna *Post órfãs* traz a quantidade de postagens órfãs encontradas na coleta. A coluna *Comp* mostra a quantidade de componentes da rede. A coluna *Comp.c. max* exibe o valor da componente máxima da rede. A coluna *Diâm.* traz o valor calculado do diâmetro da rede. A coluna *Diâm. médio* informa o valor do diâmetro médio da rede. A coluna *Densid.* exibe o valor da densidade da rede. A coluna *Coef. cluster.global* mostra o valor do coeficiente de clusterização global. A coluna *Audiência final* exibe a quantidade de audiência atingida até o momento da última postagem.

Notícia	Data coleta	Prim post	Últim post	Vért	Arest	Arest dupl	Post órfãs	Comp	Comp c.max.	Diâm	Diâm médio	Densid	Coef. cluster. global	Audiência Final
Vacina da febre amarela tem veneno mortal	26/03	20/03 21:35	26/03 01:49	206	226	24	136	130	39	2	1,6961	0,001989	0,0294435	10.507.182
Morte do cantor Arlindo Cruz	23/03	17/03 21:29	23/03 01:30	53	54	2	17	16	26	2	1,7665	0,013425	0	9.161.284
Taxista que xingou Jesus morre em acidente	31/03	26/03 15:52	31/03 15:34	40	58	25	52	35	5	2	0,5483	0,003846	0	352.405
Tapioca pode matar	02/04	28/03 00:18	31/03 22:42	15	19	2	8	7	6	3	1,2982	0,05238	0	37.195
Pink Floyd anuncia turnê no Brasil	08/04	01/04 12:14	08/04 04:18	321	334	25	285	274	22	2	1,2603	0,000467	0	343.756
Gritaram “Prendam Lula!” na Paixão de Cristo	17/04	15/04 22:20	17/04 00:47	126	126	0	4	4	122	2	1,9666	0,007746	0	330.794
Esposa do juiz Sérgio Moro está ligada a esquema de corrupção	01/05	28/04 02:10	01/05 13:01	203	344	60	56	23	162	5	3,1205	0,006681	0,31597881	436.099
Delegado que investigava a morte de Teori Zavascki foi assassinado	03/06	31/05 12:35	03/06 17:19	6489	7759	314	1332	1084	5271	12	3,9870	0,000151	0,01240548	20.187.403
Homossexuais terão cota em concursos públicos	09/04	07/04 22:48	09/04 22:29	172	186	4	87	94	28	4	1,6377	0,003365	0,06151427	198.411
Andreas Von Richthofen encontrado na “Cracolândia”	03/06	30/05 22:05	03/06 18:55	597	632	29	172	155	294	8	4,1198	0,001287	0,00543192	2.929.462

Tabela 22 – Resumo das coletas associadas às notícias falsas selecionadas.

A seleção de dez notícias falsas teve o objetivo de gerar um conjunto representativo, que trouxesse um número de casos suficiente para possibilitar a identificação de potenciais padrões nas redes representativas das coletas das postagens associadas a essas notícias. Esse objetivo não tinha como exigência que as coletas tivessem uma quantidade de postagens e usuários parecida. Entretanto, como já comentado na metodologia, notícias com um número de postagens muito reduzido (inferior a dez postagens), poderiam gerar análises distorcidas, já que representariam uma notícia falsa de baixíssima repercussão. O conjunto de notícias falsas selecionadas, portanto, foi formado por coletas de variados tamanhos (entre 19 e 7.759 postagens), sendo a menor delas a referente à notícia falsa "Tapioca pode matar".

Mesmo com a limitação de intervalo de busca de até sete dias anteriores à coleta, imposto pela empresa responsável pela rede social Twitter, foi possível observar no conjunto de notícias falsas selecionadas o potencial de disseminação dessas informações inverídicas nas redes sociais, com algumas das coletas atingindo quantidades altas de postagens em relação à média, como visto na rede gerada a partir da coleta da notícia falsa "Delegado que investigava a morte de Teori Zavascki foi assassinado", com 7.759 postagens associadas.

Apesar da quantidade de postagens ser um razoável indicativo do nível de repercussão de uma determinada notícia divulgada em redes sociais como o Twitter, a audiência final atingida é uma medida mais exata para quantificar o número de indivíduos aos quais essa informação chegou. Uma única postagem de um usuário com 100.000 seguidores, por exemplo, pode superar a quantidade de audiência de uma notícia falsa que foi tema de 100 postagens de usuários com 500 seguidores cada. Por esse motivo esse indicador foi analisado em separado em cada uma das coletas. O gráfico da Figura 38 demonstra um resumo dos resultados dessa métrica nas coletas analisadas nessa dissertação.

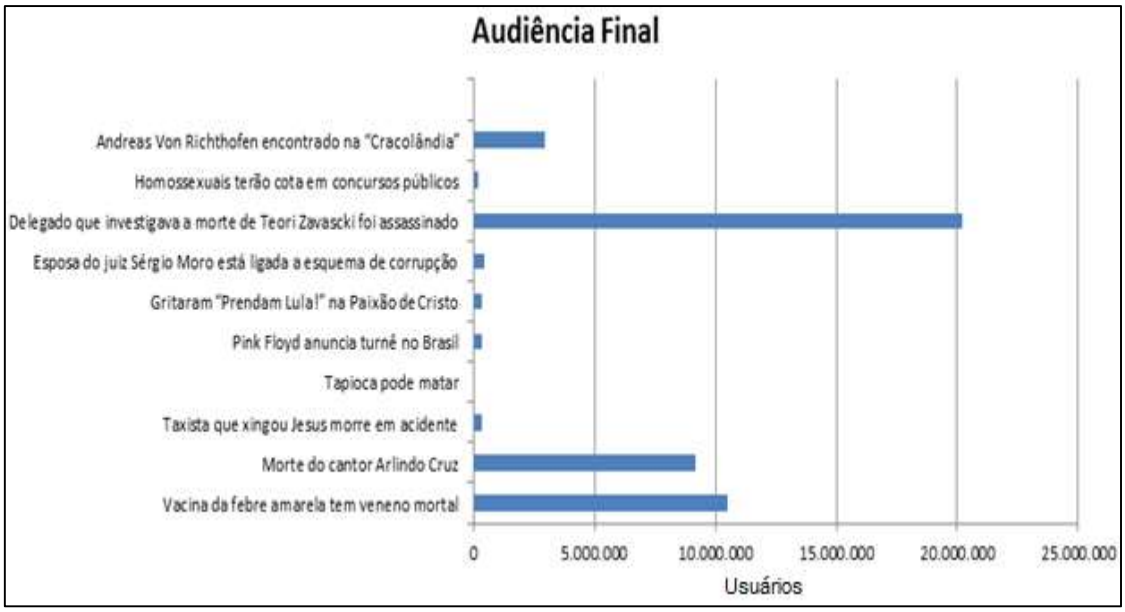

Figura 38 – Gráfico com a audiência final atingida pelas notícias falsas selecionadas (fonte: autor).

Avaliando os resultados trazidos pela soma da audiência final atingida, pôde-se observar em destaque, mais uma vez, a rede gerada a partir da coleta da notícia falsa "Delegado que investigava a morte de Teori Zavascki foi assassinado", que ultrapassou a marca de audiência de 20 milhões de usuários, seguida das notícias "Vacina da febre amarela tem veneno mortal", que ultrapassou a marca de 10 milhões, e "Morte do cantor Arlindo Cruz" que chegou próximo a esse valor.

O fato da notícia falsa com maior número de postagens ("Delegado que investigava a morte de Teori Zavascki foi assassinado", com 7.759) ter sido também a de maior audiência final poderia levar à incorreta dedução de que o número de postagens elevado implicaria em uma maior audiência. Entretanto, isso não foi observado em outras coletas, que atingiram também uma grande audiência com um número de postagens menor em relação à média observada. Essa constatação fica mais clara quando listamos as coletas ordenadas pela razão entre audiência final e o número de arestas (postagens) da rede associada à coleta. O resultado pode ser observado na Tabela 23.

Notícia	Arestas	Audiência Final	Audiência final / arestas
Morte do cantor Arlindo Cruz	54	9.161.284	169.653,41
Vacina da febre amarela tem veneno mortal	226	10.507.182	46.491,96
Taxista que xingou Jesus morre em acidente	58	352.405	6.075,95
Andreas Von Richthofen encontrado na "Cracolândia"	632	2.929.462	4.635,22

Gritaram “Prendam Lula!” na Paixão de Cristo	126	330.794	2.625,35
Delegado que investigava a morte de Teori Zavascki foi assassinado	7759	20.187.403	2.601,80
Tapioca pode matar	19	37.195	1.957,63
Esposa do juiz Sérgio Moro está ligada a esquema de corrupção	344	436.099	1.267,73
Homossexuais terão cota em concursos públicos	186	198.411	1.066,73
Pink Floyd anuncia turnê no Brasil	334	343.756	1.029,21

Tabela 23 – Coletas ordenadas pela razão entre audiência final e quantidade de arestas das redes representativas das coletas.

Percebe-se que, sob o enfoque da razão entre audiência final e arestas, a coleta relacionada à notícia “Morte do cantor Arlindo Cruz” foi a que teve uma maior “produtividade”, com uma média de aproximadamente 169.653 usuários do Twitter alcançados por cada uma das 54 postagens emitidas. Isso levou a uma audiência final alta de 9.161.284 usuários apesar da baixa quantidade de postagens, fazendo a comparação dessas duas métricas com o observado nas outras coletas. A quantidade de seguidores dos usuários emissores das postagens, portanto, pesou muito para a elevação da audiência final atingida.

A coleta “Vacina da febre amarela tem veneno mortal” também teve grande disseminação, com audiência final atingida de 10.507.182 usuários – posicionando-a em segundo lugar neste quesito – com apenas 226 postagens, também pela grande quantidade de seguidores dos usuários emissores das postagens.

É possível afirmar, pelos dados observados, que apesar do potencial de disseminação vindo da característica do modelo de rede de livre escala no qual as redes sociais estão enquadradas, que indicam um potencial de crescimento exponencial do número de indivíduos atingidos, as coletas da rede social Twitter precisam ser também analisadas em função do quantitativo de seguidores dos usuários emissores. A ausência dessa métrica, portanto, comprometeria fortemente as conclusões obtidas. Uma sugestão que poderia ser avaliada para futuros estudos seria a de inserir vértices representando os usuários seguidores na rede representativa da coleta. Essa ideia não foi posta em prática nessa dissertação pelo efeito colateral de dificultar consideravelmente a análise visual do desenho da rede através do algoritmo de FRUCHTERMAN e REINGOLD (1991). Ao invés disso, a graduação do tamanho do vértice foi a solução escolhida para que o quantitativo do número de seguidores fosse considerado no desenho.

Analisando em conjunto o desenho de todas as redes, em geral foi possível observar o formato esperado para redes de livre escala, com um vértice ou um pequeno

grupo de vértices destacado(s) ao centro, com alto grau de entrada em relação à maioria restante de vértices com zero ou baixo grau, posicionados na periferia pelo algoritmo de desenho, como mostrado na Figura 39. Esses vértices periféricos, em geral, foram também emissores de postagens que não tiveram referências de nenhum outro usuário, formando um laço visível na maioria dos grafos.

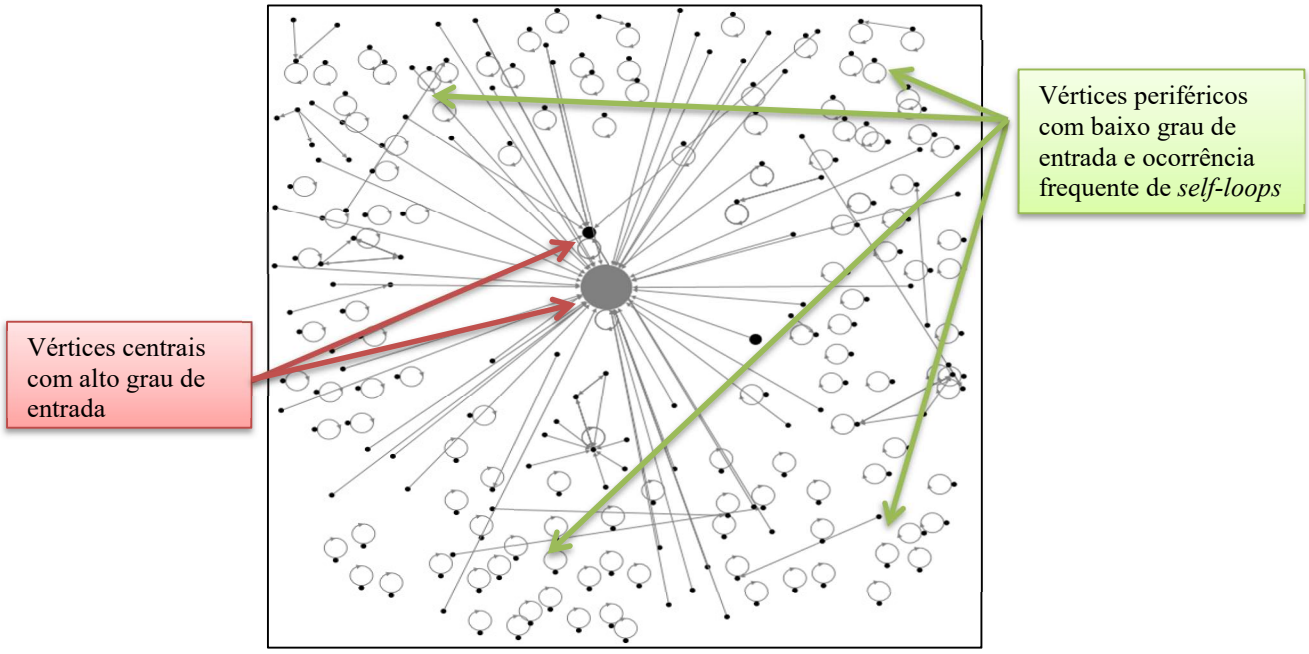

Figura 39 – Modelo de desenho observado nas redes representativas das notícias falsas analisadas, feito a partir da rede da coleta "Vacina febre amarela veneno mortal" (fonte: autor).

Esse padrão de redes de livre escala foi observado nas redes relacionadas às coletas das notícias "Vacina febre amarela veneno mortal", "Arlindo Cruz morre", "Taxista que xingou Jesus morre em acidente", "Pink Floyd anuncia turnê Brasil" e mesmo em pequenas redes como a referente à notícia "Tapioca pode matar". Entretanto, foi possível identificar algumas características que fugiram um pouco desse padrão observado e que merecem destaque, como na coleta referente à notícia "Gritaram Prendam Lula Paixão de Cristo" em que os vértices periféricos tiveram baixa quantidade de laços, indicando que apenas referenciaram as postagens dos usuários centrais, sem emitir postagens próprias. Outra coleta que apresentou diferença marcante foi a relacionada à notícia "Mulher Moro esquema corrupção APAE" que mostrou, além da mesma pouca frequência de laços nos vértices periféricos, também uma quantidade de vértices centrais maior em relação à média observada nas outras redes.

Outra informação importante, surgida a partir da observação dos desenhos das redes com o tamanho dos vértices dimensionado de acordo com a quantidade de

seguidores do usuário associado, foi que a quantidade desses seguidores não é determinante para o posicionamento do vértice ao centro da rede. Ou, dito de outra forma, a quantidade de seguidores dos usuários não esteve relacionada ao grau de entrada dos vértices associados a eles. Essa característica foi observada nas coletas relacionadas às notícias “Arlindo Cruz morre”, “Taxista que xingou Jesus morre em acidente”, “Pink Floyd anuncia turnê Brasil”, “Gritaram Prendam Lula Paixão de Cristo”, “Mulher Moro esquema corrupção APAE”, “Delegado inquérito Teori Zavascki assassinado” e “Homossexuais terão cota concursos”. Com base nessa característica observada nas redes estudadas, é possível afirmar que a quantidade de referências feitas por outras postagens – que poderia ser chamada de “popularidade” da postagem – não indica estar relacionada com o número de seguidores que um usuário possui. É frequente, portanto, que um usuário com quantidade pequena de seguidores possa emitir postagens que tenham muitas referências futuras e, consequentemente, atinja grande audiência.

Analisou-se também as métricas de número de componentes, componente máxima, diâmetro, densidade e coeficiente de agrupamento local. A expectativa era de que as redes analisadas apresentassem características esparsas, seguindo o modelo de redes de livre escala onde, em teoria, se enquadrariam. O conjunto das redes analisadas trouxe informações que foram em linha com essa expectativa, principalmente quando se consideraram as métricas de densidade e coeficiente de agrupamento local, que sempre se mostraram mais próximas de zero do que de seu limite teórico de um. As duas únicas redes onde as densidades calculadas ultrapassaram o valor de 0,01, indicando que a quantidade de arestas entre os vértices passou de 1% do limite teórico, foram as relacionadas às notícias “Tapioca pode matar” (0,05238) e “Morte do cantor Arlindo Cruz” (0,013425). Há que se considerar que essas duas redes foram exatamente as que tiveram as mais baixas quantidades de postagens e, portanto, as menores quantidades de vértices/arestas (15/19 e 53/54, respectivamente). Isso leva a crer que o pequeno tamanho dessas duas redes levou a um viés positivo nas métricas de densidade e coeficiente de agrupamento global.

O número de componentes nas redes não trouxe um padrão visível, sendo possível encontrar casos onde seu número era bastante elevado, como nas coletas relacionadas às notícias “Delegado que investigava a morte de Teori Zavascki foi assassinado” (1.084) e “Pink Floyd anuncia turnê no Brasil” (274), ou casos onde a

quantidade de componentes era relativamente baixa, como nas notícias “Gritaram “Prendam Lula!” na Paixão de Cristo” (4) e “Tapioca pode matar” (7).

A componente máxima foi outra métrica que também não apresentou padrão identificável, com situações em que agrupou quase todos os vértices da rede, como nas coletas “Delegado que investigava a morte de Teori Zavascki foi assassinado” (5.271 de 6.489 vértices), “Gritaram “Prendam Lula!” na Paixão de Cristo” (122 de 126 vértices) e “Esposa do juiz Sérgio Moro está ligada a esquema de corrupção” (162 de 203 vértices); mas também com casos onde essa quantidade de vértices conectados não chegava próximo da quantidade total de vértices da rede, como nas coletas “Vacina da febre amarela tem veneno mortal” (39 de 206 vértices), “Taxista que xingou Jesus morre em acidente” (5 de 40 vértices) e “Pink Floyd anuncia turnê no Brasil” (22 de 321 vértices).

Os diâmetros se apresentaram, em geral, condizentes com um formato mais centralizado das redes, estando em sua maioria próximos ao valor dois, indicando um máximo de duas arestas separando os vértices conectados da maioria das redes analisadas. Os únicos casos que não acompanharam esse padrão foram identificados nas coletas associadas às notícias “Delegado que investigava a morte de Teori Zavascki foi assassinado” (12) e “Andreas Von Richthofen encontrado na Cracolândia” (8). Essas duas coletas tiveram uma quantidade de postagens mais alta que a média (7.759 e 632, respectivamente), o que aparenta ter contribuido para esse valor mais alto de diâmetro observado.

O histograma de distribuição de graus foi mais uma ferramenta que mostrou a aderência das redes coletadas ao modelo de redes de livre escala. Em geral, todas as redes apresentaram uma distribuição de probabilidades de localização de vértices que seguiu próximo de um padrão exponencialmente decrescente, conforme o aumento do grau, como visto na Figura 7 da Seção 2.3 do Capítulo 2. Destaca-se que a análise dos graus de entrada e saída foi feita em conjunto, demonstrando para ambos o mesmo padrão de distribuição. O único ponto de frequente interrupção desse padrão nos gráficos observados foi nos graus de entrada e saída iguais a zero, que em geral mostraram probabilidades de localização menores que os de grau um. Entretanto, a adição do grau zero ao eixo X não é padronizado nos estudos que utilizam essa ferramenta, sendo mais comum encontrar-se o início do eixo X em um. Ainda assim, entendeu-se ser interessante a exibição das probabilidades de grau zero para análise, sem que sua inferioridade observada com relação ao grau um seja considerada neste

estudo como um fator que determine a não aderência ao padrão de decréscimo exponencial na distribuição de graus das redes de livre escala.

Os gráficos de evolução da audiência em função do tempo, nos quais o tempo foi agrupado em períodos de um dia, evidenciou saltos abruptos das audiências alcançadas pelas notícias pesquisadas, formando padrões de crescimento "em escada", variando apenas no número de degraus em que esse crescimento ocorreu. O crescimento próximo ao padrão de degrau único foi observado nas notícias "Vacina da febre amarela tem veneno mortal", "Morte do cantor Arlindo Cruz", "Taxista que xingou Jesus morre em acidente", "Tapioca pode matar" e "Delegado que investigava a morte de Teori Zavascki foi assassinado". As outras notícias apresentaram múltiplos degraus de crescimento de audiência identificáveis: "Pink Floyd anuncia turnê no Brasil", "Gritaram "Prendam Lula!" na Paixão de Cristo", "Esposa do juiz Sérgio Moro está ligada a esquema de corrupção", "Homossexuais terão cota em concursos públicos" e "Andreas Von Richthofen encontrado na "Cracolândia". Pelos gráficos analisados, um elemento que influenciou fortemente o formato dos gráficos de evolução de audiência foi a decisão de agrupar as postagens por data, sem considerar a hora da postagem, o que pode ter contribuído para concentrar muitas postagens no mesmo dia e levado aos degraus observados. Uma sugestão para futuros estudos que busquem analisar mais a fundo a evolução da audiência é procurar novas formas de agrupamento (a cada hora ou minuto, por exemplo) que procurem identificar mais detalhadamente o momento em que o crescimento mais intenso da audiência se deu.

As tabelas com os cinco usuários mais influentes, de acordo com seu índice PageRank, teve o objetivo de identificar que usuários tiveram maior importância na disseminação da notícia dentro da rede. A importância, neste caso, quantificada por uma métrica que procura destacar os vértices por sua influência global na rede. Os dados obtidos demonstraram a presença de vários tipos de usuários entre os mais influentes, desde usuários representantes de grandes redes de comunicação como o "g1" (do Grupo Globo de Comunicação) – líder na coleta relacionada à notícia "Vacina da febre amarela tem veneno mortal" – e "bandsportstv" (do grupo Bandeirantes de Comunicação), segundo colocado na coleta da notícia "Morte do cantor Arlindo Cruz"; até usuários representantes de portais que procuram desmistificar as notícias falsas, como o "efarsas" – segundo colocado na coleta relacionada à notícia "Vacina da febre amarela tem veneno mortal" – e "boatosorg", visto na segunda posição das coletas "Tapioca pode matar" e "Gritaram Prendam Lula Paixão de Cristo". Além desses tipos, também foram

identificados usuários que representam grupos de pensamento político, como "comunacritico" – primeiro colocado na rede relacionada à notícia "Delegado que investigava a morte de Teori Zavascki foi assassinado" – e "pelademocracia1", segundo colocado na notícia "Esposa do juiz Sérgio Moro está ligada a esquema de corrupção na APAE". Entretanto, também com frequência, foram observados nas primeiras posições dessas tabelas usuários representantes de pessoas físicas, algumas vezes até com baixa quantidade de seguidores, mostrando que o poder de disseminação dos usuários é diversificado e independente de sua situação de popularidade fora do Twitter (característica das grandes redes de comunicação, por exemplo) ou dentro do Twitter (número de seguidores). Essas informações evidenciam o desafio que seria implementar um bloqueio da disseminação de informações falsas, dada a dificuldade de identificar previamente os possíveis usuários-chave que seriam determinantes nessa disseminação.

5. Conclusões

Esse trabalho explorou o tema da disseminação de notícias falsas, tema atualmente presente em todas as sociedades desenvolvidas, pela importância destacada na agilidade e facilidade trazidas por mecanismos de comunicação modernos. Dada a imensa abrangência do termo "informação falsa", que agrega desde dados, estatísticas, notícias, propagandas e outros conteúdos falsos, neste estudo focou-se especificamente o subconjunto formado pelas notícias falsas divulgadas nas redes sociais de microblogs, que foram caracterizadas no contexto desse estudo como notícias recentes, com conteúdo comprovadamente desconectado da realidade ou com incorreções graves que poderiam levar o leitor a interpretações errôneas sobre os fatos e que tenham sido posteriormente esclarecidas através de checagem em fontes externas confiáveis.

A partir da definição da fonte de pesquisa – redes sociais de microblogs –, a seleção do Twitter se deu basicamente por dois motivos. Em primeiro lugar, por ser uma das redes sociais de maior prestígio ainda hoje, mais de 10 anos após sua criação. Além disso, o Twitter é uma das raras redes sociais que fazem restrições mínimas quanto ao conteúdo dos dados coletados, inclusive sobre dados de usuários. Para a operacionalização da coleta das postagens, foi escolhida e utilizada a ferramenta NodeXL

As notícias falsas foram selecionadas principalmente através dos sites E-Farsas (www.e-farsas.com) e o Boatos.org (www.boatos.org), reconhecidos na internet por fazerem esclarecimentos de notícias falsas divulgadas na rede. Esses dois sites puderam ter sua importância reconhecida neste estudo, já que efetivamente foram boas fontes de esclarecimentos sobre notícias falsas circulantes nas redes sociais, com atualizações frequentes e informações que se mostraram corretas através de consultas a outras fontes de informação de confiabilidade reconhecida. A partir da seleção das notícias, foram feitas buscas com a ferramenta NodeXL com uma sequência de termos textuais que

estavam associados ao conteúdo da notícia. Ocorreu em alguns casos que o número de postagens retornados de algumas coletas não chegaram a dez, mesmo depois de ajustes nos termos para buscar uma melhor associação com o conteúdo pretendido. Nesses casos a notícia falsa foi removida da lista. O conjunto final de notícias falsas que atenderam os requisitos mínimos de tamanho e, portanto, puderam ser analisadas detalhadamente, estão resumidas na Tabela 24:

Notícia	Data coleta	Vértices	Arestas	Audiência Final
Vacina da febre amarela tem veneno mortal	26/03	206	226	10.507.182
Morte do cantor Arlindo Cruz	23/03	53	54	9.161.284
Taxista que xingou Jesus morre em acidente	31/03	40	58	352.405
Tapioca pode matar	02/04	15	19	37.195
Pink Floyd anuncia turnê no Brasil	08/04	321	334	343.756
Gritaram “Prendam Lula!” na Paixão de Cristo	17/04	126	126	330.794
Esposa do juiz Sérgio Moro está ligada a esquema de corrupção	01/05	203	344	436.099
Delegado que investigava a morte de Teori Zavascki foi assassinado	03/06	6489	7759	20.187.403
Homossexuais terão cota em concursos públicos	09/04	172	186	198.411
Andreas Von Richthofen encontrado na “Cracolândia”	03/06	597	632	2.929.462

Tabela 24 – Resumo das notícias falsas selecionadas para estudo.

A seleção das notícias falsas foi considerada satisfatória em atender os requisitos previamente definidos, espelhando situações de disseminação que se apresentaram com diferenças marcantes em métricas como o número de usuários envolvidos (vértices), número de postagens localizadas (arestas) e audiência final atingida. A audiência final foi uma métrica tratada em destaque por entender-se que retratava mais adequadamente o potencial número de leitores da notícia falsa estudada, somando todos os seguidores dos usuários do Twitter que emitiram alguma postagem relacionada à essa notícia investigada. O conjunto final de notícias falsas analisadas montou redes representativas de notícias como “Tapioca pode matar” e “Taxista que xingou Jesus morre em acidente”, que apresentaram redes relativamente pequenas em comparação com o restante das notícias coletadas; e também trouxe notícias que, por sua maior disseminação, criaram grandes redes, como a notícia “Delegado que investigava a morte de Teori Zavascki foi assassinado” que atingiu a marca de 7.759 arestas e 6.489 vértices, e a notícia “Andreas Von Richthofen encontrado na “Cracolândia” ”, que atingiu 632 arestas e 597 vértices.

As datas de emissão das postagens identificadas nas coletas realizadas demonstraram que a limitação do universo de pesquisa em até sete dias anteriores à data

de realização da coleta, restrição imposta pela empresa responsável pelo Twitter, foi um delimitador que precisou ser considerado na análise do tamanho das redes. Algumas coletas, como as relacionadas às notícias “Vacina da febre amarela tem veneno mortal” e “Pink Floyd anuncia turnê no Brasil” trouxeram postagens mais antigas iniciando exatamente no sétimo dia anterior à data da busca e postagens mais recentes na própria data da busca, sendo essas coletas mais corretamente consideradas como um recorte cronológico do que seria uma hipotética rede com todas as postagens referentes à notícia estudada, emitidas em todas as datas. Mesmo notícias que não tiveram postagens mais antigas feitas no sétimo dia anterior à busca, ainda assim apresentaram uma interrupção abrupta das postagens na própria data da realização da coleta. Em resumo, mesmo quando a limitação de sete dias não se aplicou, havia potencial de outras postagens futuras relacionadas à notícia falsa pesquisada serem localizadas em futuras coletas.

Apesar dessas observações relacionadas ao intervalo cronológico das postagens encontradas, considerou-se que os dados obtidos foram uma representação suficiente dos universos das postagens relacionadas às notícias falsas estudadas, já que todas as notícias selecionadas tiveram como um dos requisitos que o ápice de sua divulgação tivesse acontecido poucos dias antes da realização das coletas. Contudo, a experiência obtida possibilita sugerir para estudos posteriores que se busquem outras estratégias para tentar preencher um universo mais completo das postagens encontradas.

A seguir, as questões estabelecidas como objetivo desse estudo e as conclusões que foram alcançadas com as análises desenvolvidas:

- No recorte temporal observado, quais as origens da divulgação deste conjunto de notícias falsas postadas no Twitter?
    - Em geral, não foi possível identificar um único usuário ou tipo único de usuário que seria o maior responsável pela disseminação do conjunto das notícias falsas analisadas. Por qualquer dos dois enfoques analisados – usuário emissor da primeira postagem ou usuário com maior influência na rede definida por métricas baseadas nas referências de outras postagens – o que se observou foi uma alternância de usuários representantes de grandes grupos de comunicação e usuários representantes de pessoas físicas ou grupos de pessoas. Isso indicou que, para esse grupo de notícias

falsas analisadas, não houve um usuário específico que merecesse destaque como principal responsável pela disseminação dessas notícias. Entretanto, as métricas de centralidade globais utilizadas neste estudo se mostraram adequadas para identificar o usuário mais influente na disseminação de uma única notícia.

- Como é o processo de disseminação deste conjunto de notícias falsas postadas no Twitter?
    - As redes representativas das notícias falsas analisadas apresentaram algumas características similares, mas com diferenças marcantes entre si. Seus tamanhos, por exemplo, variaram da casa de poucas dezenas a milhares de vértices/arestas. Em geral se apresentaram muitas postagens avulsas (laços), sem referências de outras postagens, ainda que se tenha encontrado um número pequeno de coletas em que isso não foi observado. Foi comum que um número pequeno de postagens recebesse um grande número de referências, sendo neste caso de grande importância na disseminação da notícia falsa analisada.
    - A análise da audiência atingida, formada não só pelos usuários emissores de postagens, mas também seus seguidores correspondentes, permitiu demonstrar e quantificar na prática os riscos trazidos pela agilidade das redes sociais como o Twitter na disseminação de notícias. Foi comum a ocorrência de saltos de milhões de usuários potenciais leitores da notícia dentro de um mesmo dia. Esses saltos foram sempre muito dependentes do número de seguidores do usuário emissor da postagem. Entretanto, o número de seguidores de um determinado usuário não demonstrou ter qualquer correlação com sua importância quantificada pelas métricas de centralidade utilizadas neste estudo. Isso implica que, pelo menos para esse grupo de notícias falsas, o usuário mais influente na disseminação não precisou obrigatoriamente ser um usuário com grande número de seguidores.

- É possível caracterizar as redes deste conjunto de notícias falsas?
    - Em geral, as notícias falsas formaram redes representativas seguindo um formato centralizado, com baixo diâmetro, seguindo o modelo de rede de livre escala esperado para as redes sociais, onde poucos vértices possuem um grau maior e o restante possui graus baixos. Esse modelo de redes indica que um número pequeno de vértices possui uma importância destacada na relação entre todos os vértices da rede. Indica também uma boa tolerância a falhas, sendo muito difícil que um usuário aleatoriamente escolhido para bloqueio impedisse a disseminação da notícia falsa.

A experiência adquirida neste estudo permite lançar algumas sugestões para trabalhos futuros. Uma delas é a realização de uma análise mais detalhada da evolução da audiência das notícias falsas em função do tempo, com um recorte de tempo menor (minuto a minuto, por exemplo) que pudesse trazer mais esclarecimentos sobre como se comporta essa evolução. Estudos posteriores poderiam também buscar soluções de como impedir que o fluxo de notícias falsas ganhe maior divulgação, apesar do desafio técnico que isso representaria. É possível também desenvolver as questões exploradas neste estudo em algum outro tema específico, como na divulgação de propaganda política ou na divulgação de eventos, em que postagens poderiam ser coletadas usando estratégia similar à utilizada neste estudo.

REFERÊNCIAS

BAVELAS, A. (1950). “Communication patterns in task-oriented groups”. *J. Acoust. Soc. Am,* 22(6):725–730.

BRANDES, U. (2001). “A faster algorithm for betweenness centrality”. *Journal of Mathematical Sociology*. 25: 163–177.

BRIN, S.; PAGE, L. (1998). “The anatomy of a large-scale hypertextual Web search engine”. *Computer Networks and ISDN Systems*. 30: 107–117

CALLAWAY, D. S., NEWMAN, M. E. J., STROGATZ, S. H. & WATTS, D. J. (2000). “Network robustness and fragility: percolation on random graphs”. *Phys. Rev. Lett.* 85, 5468–5471.

CERVI, E.U., MASSUCHIN, M.G. (2010). “O uso do twitter nas eleições de 2010: O microblog nas campanhas dos principais candidatos ao governo do Paraná”. *Contemporânea, comunicação e cultura.* 2011.

CHO, J., ROY, S. (2004). *Impact of search engines on page popularity*. 13th international conference on World Wide Web (pp. 20-29). ACM.

COHEN, R., HAVLIN, S. (2010). *Complex Networks: Structure, Robustness and Function*. Cambridge University Press.

EL PAÍS (2016). “Colômbia diz ‘não’ ao acordo de paz com as FARC”. Disponível em http://brasil.elpais.com/brasil/2016/10/02/internacional/1475420001_242063.html acessado em 10/05/2017.

ERDŐS, P. RÉNYI, A. (1959). *On Random Graphs* I. Publ. Math. Debrecen 6, p. 290–297.

EXAME (2016). “Promotoria estuda se campanha do “Brexit” foi enganosa”. Em http://exame.abril.com.br/mundo/promotoria-estuda-se-campanha-do-brexit-foi-enganosa. Acessado em 19/3/2017.

FANTINI, J.A. (2016). “Pós-verdade ou o triunfo da religião?” Em http://revistas.pucsp.br/index.php/leituraflutuante/article/view/31767/22105. Acessado em 18/3/2017.

FAUS, J. (2016). “Trump diz, sem provas, que “milhões de pessoas votaram ilegalmente” em Clinton”.
Em http://brasil.elpais.com/brasil/2016/11/27/internacional/1480278816_162209.html. El País.

FERREIRA, B.C.L. (2013). *A influência das mídias sociais na divulgação de um evento: um estudo de caso sobre o evento Experiencie*. TCC. UFRGS.

FINN S., METAXAS P. T., MUSTAFARAJ E., O'KEEFE M., TANG L., TANG S., ZENG L. (2014). *TRAILS: A System for Monitoring the Propagation of Rumors On Twitter.* Computer Science Department Wellesley College, Wellesley, MA.

FREEMAN, L. C. (1979). *Centrality in social networks: conceptual clarification*. Social Networks.

FRUCHTERMAN, T. M. J.; REINGOLD, E. M. (1991). "Graph Drawing by Force-Directed Placement". *Software – Practice & Experience.* Wiley, 21 (11): 1129–1164.

GLOBO, O. (2017). "Trump divulga informações falsas para defender bloqueio a refugiados". Em http://oglobo.globo.com/mundo/trump-divulga-informacoes-falsas-para-defender-bloqueio-refugiados-20849852. Acessado em 20/3/2017.

GOOGLE. (2011). "Por que usar o Google? Explicações sobre o método PageRank". Em https://www.google.com/intl/pt-BR/why_use.html. Acessado em 15/06/2017.

GOOGLE. (2016). "Labeling fact-check articles in Google News". Em http://blog.google/topics/journalism-news/labeling-fact-check-articles-google-news/

GUANHUA, Y., GUANLING C. EIDENBENZ S., LI N. (2011). *Malware Propagation in Online Social Networks: Nature, Dynamics, and Defense Implications*. Information Sciences (CCS-3). Los Alamos National Laboratory.

HANSEN, D.; SHNEIDERMAN, B.; SMITH, M.A. (2010). *Analyzing social media networks with NodeXL: Insights from a connected world*. umd.edu.

KEYES, R. (2004). *The Post-truth Era: Dishonesty and Deception in Contemporary Life.* St. Martin´s Press.

KIM, J., LEE, E., CHOI, J., BAE, Y., KO, M., KIM, P. (2014). "Monitoring social relationship among Twitter users by using NodeXL". *dl.acm.org*.

KWAK, H. et al. (2010). "What is Twitter, a social network or a news media?" *WWW '10 Proceedings of the 19th international conference on World wide web*, p. 591-600. North Carolina, EUA.

LAMBERSON, P.J. (2015). "Scale-free Network". Em http://social-dynamics.org/scale-free-network. Acessado em 22/10/2017.

LORO, M. (2017). "Trump Insiste Que "Milhões" De Imigrantes Ilegais Deram Vitória a Clinton no Voto Popular".
Em http://www.publico.pt/2017/01/24/mundo/noticia/trump-insiste-que-milhoes-de-imigrantes-ilegais-deram-vitoria-no-voto-popular-a-clinton-1759568. Acessado em 20/3/2017.

LUCE R.D., PERRY, A.D. (1949). “A Method of Matrix Analysis of Group Structure”. *Psychometrika* 14 (1): 95–116.

MARCHI, R. (2012). “With Facebook, Blogs, and Fake News, Teens Reject Journalistic “Objectivity” ”. *Journal of Communication Inquiry.* Vol 36, exemplar 3, 2012.

MICROSOFT. (2016). Microsoft Excel 2016. Em https://products.office.com/pt-br/excel. Acessado em 01/05/2017.

NEWMAN, M.E.J. (2003). *The structure and function of complex networks.* Department of Physics, University of Michigan.

NEWMAN, M.E.J. (2010). *Networks: An Introduction*. Oxford, UK: Oxford University Press.

OXFORD. (2016). “Word of The Year 2016”.
Em http://en.oxforddictionaries.com/word-of-the-year/word-of-the-year-2016 - Acessado em 18/03/2017 - Oxford University Press.

PRIMO, A.F.T. (2011). “Transformações no jornalismo em rede: sobre pessoas comuns, jornalistas e organizações; blogs, Twitter, Facebook e Flipboard”. *inTexto,* no. 25. Universidade Federal do Rio Grande do Sul.

ROCHAT, Y. (2009). *Closeness Centrality Extended to Unconnected Graphs: The Harmonic Centrality Index*. Institute of Applied Mathematics, University of Lausanne, Switzerland. ASNA 2009.

SHERMER, M. (2011). *The Believing Brain: From Ghosts and Gods to Politics and Conspiracies - How We Construct Beliefs and Reinforce Them as Truths*. St. Martin's Griffin.

SITUNGKIR, H. (2011). *Spread of hoax in Social Media*. Dept. Computational Sociology. Bandung Fe Institute.

SMITH, M.A, SHNEIDERMAN, B., MILIC-FRAYLING, N. (2009) *Analyzing (social media) networks with NodeXL*. dl.acm.org.

SOCIALMEDIA. (2014). NodeXL: network overview, discovery and exploration for Excel. Em http://nodexl.codeplex.com/. Acessado em 01/05/2017.

STROGATZ, S. H. (2001). “Exploring Complex Networks”. *Nature,* vol 410. P 268-276.

TWITTER INC. (2017). Busca avançada. Em https://twitter.com/search-advanced?lang=pt. Acessado em 01/05/2017.

WASSERMAN, S., FAUST, K. (1994). *Social Network Analysis: Methods and Applications.* Cambridge University Press.

WATTS D. J., STROGATZ S. H. (1998). "Collective dynamics of 'small-world' networks". *Nature.* London. 393, p. 440.

ZAGO, G. S. (2012). "Trolls e Jornalismo no Twitter". *Estudos em Jornalismo e Mídia* - Vol. 9 Nº 1 – Janeiro a Junho de 2012.

ZHIWEI, J., JUAN, C., YU-GANG, J., YONGDONG, Z. (2014). *News Credibility Evaluation on Microblog with a Hierarchical Propagation Model*. Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing, China.