UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM DE SISTEMAS COMPLEXOS

ROGÉRIO OLÍMPIO DA SILVA

Estudo da propagação de postagens de notícias no Twitter

SÃO PAULO – SP

2018

ROGÉRIO OLÍMPIO DA SILVA

Estudo da propagação de postagens de notícias no Twitter

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Modelagem de Sistemas Complexos.

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em 29 de setembro de 2017. A versão original encontra-se em acervo reservado na Biblioteca da EACH/USP e na Biblioteca Digital de Teses e Dissertações da USP (BDTD), de acordo com a Resolução CoPGr 6018, de 13 de outubro de 2011.

Área de Concentração:

Sistemas Complexos

Orientador:

Prof. Dr. Fernando Fagundes Ferreira

SÃO PAULO – SP

2018

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

CATALOGAÇÃO-NA-PUBLICAÇÃO

(Universidade de São Paulo. Escola de Artes, Ciências e Humanidades. Biblioteca)

Silva, Rogério Olímpio da
Estudo da propagação de postagens de notícias no Twitter / Rogério
Olímpio da Silva ; orientador, Fernando Fagundes Ferreira. – 2018.
84 f. : il.

Dissertação (Mestrado em Ciências) - Programa de Pós-
Graduação em Modelagem de Sistemas Complexos, Escola de
Artes, Ciências e Humanidades, Universidade de São Paulo, em
2017.

Versão corrigida.

1. Redes sociais. 2. Redes complexas. 3. Notícia -
Disseminação - Simulação computacional. 4. Twitter. I. Ferreira,
Fernando Fagundes, orient. II. Título

CDD 22.ed.— 303.4833

ROGÉRIO OLÍMPIO DA SILVA

Estudo da propagação de postagens de notícias no Twitter

Dissertação apresentada à Escola de Artes,
Ciências e Humanidades da Universidade de
São Paulo para obtenção do título de Mestre em
Ciências do Programa de Pós-Graduação em
Modelagem de Sistemas Complexos.

Área de Concentração:

Sistemas Complexos

Aprovado em: 15/09/2017.

Banca Examinadora

Prof. Dr. Ana Amélia Benedito Silva

Instituição: EACH - USP

Prof. Dr. Josmar Andrade

Instituição: EACH - USP

Prof. Dr. Nelson Lemer Barth

Instituição: FGV-SP

Dedico esse trabalho aos meus pais, Ariovaldo e Maria, os quais me ensinaram a importância de estudar sempre com seriedade, trabalhar com empenho e acreditar que sempre dá para ir além. E assim me impulsionaram a ir além até do que eu achava ser possível. Grato a Deus por ser filho de vocês!

AGRADECIMENTOS

Em primeiro lugar quero agradecer a Deus porque Dele, por Ele e para Ele são todas as coisas. A Ele seja a glória para sempre. Amém!

Meus profundos agradecimentos à minha família que imergiu neste projeto comigo nos últimos anos e suportou a adaptação para que eu pudesse focar. Minha esposa Estela por sua infundável garra e encorajamentos, à minha filha Nicole por sua participação com releituras e discussões dos temas, e ao meu filho Artur que lidou com bom humor com um pai focado no projeto.

Aos parentes, amigos e à comunidade Geração Nova Filadélfia que conviveram com minha dedicação parcial nos últimos anos devido ao foco neste projeto.

Por fim, ao meu orientador professor doutor Fernando Fagundes Ferreira, que acreditou em mim antes mesmo do projeto ser viável, acreditou que um profissional de mercado pudesse trabalhar de forma acadêmica. Obrigado!

“Ele fez tudo apropriado a seu tempo. Também pôs no coração do homem o anseio pela eternidade; mesmo assim este não consegue compreender inteiramente o que Deus fez.”

Salomão

RESUMO

SILVA, Rogerio Olimpio Da. **Estudo da propagação de postagens de notícias no Twitter**. 2017. 84 p. Dissertação (Mestrado em Ciências pelo Programa de Pós-graduação em Modelagem de Sistemas Complexos) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2018. Versão Corrigida.

O presente trabalho tem como objetivo apresentar o resultado da análise e identificação de padrões e comportamentos de postagens na rede social Twitter monitorando um termo específico em um período definido de tempo. Após a extração e tratamento dos dados obtidos do Twitter, explicitando dentre as postagens quais são oriundas de usuários geradores de notícias, foram analisadas informações de topologia da rede de usuários do experimento, inclusive com o detalhamento das informações de propagação e duração das postagens: dados e comportamentos os quais buscamos reproduzir utilizando simulação por agentes.

Palavras-chave: Complexidade. Redes sociais. Redes complexas. Propagação de notícias. Modelo de rumor. Difusão da inovação. Twitter. Simulação.

ABSTRACT

SILVA, Rogerio Olimpio Da. **Study of the propagation of news postings on Twitter**. 2017. 84 p. Master Thesis on Complex Systems Modeling – School of Arts, Science and Humanities, University of São Paulo, 2018. Corrected Version.

The present work aims to present the result of the analysis and identification of patterns and behaviors of Twitter posts monitoring a specific term in a period. After the extraction and treatment of the data obtained from Twitter, emphasizing among the posts which are generated from news agency users, were analyzed the topology information about the network of users of the experiment, including details of propagation information and duration of the posts: data and behaviors that we seek to reproduce using agent simulation.

Keywords: Complexity. Social networks. Complex networks. News spread. Rumor model. Diffusion of innovation. Twitter. Simulation.

LISTA DE FIGURAS

Figura 1 - Distribuição de graus dos usuários do Twitter.	25
Figura 2 -Número de destinatários de retuite por número de seguidores.	25
Figura 3 - Proporção e usuários participantes na árvore de retuites.	26
Figura 4 - Taxa de mensagens e Tamanho de mensagens.	27
Figura 5 - Frequência do percentual de retuites por seguidores.	36
Figura 6 - Frequência da quantidade de retuites por postagem.	37
Figura 7 - Frequência da quantidade de postagens por dia.	37
Figura 8 - Frequência da distribuição de seguidores.	38
Figura 9 - Comparação de postagens com mesmo conteúdo com a duração da postagem pioneira em horas.	39
Figura 10 – Distribuição das quantidades de postagens com o mesmo conteúdo no período de 16 dias. Classificados por notícias (as postagens originadas de agências de notícias) que estão representadas pela cor vermelha e postagens comuns (postagens originadas por usuários comuns) na cor azul.	40
Figura 11 - Quantidade de postagens com mesmo conteúdo acumuladas e quantidade de retuites paralelos acumulados.	41
Figura 12 - Comparação do contador de usuários postando em relação à duração média das postagens em horas distribuídas na data da postagem pioneira.	42
Figura 13 - Relação das postagens e dos usuários que postaram as 20 postagens que mais duraram na rede social dentro do experimento.	43
Figura 14 - Quantidade de postagens pela quantidade de horas de duração das primeiras postagens, tanto para postagens originadas de agências de notícias (em vermelho) quanto para as de usuários comuns (em azul).	44
Figura 15- Sequência de postagem de notícias com mesmo conteúdo, onde as setas representam a continuidade da notícia no tempo.	46
Figura 16 - Sequência de postagens comuns (não são notícias) com mesmo conteúdo, onde as setas representam a continuidade da postagem no tempo.	46
Figura 17 - Processo de simulação de modelagem por agentes.	50
Figura 18 - Distribuição de graus dos dados da rede empírica.	52
Figura 19 - Distribuição de graus dos dados da Rede Simulada.	53
Figura 20 - Comparação entre dados empíricos e simulados considerando quantidade de postagens e quantidades de usuários.	54
Figura 21 - Forma como a probabilidade P afeta a existência ou não da rede mundo pequeno.	65
Figura 22 - Podemos verificar a diferença de topologia de uma rede Livre de Escala e uma rede aleatória.	66
Figura 23 - Processo de preparo dos dados empíricos.	71

LISTA DE TABELAS

Tabela 1 - Estatísticas de Mensuração do Twitter.....	18
Tabela 2 - Tabela de Dados 'Posts'	32
Tabela 3 - Tabela de métricas da rede de usuários chamados (U1) e chamadores (U2). Extraídos das ferramentas Python e R após importação da tabela como <i>dataset</i>	34
Tabela 4 - Informações estatísticas da rede de usuários.....	35
Tabela 5 – Descrição das variáveis.	36
Tabela 6 - Tabela com detalhamento dos valores das Figuras 15 e 16.	47
Tabela 7 - Tabela de Dados 'Usuario'.....	73
Tabela 8 - Tabela de Dados 'Analitico'.....	73
Tabela 9 - Tabela de Dados 'Chamados'.....	74
Tabela 10 - Tabela de Dados 'Hashtags'.....	74

SUMÁRIO

1	INTRODUÇÃO	13
1.1	DESCRIÇÃO DO PROBLEMA	15
1.2	OBJETIVOS	16
1.2.1	Objetivo Principal	16
1.2.2	Objetivos Secundários	16
1.3	ESTRUTURA DA DISSERTAÇÃO	16
2	REVISÃO DA LITERATURA	17
2.1	PROPAGAÇÃO DE INFORMAÇÕES NO TWITTER	20
2.2	TEORIA DO RUMOR	28
3	ANÁLISE EMPÍRICA DA REDE DE USUÁRIOS DO TWITTER: UM ESTUDO DE CASO	32
3.1	DESCRIÇÃO DA REDE	32
3.2	ANÁLISE DA REDE DE USUÁRIOS E POSTAGENS	35
4	MODELO E SIMULAÇÃO	49
4.1	MODELO	51
4.2	RESULTADOS DA SIMULAÇÃO	52
4.3	DISCUSSÃO	54
5	CONSIDERAÇÕES FINAIS	56
	REFERÊNCIAS	59
	APÊNDICE A – REDES COMPLEXAS	64
	APÊNDICE B – GLOSSÁRIO DE TERMOS	68
	APÊNDICE C – MÉTODO DE PESQUISA	71
	APÊNDICE D - TABELAS AUXILIARES	73
	APÊNDICE E – CRIAÇÃO DAS TABELAS DE DADOS	75
	APÊNDICE F – SIMULAÇÃO POR MODELAGEM BASEADA EM AGENTES	78

1 INTRODUÇÃO

A interação entre pessoas está se tornando cada vez mais dependente de tecnologia. Diversos tipos de mensagens são compartilhados ao redor do mundo por usuários que regularmente usam as plataformas de redes sociais. Os jornais são mais vistos de forma eletrônica do que impressa, e a mídia utiliza múltiplos canais digitais incluindo redes sociais para disseminar informações, conceitos, notícias e entretenimento, apenas para citar algumas (AIRSTRIP, 2017).

Uma rede social é composta por um número finito de atores e as relações existentes entre eles, sendo que entre dois atores podem existir diferentes tipos e qualidades de ligações: recíprocas ou não (KWAK, HAEWOON, et al., 2010). Um estudo do comScore (COMSCORE, 2017) diz que os brasileiros gastaram em média 29,7 horas por mês nas redes sociais em 2014, 7 horas a mais do que a média mundial. A cada dia é mais raro encontrar um jovem que não tenha contas em diversas redes sociais, sendo sua principal forma de se expressar e interagir.

As plataformas de redes sociais surgiram no final do século XX, e no Brasil a primeira plataforma a ter mais destaque foi o MySpace (2003), que permitia localizar e trocar mensagens com pessoas com quem não se tinha acesso direto bem como artistas e celebridades. Em paralelo se iniciava o uso de uma plataforma de redes sociais para empresários e profissionais chamada LinkedIn (2003), que só teria seu alcance ampliado posteriormente. Em seguida figurou o Orkut (2004) – grande boom em abrangência nas redes sociais – este oferecia a possibilidade de reencontrar pessoas há muito não vistas e compartilhar informações em grupos de interesse. Seguindo os mesmos objetivos de conectividade chega o Facebook (2004), com seu padrão simplificado de se compartilhar de forma direcionada e manter conversações incluindo contatos do MSN e do Skype. A próxima geração de redes sociais viria algum tempo depois, com outra plataforma sendo utilizada de forma crescente, o Twitter (2006). Criando um ambiente menos invasivo para que as celebridades pudessem expor seus pensamentos sem necessariamente serem obrigadas a visualizar o retorno. E, por fim, o Whatsapp (2012) surgiu para possibilitar o contato entre pessoas que tivessem o aplicativo, desde que conhecessem o número de celular do usuário com quem quisessem conversar, o que restringia mais o acesso aos perfis pessoais do que as outras redes sociais (TECMUNDO-1, 2017), (TECMUNDO-2, 2017).

Com a disseminação do uso de redes sociais digitais, as pessoas acostumam-se a se relacionar desta forma. Até mesmo empresas buscam tirar proveito deste momento, criando

variações e aplicativos similares a redes sociais para razões profissionais como, por exemplo, a facilitação de comunicação entre colaboradores em redes sociais internas de intranet. A influência das redes sociais na vida das pessoas é imensa e tende a crescer, ditando estilos de vida e moldando opiniões. O acesso à informação tem sido facilitado e generalizado, atualizações são disponibilizadas em diversos canais e dispositivos.

As companhias de mídia de larga escala - jornais, revistas e entretenimento - ao identificar o potencial deste novo canal (meio de comunicação) para a veiculação de notícias, disseminação de rumores e construção de ideias, focam-se então no investimento em seguidores dispostos a receber a notícia ‘fresca’ no momento em que ela for redigida. Em vez de ser necessário entrar em um site para ler uma notícia, agora basta assinar um canal que “a notícia vem até você” pelas atualizações de postagens deste, seja por e-mail ou até por *push* (notificações instantâneas) de celular.

Uma nova funcionalidade é disponibilizada por quase todas as redes sociais: a possibilidade de que o público leia uma matéria, um texto ou veja uma foto em algum canal online, como um site de internet, e os possa compartilhar em suas redes sociais com apenas um ou dois cliques, agindo como um propagador de conteúdo. Considerando a credibilidade dos canais de notícia e a do usuário que compartilha a notícia, toda a postagem passa a ter incrível potencial de alcançar alto número de pessoas, seja esta notícia confiável, meia-verdade ou um rumor, rapidamente circula pelos meios digitais. Tendo-se milhares ou até milhões de visualizações a determinada informação, pode-se facilmente definir quais canais e hubs sociais (pessoas que centralizam em si múltiplos contatos de rede - também chamadas nesse trabalho de celebridades de rede social - seriam mais valiosos como reprodutores das informações a certos grupos de pessoas. (KWAK, HAEWOON, et al., 2010).

A plataforma de *microblogging* Twitter é amplamente conhecida e divulgada, usada por centenas de milhões de pessoas em todo o mundo. Apresenta características de rede assimétrica e dirigida, permitindo que qualquer pessoa tenha acesso ao que os outros postam, sem necessidade de autorizações ou permissões adicionais. Pela facilidade na propagação de informações, tem sido usada tanto para a divulgação de ideias, quanto de notícias e rumores permitindo visibilidade de uma postagem para grandes audiências.

O papel do Twitter foi fundamental em eventos marcantes como a "Primavera Árabe" e os "Protestos de Londres", facilitando e fortalecendo os ajuntamentos e unificação de ideias que se propagam rapidamente (DOERR, et al., 2012).

Alguns dos eventos de protesto social e mobilizações políticas mais importantes dos últimos anos tiveram o Twitter como principal ferramenta de divulgação e ativação.

Ao estudar o crescente uso das redes sociais, bem como seu poder de propagação de informações, diante da quantidade de postagens realizadas diariamente por canais de notícias com alto número de seguidores, surge a motivação para este projeto: analisar a propagação e o alcance de notícias postadas na rede social Twitter, observar a similaridade entre a circulação das postagens e os estudos do modelo de rumor.

1.1 DESCRIÇÃO DO PROBLEMA

Cada vez que uma notícia é divulgada nos canais digitais de comunicação, sua promoção é feita nas redes sociais, principalmente no Twitter, por meio da divulgação da manchete da notícia. Para AARTS, OLAV, et al. (2011) e KWAK, HAEWOON, et al. (2010), 85% das postagens com maior propagação na rede social Twitter são manchetes ou derivações de notícias que persistem naturalmente na rede, se propagando sem necessidade de nova interferência do emissor, razão pela qual as grandes redes de mídia têm se utilizado do Twitter. No mesmo estudo é dito que o sucesso da popularidade (*buzz*) das postagens, mesmo as enviadas por celebridades ou usuários populares com alto alcance, necessita da reprodução (retuite ou menção) dos usuários não populares. O número de seguidores é importante, mas não é o único ponto a ser considerado, existem outras informações também importantes para se avaliar o valor de uma postagem, como a quantidade de retuites, que demonstra a influência da postagem, e a quantidade de menções, que demonstra os ruídos causados pela mesma (ROSENMAN E EVAN, 2012).

Diversos fatores contribuem para que uma notícia se torne um *hit* ou um sucesso de divulgação e alcance o maior número de pessoas dentro e fora das redes sociais. Continuamente os jornais, revistas e emissoras de TV buscam identificar quais seriam as formas mais efetivas para que sua notícia se propague com maior alcance. Usuários de redes sociais, anônimos e famosos, valorizam o “número de seguidores” como uma medida de poder e de maior vitrine para demonstrar sua influência social, o que inclusive resulta em contratos com remuneração para celebridades do mundo virtual, valendo o esforço de ampliar sua visibilidade na rede. Em alguns casos, vale a pena manter um trabalho profissional especializado na administração das redes sociais. O foco desta pesquisa consiste na descrição e análise da rede de usuários do Twitter, análise da propagação de postagens na rede social Twitter e a influência dos usuários e suas postagens.

1.2 OBJETIVOS

1.2.1 Objetivo Principal

O propósito deste trabalho é analisar as postagens na rede social Twitter e suas características de propagação, taxa de retransmissão e duração das postagens. As postagens e repostagens de notícias ou rumores são oriundas de usuários que são agências formais de notícias e usuários que não são agências formais de notícias.

1.2.2 Objetivos Secundários

Os objetivos mais específicos deste trabalho envolvem analisar o impacto e influência dos usuários de redes sociais, e o quanto uma notícia, informação, novidade ou rumor persiste na rede.

Diante desse cenário proposto, as seguintes questões tornam-se pertinentes:

- Há diferença na duração para postagens realizadas por usuários de alto alcance ou usuários geradores de notícias?
- Qual seria o impulso de um retuite para divulgação de notícias?
- Qual a topologia de redes complexas com que a rede social Twitter mais se assemelha?

1.3 ESTRUTURA DA DISSERTAÇÃO

A dissertação está estruturada da seguinte maneira: no capítulo 1 há a contextualização do Twitter, história das redes sociais, descrição breve da problemática e descrição dos objetivos principal e secundários. No capítulo 2, apresentamos a revisão da literatura sobre propagação de informação na rede social Twitter e teoria do rumor. No capítulo 3, a análise empírica da rede de usuários. No capítulo 4, temos a simulação e a construção do modelo. E por fim, a conclusão e as considerações finais encontram-se no capítulo 5.

2 REVISÃO DA LITERATURA

"Twitter é uma rede social e um servidor para *microblogging* (blog para postagens curtas), que permite aos usuários enviar e receber atualizações pessoais de outros contatos, através de mensagens curtas de até 140 caracteres, conhecidas como 'tweets', por meio do website do serviço, e por softwares específicos de gerenciamento. Criado em março de 2006 por Jack Dorsey, Evan Williams, Biz Stone e Noah Glass, e lançado em julho de 2006 nos EUA" (TWITTER, 2017).

Inicialmente chamada Twtr (sem vogais), o nome da rede social, em inglês, referenciava a reprodução dos sons emitidos pelos pássaros. A ideia inicial seria de que os usuários postassem para chamar atenção ao que quisessem dizer pela internet. O ambiente do Twitter compreende uma estrutura em que cada usuário possui uma conta e uma linha do tempo. Em sua conta há informações sobre a data de abertura da conta, a rede de usuários seguidos e rede de usuários seguidores. Na linha do tempo de cada usuário aparecem as postagens do próprio usuário, dos usuários a quem ele segue e as em que ele foi referenciado, aparecendo as mais recentes no topo.

As postagens na linha do tempo são exibidas em ordem decrescente de data e hora, ou seja, as mais novas ocupam o topo da linha do tempo, tomando o lugar e a evidência das postagens anteriores. Para ler as mensagens anteriores o usuário precisa rolar a tela quantas vezes forem necessárias. Sendo assim, uma postagem se torna visível na tela do usuário por certo período de tempo até que cheguem outras novas (TECMUNDO-1).

Das plataformas atualmente disponíveis de *microblogging* o Twitter é indubitavelmente uma das mais conhecidas, usada por centenas de milhões de pessoas em todo o mundo. Os usuários atualizam suas atividades diárias através de computadores ou telefones móveis, transmitindo o que acontece em seu cotidiano. Essencialmente, os relacionamentos no Twitter são assimétricos, um usuário é seguido por outro, o qual não tem necessariamente que retribuir seguindo-o de volta. Algumas redes sociais adotam relações simétricas. Ao passo que no Facebook, por exemplo, uma relação é estabelecida quando um pedido de amizade é aceito por um usuário, o que aumenta cada uma das listas de contatos e, se um usuário remove o outro, o relacionamento é quebrado. Por conseguinte, um diferencial importante na rede Twitter é ser dirigida, enquanto que outras, como o Facebook são bidirecionais, e boa parte das informações pode ser acessada sem identificação. Isto é, no Twitter, um usuário pode acessar todas as informações de outro usuário sem a necessidade de autorizações ou permissões adicionais.

Em comparação com os meios tradicionais, como jornais e televisão, as redes sociais permitem a criação e troca de conteúdo gerado pelo usuário, sendo que cada usuário pode produzir e distribuir mensagens. Isto resulta em um crescimento explosivo na quantidade de informações e faz com que muitas redes sociais se tornem saturadas. Devido ao potencial de marketing e publicidade, o Twitter e outras redes sociais são considerados meios eficientes para estimular a utilização de produtos ou serviços. Pela facilidade com que informações são propagadas, profissionais de jornalismo e marketing costumam procurar as redes sociais para construir notícias, rumores e disseminar ideias, valendo-se da grande exposição que as redes sociais proporcionam.

São chamados de ‘celebridades de redes sociais’ as pessoas que tem um alto grau de conectividade (AARTS, OLAV, et al., 2012), seja por terem muitos seguidores ou por ter entre seus seguidores outros usuários com muita conectividade e que sejam ativistas em compartilhar suas postagens (DOERR, et al., 2012). Os usuários com menor conectividade também têm um papel importante na propagação e disseminação das postagens, já que fazem a conexão entre usuários de outros agrupamentos ou áreas de interesse.

No entanto, devido às limitações da capacidade de processamento das informações pelas pessoas, se as mensagens chegam em número maior do que os usuários podem processar, algumas são perdidas, sem prender a atenção dos usuários (GONÇALVES, et al., 2011), de forma que os usuários tendem a buscar mais conteúdo qualitativo do que quantitativo (CHA, MEEYOUNG, et al., 2012). Os usuários compartilham tanto informações reais e notícias importantes quanto boatos, podendo citar ou não as fontes e expressar opiniões pessoais. Entretanto, as informações têm maior confiabilidade quando são compartilhadas por canais geradores de notícias (DOERR, et al., 2012), em que a manchete é postada e costuma haver um link levando para a matéria completa sobre o assunto.

Tabela 1- Estatísticas de Mensuração do Twitter.

<i>Informações Estatísticas do Twitter (Dados de 30 de junho de 2016)</i>	
<i>Usuários mensais ativos</i>	313 x 10 ⁶
<i>Visitantes com ao menos um acesso a sites usando dados do Twitter</i>	1 x 10 ⁷
<i>Usuários regulares em celular</i>	82%
<i>Empregos em nível global</i>	3.860
<i>Escritórios ao redor do mundo</i>	+35
<i>Usuários fora dos EUA</i>	79%

<i>Suporte a diferentes idiomas</i>	+40
<i>Empregados em papéis técnicos</i>	40%

Fonte: TWITTER, 2017

A tabela 1 apresenta informações estatísticas e de mensuração do Twitter, contendo dados de métrica e as datas em que as informações foram levantadas. Esses dados crescentes são constantemente atualizados e disponibilizados em áreas específicas da ferramenta para consulta.

Há termos e convenções bem definidas e estritas no Twitter, conforme segue abaixo:

- Twitter: Aplicativo e plataforma de *microblogging*
- Tuíte (*Tweet*): Postagem por mensagem na Rede Social Twitter, com até 140 caracteres, enviado para todos os seguidores do usuário que assinaram para acompanhar suas publicações.
- Retuite ou RT (*Retweet*): Ação de um usuário que compartilha a postagem de outro usuário, enviando desta forma a postagem original para sua rede de seguidores, amplificando assim a propagação da postagem original do usuário inicial.
- Menção (*Mention*): Quando um usuário cita outro em seu tuíte através do uso do operador '@', ao escrever '@' seguido do nome de um usuário válido. A ferramenta faz o vínculo com o usuário citado, aparecendo na linha do tempo do mesmo.
- Seguidor (*Follower*): Referência aos usuários que seguem as postagens do usuário em questão.
- Seguindo (*Following*): Referência aos usuários seguidos pelo usuário em questão.
- Hashtags: Quando um usuário escreve o tuíte fazendo uso do operador '#'. Usa-se o operador *hashtag* seguido de uma palavra, que se torna especial, fazendo parte de um conjunto com outros tuites em que o mesmo *hashtag* é utilizado.
- Assuntos do Momento (*Trending Topics*): Na interface da página do Twitter, aparecem os *Hashtags* que estão sendo mais utilizados atualmente em ordem decrescente da quantidade de utilizações.
- Mensagem Direta (*DirectMessage*): Mensagem direta a um usuário sem visibilidade dos demais, como um e-mail interno.

- Resposta (*Reply*): Quando um usuário menciona (@) outro publicamente após ter sido mencionado por ele.

Uma descrição completa dos termos utilizados neste trabalho em relação ao Twitter está no Apêndice B.

Informações relevantes sobre os tuites conforme BOYD, et al. (2010):

- 36% mencionam outros usuários
- 86% dos retuites iniciam com menção a um usuário, com formato de resposta pública.
- 5% contém ao menos um *hashtag* (#) e desses 41% tem também um endereço de URL (endereço de site de internet)
- 22% incluem um endereço de URL
- 3% parecem ser retuites, por conterem na mensagem os termos “RT” (88%), “via” (11%) ou “retweet” (5%)

2.1 PROPAGAÇÃO DE INFORMAÇÕES NO TWITTER

No estudo de KWAK, HAEWOON, et al. (2010) com o objetivo de analisar as características topológicas do Twitter e seu poder de propagação como canal de compartilhamento de informações de mídia, observaram uma distribuição não-linear com ligação preferencial de seguidores, curto diâmetro efetivo e relacionamentos unidirecionais de baixa reciprocidade. Todas características comuns das redes sociais humanas não virtuais (NEWMAN, 2003).

Geralmente uma das formas mais comuns de se medir a importância que uma postagem atingiu é observando a quantidade de retuites. E para se medir a popularidade de um usuário se observa a quantidade de retuites que suas postagens atingiram. Em alguns estudos os usuários e as postagens são classificados por dados de retuites, como por exemplo, no estudo de SUH (2010).

Características de difusão de inovação foram observadas no estudo de LESKOVEC (2009) ao analisar rastreamento de memes em ciclos de notícias na internet, com elementos que demonstram larga escala de difusão. Ao analisar a árvore de retuites como canais de comunicação de difusão da inovação, ROSENMAN (2012) atesta que retuites alcançam larga audiência e se espalham rapidamente.

Para CHA, MEEYOUNG, et al. (2012), o Twitter representa uma oportunidade única para responder novas questões sobre ciências sociais e a interação entre indivíduos de diferentes tipos de popularidade, observando as dinâmicas de canais de notícia e influência para processos de difusão da inovação. O que está de acordo com os estudos de ROGERS (2010) e WATTS e DODDS (2007).

Acerca dos conteúdos das postagens, segundo AARTS, OLAV, et al. (2011), postagens de tópicos gerais possuem maior visibilidade dos seguidores e tem maior probabilidade de ser retuitados do que mensagens de interesse pessoal. Da mesma forma mensagens que contenham hashtags, nomes de usuários, URLs, interrogação e tuites intensos são mais prováveis de serem retuitadas do que mensagens semelhantes com pontos de exclamação e *emojicons* positivos. Da mesma forma, a influência e reputação do usuário originador da influência trazem peso sobre a credibilidade da mensagem nas interações pessoais, mesmo nas redes sociais não virtuais (KATZ, et al., 1966).

ROSENMAN e EVAN (2012) ao descrever vários aspectos da rede Twitter evidenciam os seguintes pontos:

- 22% dos relacionamentos do Twitter são mútuos
- Geralmente os usuários se conectam a outros sem interesse em comunicação direta
- A maioria dos tuites recebem pouca atenção
- Uns poucos recebem muita atenção, apresentando uma distribuição de lei potência com cauda longa
- Influência em redes sociais não é uma simples questão de número de seguidores, deve ser levado em conta o número de retuites e o número de menções
- O número de conexões de grau, como retuites e menções, são indicadores mais preditivos do que número de seguidores.
- Acerca das 60 celebridades com maior número de seguidores, ROSENMAN e EVAN (2012) apontam:
 - O maior impacto do quanto uma celebridade é falada nas redes sociais está associado a quantidade de retuites circulando com sua menção (Spearman 0,70)
 - Tuites enviados e contagem de retuites estão positivamente correlacionados (Spearman 0,44)
 - Retuites são significativamente correlacionados com o número de seguidores (Spearman 0,58)

- Não há correlação entre o tamanho da audiência e frequência de postagens (seguidores não são atraídos por uma larga quantidade de geração de conteúdo)
- Não há correlação entre a quantidade de citações e referências e a quantidade de postagens

No trabalho de KWAK, HAEWOON, et al. (2010), ao classificar os *Assuntos do Momento* (*trending topics*) e os tuites num período, pôde ser observado que a maioria (85%) dos tópicos são normalmente manchetes de notícias ou notícias persistentes, ou seja, postagens que voltam a ser veiculadas ou que são citadas por outras depois de um período. Olhando mais de perto os retuites revelam que qualquer retuite pode atingir 1000 usuários em média, independentemente do número de seguidores que tenha o tuite original, uma vez retuitado em no máximo 1 hora, significa rápida difusão da inovação depois do primeiro retuite (AARTS, OLAV, et al., 2012). Existem postagens que são retuitadas algumas vezes logo após a primeira postagem, demonstrando alta difusão após o primeiro retuite. Quanto mais recente a postagem maior a chance de a mesma receber a impulsão através de retuites.

O Twitter tem emergido como um meio popular para a discussão de acontecimentos marcantes por todo o mundo, tendo recebido a atenção desde usuários comuns a usuários profissionais da mídia, e até de canais de televisão, geradores de notícias e usuários famosos e influentes (WATTS E DODDS, 2007). Através do uso do Twitter qualquer usuário se assemelha a um repórter divulgando o vídeo de um furo jornalístico, sendo o responsável pela propagação de informações de forma simples para grandes audiências.

Na literatura recente o Twitter tem recebido atenção em pesquisas como sobre a motivação qualitativa do uso da plataforma (ZHAO, et al., 2009), em que os seguidores de um usuário geralmente não possuem vínculo com o usuário seguido (HUBERMAN, et al., 2008), e sobre analisar o boca-a-boca e as discussões públicas que marcam as linhas do tempo (JANSEN, BERNARD J., et al. 2009).

Analisando a abrangência de tuite e a ampliação de propagação causada pelo retuite, KWAK, HAEWOON, et al. (2010) observaram que quando um tuite é postado por um usuário, é visto e disponível para ser visto por todos os seguidores do usuário que postou. Havendo uma valorização adicional de quantos usuários fora da rede inicial de seguidores foi alcançada por essa postagem. Em concordância com essa afirmação, GRABOWICZ, et al. (2012) demonstram que retuites transcendem a comunidade de amigos do usuário originador e servem como ponte para que mensagens sejam propagadas por toda a rede (BORONDO et al., 2012).

No artigo de GONÇALVES, et al. (2011) foi analisada uma base de dados de conversas em Twitter coletadas em seis meses, envolvendo mais de 1,7 milhão de indivíduos e testando o limite cognitivo teórico sobre o número de relações sociais estáveis conhecidos como o número de Dunbar. DUNBAR (1992) mediu a correlação entre o volume neocortical e o tamanho de um grupo social típico em uma ampla gama de primatas e comunidades humanas. Com um resultado surpreendente e de longo alcance, concluiu-se que a capacidade de processamento neocortical aparece para definir o número de indivíduos com os quais é possível manter relacionamentos interpessoais estáveis.

Portanto, o tamanho do neocórtex do cérebro representa uma restrição biológica na interação social que limita o tamanho da rede social dos seres humanos entre 100 e 200 indivíduos (DUNBAR, 1998), esse número é chamado de número de Dunbar. No artigo em questão, uma das conclusões, pela análise de dados do mundo real, argumenta que a facilidade fornecida de contatos e conexões pelas redes sociais digitais, no caso o Twitter, não fornece maneiras de extrapolar o limite de interações sociais do número de Dunbar (GONÇALVES, et al., 2011). Restringindo assim, a quantidade de usuários que podem ser acompanhados por um usuário seguidor.

O processo de saturação faz com que alguns relacionamentos sejam mais valorizados do que outros. Como cada indivíduo tenta otimizar seus recursos, priorizando essas interações e assim qualificando a força dessas interações, estudando a quantidade ωi^{out} , definida como a força social média dos relacionamentos inicialmente ativos, sendo $\omega i^{out}(T) \equiv \frac{\sum_j w_{ij}(T)}{k_i^{out}}$.

Onde:

T é o tempo;

ω_{ij} é a correspondência do peso

j é o receptor da mensagem, incrementado por 1 a cada rodada

k_i^{out} é o número de conexões recíprocas, em função do *in-degree*, com valor entre 200 e 300 embora o número de conexões entrantes continue a crescer.

Conforme GONÇALVES, et al. (2011), a cada iteração do agente em sua fila de execução, a seguinte operação é realizada:

- O agente responde a uma quantidade aleatória de mensagens entre 0 e o número de mensagens q_{max} na fila.

- As mensagens a serem respondidas são selecionadas proporcionalmente pela prioridade do agente remetente (com base em seu grau de alcance total).
- A cada mensagem sendo enviada para j , o nó é incrementado após a resposta.
- Mensagens que o agente tem respondido são excluídas da fila e as recém-chegadas são trazidas para a fila em ordem de prioridade até que se alcance q_{max} , e as mensagens restantes são descartadas.

Para um número de contatos maior do que o tamanho da fila, o usuário será incapaz de responder a todas as mensagens na fila. Uma vez que o resultado da saturação é alcançado, geralmente o usuário responde à mesma quantidade de mensagens a cada iteração. O número médio de respostas por conexão segue a escala $\omega_i^{out} \propto \frac{q_{max,i}}{k_i}$.

Para GONÇALVES, et al. (2011), é possível encontrar uma complexidade ao observar padrões empíricos nos dados do Twitter, oferecendo suporte à hipótese de Dunbar de limitação biológica ao número de relacionamentos. O mecanismo de interação social de atenção limitada e priorização individual de importância dos relacionamentos e interações sociais pode ser estendida para outras áreas de cenários sociais mais amplos.

De acordo com a pesquisa de CHA, MEEYOUNG, et al. (2012), usando dados empíricos de notícias espalhadas acerca de manchetes internacionais e tópicos mais simples, foram estudados os papéis relativos de três tipos de usuários difusores de informações na rede social Twitter:

- 1) Meios de comunicação de massa, como jornais, emissoras e canais de notícia,
- 2) Difusores de base, composta por usuários comuns; e
- 3) Difusores evangelistas, consistindo de líderes de opinião, políticos, celebridades e de empresas locais.

Os meios de comunicação de massa desempenham um papel vital no alcance da maioria da audiência em qualquer dos principais tópicos. Os evangelistas complementam apresentando os tópicos para maiores e menores audiências que estão mais distantes do centro da rede e

seriam quase inalcançáveis de outra forma. Os difusores de base são relativamente passivos no espalhar das notícias mesmo representando 98% da rede.

Figura 1 - Distribuição de graus dos usuários do Twitter. Fonte: CHA, MEEYOUNG, et al. (2012)

A Figura 1 mostra a fração dos usuários na rede com determinado *in-degree* e *out-degree* maior do que x , e os graus dos usuários representando seu poder de conectividade. As duas distribuições são similares exceto pelo fato de duas quedas análogas no *out-degree* por volta dos valores 20 e 2000. As tendências de conectividade entre os usuários diferenciam o Twitter das redes sociais convencionais, em direção a ser uma ferramenta colaborativa de publicação de notícias e rumores.

Figura 2 - Número de destinatários de retuite por número de seguidores.

Considerando-se os números destinatários secundários acessados por retuite, mostrando a média e mediana por tuite em relação ao número de seguidores do usuário originador do tuite. Fonte: FILOTAS, et al (2014)

A mediana encontra-se quase sempre abaixo da média, indicando que muitos tuites tem um largo número de destinatários adicionais. Até 1.000 seguidores, o número médio de destinatários adicionais não é afetado pelo número de seguidores do tuite original. Ou seja, não importa quantos seguidores um usuário tenha, o tuite deve atingir um certo número de audiência, uma vez que o tuite comece a se espalhar por retuite.

Figura 3 - Proporção e usuários participantes na árvore de retuites.

Foram plotadas a CCDFs (função de distribuição cumulativa complementar) da altura da árvore de retuites e o número de usuários na árvore de retuites. Fonte: (KWAK, HAEWOON, et al., 2010)

A altura igual a um é o mais comum abrangendo 95,8%. Como 97,6% da paridade dos nós tem menos de seis graus de separação, todas as árvores de retuites, senão para um punhado com uma altura menor do que seis, e nenhuma árvore foi além de 11 ciclos. A distribuição dos usuários na árvore de retuite segue uma lei de potência (KWAK, HAEWOON, et al., 2010). Essa análise da árvore de retuite demonstra como retuites propagam e quantos usuários são envolvidos.

Pesquisadores defendem que rumores se espalhem na presença de um dos 3Cs¹ e a explicação mais comum para esse fenômeno tem sido ligado a ansiedade de prever resultados negativos (AARTS, OLAV, et al., 2012). Tais teorias podem explicar porque alguns rumores se propagam mais rapidamente em tempos de pânico, ainda que não sejam úteis na explicação mais considerada para rumores em situações positivas, como um festival ou um evento social importante.

¹ 3Cs: Conflito, crises e catástrofes. Quando tensões e problemas sociais evoluem e se tornam complicações de maiores magnitudes como conflitos, crises e catástrofes. Situações propícias para que rumores sejam espalhados.

Na pesquisa de MORALES, et al. (2012), ao analisar dados coletados na API do Twitter com o objetivo de descrever a estrutura e as dinâmicas emergentes da rede social, em um evento de protesto venezuelano exclusivamente organizado pelo Twitter em dezembro de 2010, foi observada a influência das redes sociais em eventos sociais como o protesto e o impacto dos canais de mídia no Twitter. Confirmando o que já foi observado em estudos anteriores que, embora as redes sociais possam parecer um ambiente puramente social e virtual, os canais tradicionais de mídia ainda têm cargas de influência na rede (MORALES, et al., 2009).

Figura 4 - Taxa de mensagens e Tamanho de mensagens.

Painel Superior: Variações temporais da taxa de mensagens (mensagens/minuto) do Twitter no venezuelano #SOSInternetVE. Setas indicam alguns dos momentos de convocação de participantes ao protesto.

Painel Inferior: Sistema cresce no tempo em termos do percentual de mensagens (linha tracejada) e o percentual de usuários participantes (linha sólida).

Fonte: BORONDO et al. (2012)

É notável o fato de que 42% das postagens são retuites e 60% das mesmas foram enviadas de smartphones, o que está em concordância com pesquisas de tendências de usuários, em que a maioria dos usuários participa de redes sociais com smartphones. Analisando as mensagens ao se aproximar da data do protesto, a forma da evolução se assemelha com a modelagem das séries do Twitter no estudo de YANG, et al. (2011).

2.2 TEORIA DO RUMOR

O conceito de rumor é definido por uma afirmação não verificada de um evento no momento da circulação da informação, quando um rumor possui ambiguidade em sua composição, tem uma ampliação de sua exibição pelo aumento de interesse dos expectadores (PETERSON, et al., 1951).

Ao analisar a propagação de rumores em *microblogging* como o Twitter, KWON, SEJEONG, et al. (2013a) dizem que os usuários tendem a identificar informações entre verificadas e não-verificadas, baseado em seus próprios julgamentos. Na psicologia, a motivação para a propagação de uma informação pejorativa ou um rumor está vinculada ao fato da pessoa se sentir superior ao depreciar outros. Logo quando um rumor é recebido, o receptor da informação pode aceitar, rejeitar ou propagar dependendo da confiabilidade da origem (CANTRIL, 1940, FESTINGER, 1948).

O crescimento da popularidade das redes sociais online e de seu potencial para propagação de informações e rumores, a análise das características de publicações e postagens, bem como o alcance das propagações, têm sido importantes fatores de estudo. No artigo de KWON, SEJEONG, et al. (2013b) é proposta uma nova abordagem para identificar rumores baseados em propriedades temporais, estruturais e linguísticas de propagação de rumor. Considerando que as estruturas possibilitam inclusive informações preditivas, em um trabalho baseado em dados reais, analisando três anos e meio de dados quase totais da base do Twitter, foram extraídos 104 eventos virais, cada um deles envolvendo ao menos 60 postagens.

Ao estudarmos as dinâmicas que influenciam a opção de novas conexões nos deparamos com a característica de ligação preferencial (PA - *Preferential Attachment*), característica de redes em que os nós que estão na rede há mais tempo ou que possuem maior conectividade tendem a ser preferidos em relação aos outros para receber novas conexões (DOERR, et al., 2012), formando agrupamentos (*clusters*) em torno dos usuários com maior ligação preferencial gerando uma rede com característica "rico fica mais rico" (*rich get richer*). A conectividade do nó segue uma lei de potência, onde um nó com k número de vizinhos é inversamente proporcional ao polinômio em k . Ao explicar isso, BARABÁSI, et al. (1999) sugerem o modelo de ligação preferencial para redes no mundo real, que mostra uma lei de potência k^{-3} matematicamente, confirmada posteriormente por BOLOBÁS, et al. (2001).

Conforme dito por DOERR, et al. (2012), ao analisar as postagens como notícias que se propagam nas redes sociais, foi realizada uma pesquisa no processo de simulação da propagação de uma mensagem simples em diferentes topologias de rede. Foi observado que as notícias

propagam muito mais rapidamente em topologias de redes sociais, considerando a importância tanto de nós com alta conectividade quanto de nós com baixa conectividade. Pelo fato de que usuários com alta conectividade exercem o papel de Hub levando a informação para grandes audiências de uma mesma linha ou agrupamento, enquanto usuários de baixa conectividade levam a informação para outros agrupamentos, por isso que notícias e rumores se espalham rapidamente em redes sociais (DOERR, et al., 2012, PFEIFFER III, et al., 2011).

Para DOERR, et al. (2012), um rumor que inicia num nó aleatório na rede Twitter alcança em média 45,6 milhões de nós de um total de 51,2 milhões de nós da malha total em apenas 8 rodadas de comunicação. Valorizando-se assim os nós com poucas conexões como agentes cruciais para rápidas disseminações.

Explicando parcialmente porque pode-se observar que as redes sociais propagam informação rapidamente, mesmo que o processo não seja organizado centralmente, e a rede não tenha sido projetada com escolhas propositalmente das conexões. É crucial a interação multiplicadora entre agrupamentos com muitas conexões e usuários médios com poucas conexões. Os agrupamentos tornam as notícias disponíveis para grandes audiências, ao passo que usuários médios rapidamente levam a informação de um vizinho para outro.

No trabalho de LIU et al. (2011) observamos que o processo de rumor é uma boa maneira de analisar o método de modelagem baseado em agentes. Nesse trabalho eles simularam uma rede Livre de Escala, com comportamento similar ao do Twitter, e usaram a ferramenta Netlogo como plataforma de simulação da propagação do modelo de SIR. Observando a propagação de rumor em sites como o Twitter, considerando o fluir das informações, o tempo máximo de propagação, e os mecanismos que levam as pessoas na rede a acreditarem no rumor, os autores do artigo chegaram à conclusão de que o rumor se propaga de forma bem eficiente em redes desta topologia, mas de forma imprevisível, e com tendência de que a propagação do rumor reinicie por si só sem nova inserção na rede.

MORENO, et al. (2004) realizaram estudos de propagação de rumor em rede Livre de Escala, separando os usuários em três grupos: Ignorantes: Não conhecem a informação - viram transmissores com possibilidade a . Transmissores: Conhecem e propagam a informação - viram abafadores com possibilidade a onde a é o número de vizinhos Transmissores. Abafadores (*Stifler*): Conhecem a informação e não propagam – Probabilidade de virarem transmissores.

De acordo com LIU et al. (2011), a quantidade atingida pela propagação do rumor em redes sociais online como o Twitter pode ser analisada pela importância da informação, os níveis de influência dos envolvidos e a complexidade da malha de rede em questão. Ao falar

sobre a importância da informação, o rumor tem maior alcance quando a informação é de interesse de um público maior. A influência dos envolvidos refere-se ao quanto o assunto atrai a atenção de celebridades nas redes sociais, causando o engajamento desses usuários, o que aumenta o interesse e a credibilidade na rede.

A complexidade da malha refere-se à densidade da rede e a quantidade de agrupamentos (*clusters*) pelo qual a mensagem de rumor irá circular. Na pesquisa em questão de LIU, et.al. (2011), foi simulada uma rede Livre de Escala, a pesquisa foi realizada por modelagem baseada em agentes com a ferramenta Netlogo, considerando a análise do processo de rumor pelo modelo de propagação SIR.

Ao analisar o importante papel de redes sociais como Facebook e Twitter na "Primavera Árabe" e nos "Protestos de Londres", tanto viabilizando os ajuntamentos quanto no fortalecimento da unificação de ideias, pode-se observar o quanto as redes sociais propagam notícias e rumores muito rapidamente (DOERR, et al., 2012), em comparação com outros tipos e topologias de rede.

No artigo de KOSTKA, et al. (2012), a propagação de rumor é considerado como um jogo (*Rumor Game*), modelo de disseminação de competição da informação em rede, de forma que dois nós escolhidos aleatoriamente na rede são simulados como postando uma mesma informação, um rumor, sob as mesmas condições, analisando a capacidade de disseminação da informação como uma competição entre eles, sendo que a persuasão do rumor depende das mensagens de rumor em relação ao total geral de mensagens.

O entendimento das possíveis formas de propagação e comportamentos influenciam o valor de persuasão atribuídos ao rumor e a decisão dos receptores em repassar ou não com possibilidade P_{rumor_i} , sendo: $P_{rumor_i} = \frac{\# mensagens\ rumor_i}{\# mensagens}$.

Considerando a propagação do rumor como dependente da persuasão de forma linear (KOSTKA, et al., 2012), um modelo mais complexo como o 'modelo limiar linear' (*linear threshold model*) ou o "modelo de cascata independente" (*independent cascade model*) poderiam ser implementados. O modelo linear foi introduzido por GRANOVETTER (1978) e SCHELLING (1978), que publicaram as primeiras definições para um modelo que manipula a propagação da informação em redes. O 'modelo de cascata independente' tem sido proposto no contexto de marketing por GOLDENBERG, et al. (2001).

Onde um nó u recebe o rumor em uma oportunidade de propagar o rumor i para o vizinho v com probabilidade pu de v . Nesse modelo, um nó u encaminha um rumor, e não há mais

tentativas do nó u convencer o nó v a participar na propagação. KEMPE, et al. (2005) mostram que esses dois modelos podem ser generalizados posteriormente e por fim, equivalentes. A forma com que os rumores se propagam em redes complexas tem sido alvo de diversos estudos, a seguir estudaremos com maior profundidade as características das redes complexas que agregam valor a essa disseminação. Informações e bibliografia sobre Redes Complexas podem ser encontradas no Apêndice A.

3 ANÁLISE EMPÍRICA DA REDE DE USUÁRIOS DO TWITTER: UM ESTUDO DE CASO

Neste capítulo apresentamos a descrição e os resultados da análise da rede Twitter empírica, obtida para estudar as ocorrências do termo “Dilma”, nas semanas que antecedem as manifestações de março de 2015. A descrição do processo do método técnico utilizado neste trabalho pode ser encontrada no Apêndice C.

3.1 DESCRIÇÃO DA REDE

Para a realização deste trabalho, foi monitorado o termo “Dilma” na rede social Twitter no período de 01 a 16 de março de 2015. Este período abrange uma importante manifestação popular na forma de protestos nas ruas no dia 15 de março do mesmo ano. Foram armazenadas informações de mais de 1,6 milhões de postagens. Após a limpeza e o tratamento dos dados, obtivemos um banco de dados com 1.522.716 registros, contendo os atributos listados na tabela 2.

Tabela 2 - Tabela de Dados 'Posts'

Tabela:	Posts
Atributo	Dicionário
Contador	Número sequência de contagem do número da linha
Data_Post e Hora	Data e Hora que a postagem foi publicada no Twitter
Nome_usuario	Nome de cadastro do usuário
Perfil_usuario	Usuário público que é visível para ser seguido e referenciado
ID_usuario	Identificador único da postagem codificado do Twitter
Posts_tot_usuario	Quantidade de Postagens totais do usuário no Twitter no momento desta postagem
ID_post	Identificador único da postagem codificado do Twitter
Texto do Post	Texto original da postagem
Seguidores	Quantidade de seguidores do usuário
Dispositivo	Tipo de dispositivo que fez a postagem: Móvel ou Computador

Na tabela 2 temos os dados coletados no monitoramento da rede social Twitter no período escolhido e com os critérios de nossa pesquisa. A lista exhaustiva de todas as tabelas e campos trabalhados tanto para análise quanto para geração dos dados empíricos pode ser vista no Apêndice D. As estruturas e os códigos-fonte da criação das tabelas encontram-se no Apêndice E. Após o tratamento dos dados algumas tabelas auxiliares foram criadas, sendo a principal delas a TABELA DE USUÁRIOS. Para montar a rede de usuários do banco de dados elegemos como nós, os usuários que postam e usuários referenciados ou citados na postagem. O nó pode ser classificado como seguidor, seguido ou audiência. As arestas da rede são as referências (retuites) ou citações (@s). É difícil fazer uma análise visual da rede devido ao seu tamanho. No entanto, podemos estudar os aspectos mais relevantes por meio das métricas de rede.

A quantidade de retuites (ou repostagens) na base de dados é de 932.690. Ao considerar o número médio de retuites de postagens que tiveram ao menos um retuite temos a média para notícias de 68,37 retuites e para as que não são notícias de 27,25 retuites.

A quantidade de usuários geradores de notícias é de 1.244, e a quantidade de postagens por usuários não originadores de notícia teve o máximo de 2.511 e em média 5,1312, enquanto de postagens por usuários originadores de notícia foi de no máximo 285 e em média 34,5556.

A quantidade média de postagens em que outro usuário é mencionado *in-degree*, vindos de usuários não originadores de notícia foi de 20 postagens, e vindos de usuários originadores de notícia foi de 214.314. Quando falamos de postagens nas quais outros usuários mencionam *out-degree*, dos não originadores de notícia foram 46.006, e dos originadores de notícias foram 35. Por fim, quando falamos de número de seguidores dos usuários, para os não geradores de notícias temos o máximo de 7.081.306 e média de 1.831, e para os geradores de notícias temos o máximo de 9.291.153 e a média de 2.017.

Diante dessas informações obtidas e trabalhadas foi realizado um processo de extração de informações empíricas de observação dos dados e seus resultados que possibilitou uma análise mais detalhada dos comportamentos conforme demonstrado na sessão seguinte.

Na próxima sessão apresentaremos as métricas de rede em relação aos dados dos usuários do experimento.

Tabela 3 - Tabela de métricas da rede de usuários chamados (U1) e chamadores (U2). Extraídos das ferramentas Python e R após importação da tabela como *dataset*.

Tamanho total da rede de usuários	1.046.999
Quantidade de usuários chamadores únicos	210.373
Quantidade de usuários chamados únicos	64.493
Usuário maior chamador (U1) - "Ary_AntiPT"	1.325
Usuário mais chamado (U2) - "dilmabr"	46.185
Arestas ou Links	1,05E+06
Arestas Únicas Bidirecionais	928.095
Densidade da Rede	1,80E-05
Diâmetro da Rede	240.982
Distância geodésica média (não dirigido)	229,3023
Distância geodésica média (dirigido)	404,5793
Grau Médio	7,7030
Grau Máximo	1,34E+04
Distribuição de Graus - Média	7,49E-05
Distribuição de Graus - Máxima	0,5332

Fonte: SILVA, Rogério Olímpio da.

O tamanho total da rede importada de usuários é de 1.046.999 usuários em que temos 210.373 usuários U1 chamados ao menos uma vez e 64.493 usuários U2 chamados ao menos uma vez. O usuário com o maior número de citações é o “Ary_AntiPT” com 1.325 citações no período, e o usuário citado no maior número de vezes foi o usuário “dilmabr” com 46.185 citações ou referências.

A densidade da rede é de 18×10^{-5} , a qual representa uma densidade muito baixa, ou seja, os usuários possuem baixa interatividade recorrente, sendo uma rede pouco densa e sem muitas conexões entre si. O diâmetro máximo da rede é de 240.982 nós, os quais precisam de um número muito alto de conexões para que a informação propague por toda a rede e alcance as extremidades da rede. A quantidade de nós do diâmetro corresponde a 81,24% dos nós totais únicos da rede.

A distância média entre os nós de forma não-dirigida é de 229,30 nós, que representa uma quantidade baixa de usuários para que um segundo usuário da mesma rede possa vir a conhecer a informação. Já a distância média entre os nós de forma-dirigida é de 404,58 nós, que é uma quantidade alta de conexões para que um segundo nó tenha acesso à informação compartilhada. Apesar do Twitter ser uma rede dirigida, em que um usuário não tem necessidade de reciprocidade com seus seguidores, a informação aparece na linha do tempo do usuário seguido mesmo sem a reciprocidade com o seguidor. Sendo assim uma rede dirigida,

mas com acesso à informação de forma não-dirigida. Por esta razão se torna interessante observar ambas as distâncias médias.

A análise de graus em uma rede direcionada como o Twitter se baseia na quantidade de seguidores como grau de entrada (*in-degree*) que um usuário possui e na quantidade de seguidos como grau de saída (*out-degree*) que um usuário possui. Nesta rede de usuários em questão, serão considerados como grau de entrada a quantidade em que os usuários são chamados e mencionados, e como grau de saída a quantidade de chamadas a outros que os usuários possuem em média. Contando, portanto, como grau a soma do grau de entrada e de saída das postagens registradas nessa rede.

3.2 ANÁLISE DA REDE DE USUÁRIOS E POSTAGENS

Nesta seção vamos estudar as relações sobre as quantidades de notícias em que aparece o termo monitorado e o quanto a notícia permaneceu nas redes sociais. Além de apresentar o impacto que o alcance de um usuário que posta a notícia causa sobre a duração do rumor. Desta forma, podemos reunir informações que nos possibilitam entender melhor os dados empíricos.

Para a realização desta análise utilizamos a linguagem Python com as bibliotecas *networkx*, *panda* e *numpy*.

Abaixo detalhamos os dados da rede de usuários analisando as informações de rede, métricas e estatísticas.

Tabela 4 - Informações estatísticas da rede de usuários. Fonte: SILVA, Rogério Olímpio da

Usuários	Totais	Mínimo	Média	Máximo
Seguidores	598.302.518	0	2.017	9.291.153
In-degree (menções e retuites)	996.324	0	3	51.385
Out-degree	1.082.103	0	4	1.376
Retuites	932.881	0	3	2.193
Qtde de postagens	1.523.056	0	5	2.511
RTs por postagem	61,25%	0,00%	61,25%	87,34%
Seguidores que tiveram alguma ação de indegree	0,17%	0,00%	0,17%	0,55%
Seguidores que Retuitaram	0,16%	0,00%	0,16%	0,02%
Postagens por dia	95.191	-	0,32	156,94

Tabela 5 – Descrição das variáveis. Fonte: SILVA, Rogério Olímpio da

in-degree (menções e retuites)	Quantidade de retuites ou citações que outras pessoas fizeram desse usuário
out-degree	Quantidade de postagens em que houve uma citação ou retuite de outras pessoas
retuites	Quantidade de postagens do usuário por período
qtde postagens	Quantidade de postagens do usuário por período

Na tabela 4 podemos observar que a quantidade de out-degree é maior do que o in-degree, o que sugere que os usuários que postaram durante o período do experimento, na maioria das vezes, citaram outros mais do que foram citados. Ainda há os grandes canais de notícias que foram muito citados, tendo o número máximo de in-degree. E mesmo que uma pequena quantidade de seguidores teve alguma ação de in-degree (0,17%) e apenas uma pequena parcela de seguidores retuitou (0,16%), uma grande parte das postagens foi retuitada em algum momento (61,25%). Estando claro que alguns poucos usuários têm alto grau, e que suas postagens representam grande parte da rede. Na tabela 5 explicamos as colunas de grau da tabela 4.

Figura 5 - Frequência do percentual de retuites por seguidores. Fonte: SILVA, Rogério Olímpio da

A Figura 5 mostra que à medida que a taxa de retuites por seguidor aumenta, a frequência diminui. No entanto, na cauda aparece um pico contrariando a tendência. Isto ocorre porque existem poucos usuários que apesar de não terem muitos seguidores, eles mesmos promovem uma alta taxa de retuites (aproximadamente 100%).

Figura 6 - Frequência da quantidade de retuites por postagem. Fonte: SILVA, Rogério Olímpio da

A Figura 6 mostra que à medida que a taxa de retuites por postagem aumenta, a frequência também aumenta. Quanto mais retuites por postagem, maior é o engajamento dos usuários com postagens em geral.

Interessante observar que temos cerca de 5.000 postagens com 10% de retuites e temos 50.000 postagens com 100% de retuites. Acreditamos que isso deve ocorrer talvez pelo fato de fatores externos afetarem o Twitter, tais como programas de TV que pedem postagens, eventos esportivos e notícias que amplificam postagens. Refletindo o aspecto de que quando a rede está sendo muito utilizada temos maior número de retuites por postagens.

Figura 7 - Frequência da quantidade de postagens por dia. Fonte: SILVA, Rogério Olímpio da

A Figura 7 representa a distribuição do número de postagens por dia, de forma que possamos ver que a frequência decai com o número de postagens. Segundo os autores CHA, MEEYOUNG, et al. (2012), os usuários tendem a aprender que gerar uma grande quantidade

de postagens por dia torna seu conteúdo irrelevante e perde a atenção dos seguidores. O que pode ser uma possível explicação para esta distribuição.

Figura 8 - Frequência da distribuição de seguidores. Fonte: SILVA, Rogério Olímpio da

A Figura 8 mostra dois comportamentos ou tendências distintas. Ela é crescente atingindo o máximo em torno de 500 seguidores e decresce. Quando a quantidade de seguidores atinge o máximo, a frequência passa a ser descrita por uma lei de potência. Isto provavelmente é resultado da topologia da rede com ligação preferencial.

Interessante notar que para a região onde o número de seguidores está entre 1 seguidor e aproximadamente 500 seguidores, a frequência tem um comportamento inverso, ela cresce com o número de seguidores. Esta região mostra que é mais provável ter 500 seguidores do que um número inferior como 100 ou 10 seguidores.

Talvez seja o efeito de usuários que ingressaram recentemente e se encontram num momento transiente. Se deixarmos evoluir mais tempo, o comportamento pode convergir para ligação preferencial.

Figura 9 - Comparação de postagens com mesmo conteúdo com a duração da postagem pioneira em horas. Fonte: SILVA, Rogério Olímpio da

A Figura 9 traz o contraste entre as postagens pioneiras e postagens retuitadas (RTs), apresentando a duração da postagem pioneira² em horas e o número de postagens com o mesmo conteúdo da postagem pioneira (escala log x log).

Pode ser observado nessa Figura que as postagens pioneiras que duraram mais tempo foram as postagens que tiveram o maior número de outras postagens com o mesmo conteúdo. Principalmente no caso dos retuites, nos quais os retuites tiveram maiores postagens com o mesmo conteúdo, o que é verificável pelo pequeno aumento na variação da inclinação da reta dos retuites em relação às postagens comuns.

Podemos ver duas curvas “lei de potência” representadas na Figura nas cores amarelo e verde, com um leve aumento para a curva amarela do retuite que representa dados mais duradouros na rede. Sugerindo uma tendência de duração das postagens originadas com retuite na mesma distribuição das mensagens compartilhadas por retuites mas com proporções diferentes das mensagens não retuitadas que duram mais tempo do que as mensagens pioneiras, sugerindo uma tendência em que mensagens com mais de 100 ocorrências com o mesmo conteúdo e que estejam a mais de 60 horas e sejam compartilhadas tenham maior probabilidade de serem vistas pelos usuários, afinal as mensagens não duram muito na linha de tempo de um usuário e precisariam ser compartilhadas novamente para que o tempo de duração fosse estendido.

² Postagem pioneira: A primeira postagem a apresentar um conteúdo exclusivo e que será em seguida retuitada ou compartilhada por outros.

Vale explicar que as postagens com o mesmo conteúdo e que não são retuites são aquelas que foram copiadas de outros usuários ou usadas advindas de algum site ou ferramentas que possuem a funcionalidade de compartilhar na rede social Twitter.

Na Figura 10 pode ser observada a quantidade de postagens com o mesmo conteúdo versus a duração de 16 dias. Nesta Figura está sendo exibido o número de postagens originadas por agências de notícia, deixando em evidência que o volume das notícias é muito menor que o de postagens com mesmo conteúdo. Isso mostra a importância da rede social frente às agências de notícias no processo de propagação de rumores.

Figura 10 – Distribuição das quantidades de postagens com o mesmo conteúdo no período de 16 dias. Classificados por notícias (as postagens originadas de agências de notícias) que estão representadas pela cor vermelha e postagens comuns (postagens originadas por usuários comuns) na cor azul. Fonte: SILVA, Rogério Olímpio da

Dessa forma, podemos reforçar que os grandes canais de notícia estão se utilizando de interfaces para que os usuários postem seu conteúdo, como se fosse um conteúdo próprio do usuário, inclusive das próprias manchetes, fazendo referência ao site de origem na postagem. No caso dos rumores ou manchetes com conteúdo duvidoso ou ambíguo, o usuário irá postar o texto que leva em si o endereço do site que tem mais detalhes sobre o assunto, sem que o usuário seja o responsável pela informação.

Para o canal de notícia, o mais importante será o número de usuários postando seu conteúdo, e principalmente fazendo referência à origem da informação que está levando para

frente, e se em algum momento a notícia for questionada, negada ou sofrer retratação, o usuário não terá meios automáticos de corrigir o que foi postado e a notícia ou rumor já terá sido espalhado, inclusive até poderá ter sido retuitado novamente, levando a informação não confirmada para outras audiências.

Vale ressaltar que existe um pequeno *delay* de minutos ou horas entre as notícias postadas e os retuites pelos usuários comuns, no gráfico parecem ser simultâneos porque o gráfico só mostra agrupamentos diários. Porém, as postagens de notícias impulsionam as postagens normais em assuntos como o monitorado, principalmente em momentos de protestos políticos (BORONDO et al., 2012). Portanto, como nosso gráfico agrupa postagens por dia, temos que considerar o aspecto de que algumas postagens precederam e alavancaram postagens posteriores.

Figura 11 - Quantidade de postagens com mesmo conteúdo acumuladas e quantidade de retuites paralelos acumulados. Fonte: SILVA, Rogério Olímpio da

Na Figura 11 são apresentadas a quantidade acumulada de postagens com mesmo conteúdo à esquerda, e quantidade de retuites à direita, durante 16 dias.

Ao analisar esta Figura podemos observar que a quantidade de postagens é maior que a quantidade de retuites, mas positivamente correlacionados. Este resultado foi observado por FILOTAS, et al. (2014) e KWAK, HAEWOON, et al. (2010), que consideram os retuites como amplificadores das mensagens comuns.

Os seguidores podem retuitar a partir da mensagem em curso ou voltar na origem para compartilhar seu retuite. Como a quantidade de retuites é menor que a postagem acumulada, a

tendência é que estejam voltando à origem para retuitar. A origem pode ser o primeiro tuite ou a notícia que originou o tuite da manchete no site de notícias.

Figura 12 - Comparação do contador de usuários postando em relação à duração média das postagens em horas distribuídas na data da postagem pioneira. Fonte: SILVA, Rogério Olímpio da

Na Figura 12 está representada a comparação entre contador de usuários postando e a duração média em horas, distribuídos nas datas das postagens pioneiras com as quantidades totais de postagens dia-a-dia.

Fatos a serem considerados para o entendimento deste gráfico:

- O dia 15 foi o dia de protesto
- Duas semanas antes do protesto haviam muitos usuários postando o termo “Dilma”.
- No final de semana que antecede o protesto, o número de postagens atinge o valor máximo e amplifica a duração média das postagens.
- Embora diminua o número de usuários postando, os tempos de duração das postagens continuam maiores do que na semana anterior. Seria este um padrão?

Podemos observar que, aparentemente, a mudança no contador de usuários postando tem comportamento inverso às médias das quantidades de horas, na linha do primeiro dia de postagem, também podemos observar, que a maioria das postagens que duram mais tempo ocorrem nos finais de semana (MORALES, et al., 2009).

No período monitorado, apesar de ir aumentando o número de mensagens postadas a cada dia em média, as mensagens foram durando menos tempo, e isso não se refere ao fato do monitoramento parar no dia 16 e sim devido ao surgimento de novas mensagens ofuscarem as anteriores. Quanto maior o número de postagens em um período, menor será a atenção recebida por cada postagem, o que nos reforça a ideia da atenção compartilhada pelos usuários de redes sociais (CHA, MEEYOUNG, et al., 2012).

Podemos também considerar a parte superior do gráfico, na cor azul, onde estão as quantidades de postagens totais de cada dia para comparação do volume de usuários postando e a duração das mensagens, de forma que a quantidade total de postagens por dia segue um padrão similar à quantidade de usuários postando.

Vale lembrar que a quantidade de horas da duração das mensagens é maior no primeiro dia pelo fato de que, mesmo as mensagens que foram postadas antes do período de coleta, ao serem postadas novamente no primeiro dia serão consideradas como parte deste primeiro dia como sua primeira aparição, sendo considerada como mensagem pioneira entrando nos cálculos desde o primeiro dia.

Figura 13 - Relação das postagens e dos usuários que postaram as 20 postagens que mais duraram na rede social dentro do experimento. Fonte: SILVA, Rogério Olímpio da

Na Figura 13 temos a relação dos usuários que realizaram as 20 postagens que mais persistiram na rede, apresentando o nome do usuário, o número de seguidores, na coluna da esquerda as agências de notícias e na coluna da direita os usuários comuns, por fim a quantidade de horas de duração.

Analisando essa Figura podemos observar que em média as postagens originadas por agências de notícias persistem mais na rede do que as postagens comuns e como o número dos seguidores das agências de notícia é muito maior do que os usuários comuns, esse tempo de persistência maior é significativo. O número de seguidores é importante, mas não é determinante para a maior duração das postagens como podemos ver na Figura.

Ao analisar os canais de notícia nesta Figura observamos a média de 528,90 horas de duração média das notícias, e 484,90 horas a média dos usuários que não são agências de notícias. A confiabilidade das agências de notícia facilita a propagação de suas postagens, seja a informação confirmada ou rumor. Ainda que haja uma quantidade muito maior de postagens que não sejam advindas de agências de notícias no experimento como um todo, as postagens de agências de notícias tem durado muito mais em média e tem um alcance muito maior. Ainda sobre os canais de notícias, como observamos na Figura, não existe repetição de usuários que postaram entre os 20 principais, o que pode significar que os usuários ainda não descobriram quais elementos levam a ter uma postagem com grande propagação a ponto de obterem esse resultado.

Figura 14 - Quantidade de postagens pela quantidade de horas de duração das primeiras postagens, tanto para postagens originadas de agências de notícias (em vermelho) quanto para as de usuários comuns (em azul). Fonte: SILVA, Rogério Olímpio da

Na Figura 14 pode ser visto a distribuição das quantidades de postagens, tanto das originadas por agências de notícias, quanto pelas de usuários comuns, apresentadas num histograma das quantidades de horas que as primeiras postagens que foram compartilhadas duraram.

Podemos observar que a grande maioria das postagens duraram até 50 horas na rede, ocorrem rápido decaimento da frequência de postagem para durações que podem chegar a aproximadamente 400 horas.

As agências formais de notícias além de ter um grande alcance tendo de milhares a milhões de usuários que consomem seu conteúdo, também apresentam uma grande quantidade de seguidores que repassam e veiculam suas postagens, sendo elas notícias, manchetes ou rumores. Inclusive dias após a primeira postagem, havia pessoas engajadas em propagação de mensagens porque acreditam e até as que propagam contestando. Para o usuário gerador de grandes conteúdos até os questionamentos apresentam características positivas de visibilidade, porque o usuário que contesta necessita publicar novamente para contestar, mantendo a postagem com uma nova visibilidade, o que proporcionará que outros usuários tenham acesso a informação concordando ou discordando.

Outro fato interessante, aqui citado pela literatura do assunto, é que para que uma postagem fique ativa durante tanto tempo é necessário que outros usuários a mantenham na memória ativa e linha do tempo de outros usuários (GONÇALVES, et al., 2011), o que é chamado de notícia ou postagem persistente, e nesta Figura estamos apresentando algumas que ficaram muitos dias sendo repostadas, algumas chegando a 10 dias de novo compartilhamento.

Figura 15- Sequência de postagem de notícias com mesmo conteúdo, onde as setas representam a continuidade da notícia no tempo. Fonte: SILVA, Rogério Olímpio da

Na Figura 15, selecionamos as 4 postagens de notícias que tiveram mais repostagens como amostragem para analisar a sequência de postagens que as mesmas tiveram no dia-a-dia. Cada quadro é uma postagem com a informação de quantidades de postagem por dia e marcadores nas linhas que representam quantas vezes essas postagens foram repetidas nos demais dias.

Analisando os gráficos gerados pelos marcadores, podemos observar que as notícias tendem a iniciar num patamar e ir reduzindo e sendo esgotada a caminho de um formato de finalização como “cauda curta”, e exceto os casos em que há um ressurgimento, ela para de existir depois de alguns dias.

Caso ela ressurja e se mantenha persistente por mais tempo, significa que alguns usuários a repostaram iniciando um novo ciclo de vida da notícia, iniciando um novo processo de esgotamento em nova “cauda curta”.

Nesta Figura, apenas uma notícia teve uma “cauda longa”, o restante todo foi rápido esgotando-se em períodos curtos.

Figura 16 - Sequência de postagens comuns (não são notícias) com mesmo conteúdo, onde as setas representam a continuidade da postagem no tempo. Fonte: SILVA, Rogério Olímpio da

Na Figura 16 temos a mesma seleção que na Figura 15, mas para postagens que são de usuários comuns, extraindo uma amostragem das postagens que tiveram mais repostagens para

analisar a sequência de dia-a-dia. As apresentações gráficas das sequências das postagens que não foram postadas por agências de notícia são bem diferentes das de notícias, apresentando variações e dificultando que se encontre similaridade. Algo que podemos observar é que estas postagens apresentaram alguns picos em sua maioria, ao invés de só um pico, o que mostra que outros usuários retomam a postagem reativando sua visibilidade novamente. O que é muito comum de acontecer quando a postagem inicial é gerada pelo compartilhar informações de um site ou matéria, e até mesmo notícias ou rumores, porque o usuário que está postando algo está contando com a ajuda de um mecanismo automático, e quando o assunto ressurgir, é novamente citado na mídia, as pessoas voltam a reler as mesmas matérias, e acabam compartilhando novamente o assunto que não é mais novidade.

Tabela 6 - Tabela com detalhamento dos valores das Figuras 15 e 16.

Postagem	Agência (A) / Usuário (U)	Duração Em Horas	Quantas Postagens (U)	Quantas Postagens RTs	RT/T	Quantos Usuários	U/T	Soma de Seguidores dos Usuários	Quantas com Usuários Repetidos	Quantas com Usuário de Origem
A	A	383	1432	1294	90%	1398	98%	3,7M	34	2
B	A	248	548	363	66%	490	89%	13,9M	58	9
C	A	255	207	93	45%	197	95%	5,4M	10	1
D	A	383	377	206	55%	334	89%	5,6M	43	1
E	U	191	1210	489	40%	604	50%	7M	606	1
F	U	383	1415	1407	99%	1408	100%	318K	7	1
G	U	196	950	465	49%	908	96%	1,7M	42	1
H	U	383	1650	616	37%	656	40%	3M	994	1

Ao analisar com maior profundidade os dados das Figuras 15 e 16 que possuem as postagens com maiores durações de persistência na rede, podemos observar que o percentual de retuites é muito alto, tendo uma grande quantidade de retuites para as postagens que mais duraram na rede. Outro aspecto interessante é que a quantidade de usuários é quase a mesma da quantidade de postagens para o caso das agências de notícia, tendo poucos usuários que compartilharam mais de uma vez as mesmas postagens, o que é diferente para o caso das

postagens que não são originadas pelas agências de notícias, que já tem uma característica de mais usuários compartilharem as mesmas publicações.

Uma outra informação que merece atenção é a quantidade de seguidores acumulado por cada postagem, em que foram calculados a soma dos seguidores de todos usuários que compartilharam a publicação, medindo-se a influência total da postagem inicial com seus compartilhamentos e retuites. Demonstrando que após alguns poucos retuites as postagens alcançam milhões de usuários (AARTS, OLAV, et al., 2012), sendo o retuite responsável pelo maior alcance das publicações (LIU et al., 2011).

O esforço de alguns usuários para que uma postagem tenha maior alcance também chama a atenção, ao observarmos que alguns usuários repostaram ou retuitaram a mesma postagem alguns meses, promovendo a postagem para sua rede mais uma vez a cada retuite. Dois casos dos quatro de notícias apresentam a característica da agência retuitar sua própria publicação, dobrando o alcance de sua própria publicação e se autopromovendo, o que não é uma característica comum pelas agências de notícias.

As postagens de retuites duram mais do que as postagens compartilhadas sem retuite, sendo que as originadas por agências de notícias duram ainda mais do que as de usuários comuns. O volume de postagens por usuários comuns é muito maior do que das agências de notícias, o que mostra a importância das redes sociais na propagação de informações. Um fato curioso é que o retuite leva muitos usuários a retornarem na fonte (Figura 13). A duração das postagens não tem dependência completa do número de seguidores (Figura 15), embora a média de duração das postagens das agências de notícias dure mais do que as postagens comuns, apesar de ter um volume menor.

Baseado nos dados empíricos será construído um modelo de simulação com o objetivo de reproduzir características importantes dos dados encontrados nos dados empíricos, dados os quais servirão para calibrar o modelo de simulação.

4 MODELO E SIMULAÇÃO

Alguns trabalhos têm utilizado a modelagem baseada em agentes para simular e analisar o comportamento de usuários no Twitter, bem como suas interações e propagações. Nos trabalhos de LIU et al. (2011) e DOERR, et al. (2012) foram propostas redes de Livre de Escala para estudar como as postagens se propagam e foram analisadas as persistências das mensagens. A partir destes trabalhos, propomos um modelo para simular postagens de mensagens em uma rede de Livre de Escala, direcionada. Esta rede simulada foi calibrada utilizando as informações conhecidas pela análise dos dados empíricos apresentada no capítulo anterior.

Neste trabalho, como uma forma de simplificação, a descrição do algoritmo foi dividida em blocos. No primeiro bloco temos a formação da rede de Livre de Escala com agentes que apresentam características de ligação preferencial, no segundo bloco a definição de quem seriam os usuários postando na rodada de simulação, de forma que cada seguidor destes usuários possa ter acesso à postagem. No terceiro bloco, se os seguidores estiverem online, podem ler a postagem e decidem se vão fazer ou não um retuite, simulando desta forma os retuites consecutivamente até o final de todas as postagens.

O modelo permite que uma postagem originada por um usuário pode vir novamente para o mesmo usuário em uma rodada posterior da simulação, causando um efeito de retroalimentação da informação. Outro aspecto que vale salientar, é que para manter as características de ligação preferencial, os usuários tendem a seguir os usuários com maior número de seguidores (BARABÁSI, et al., 2014).

Figura 17 - Processo de simulação de modelagem por agentes. Fonte: SILVA, Rogério Olímpio da

Na Figura 17 apresentamos a sequência das etapas de construção do modelo. Tendo em vista que o foco desta simulação está na avaliação das características da rede e das durações das postagens, não consideraremos outras variáveis como horário, dia da semana, credibilidade do usuário de origem e apelo emotivo da postagem, para manter o modelo o mais simples possível.

Com a rede gerada, é realizada uma iteração entre todos os nós para saber quem irá postar nesta rodada. Cada postagem de cada agente é fornecida ou tornada possível de ser visualizada por todos os seguidores de quem fez a postagem. Então para cada seguidor um novo valor aleatório é simulado, se o valor estiver dentro do percentual acontece o retuite e a postagem ganha mais uma possibilidade de propagação na mesma rodada, e isso segue até que todos itens da lista de postagens programadas sejam realizados. Quando encerradas as postagens programadas inclusive com os retuites que foram adicionados ao final da lista, a simulação entra nas fases de cálculo das métricas de postagens, durações e quantidades finais pelos agentes.

4.1 MODELO

O modelo de simulação foi definido da seguinte forma:

- 1) Inicialização das bibliotecas Python e variáveis:
 - a) Bibliotecas: networkx, numpy, pandas
 - b) p_{postar} = probabilidade de um usuário realizar uma postagem
 - c) p_{retuitar} = probabilidade de um usuário retuitar uma postagem
- 2) Parâmetros da rede:
 - a) N = tamanho da Rede
 - b) Gama = Probabilidade de adicionar um novo nó conectado a um nó existente escolhido aleatoriamente de acordo com a distribuição out-degree.
 - c) Alpha: Probabilidade de adicionar um novo nó conectado a um nó existente escolhido aleatoriamente de acordo com a distribuição in-degree.
 - d) Beta: Probabilidade de adicionar uma conexão entre dois nós existentes. Um nó existente é escolhido aleatoriamente de acordo com a distribuição in-degree e o outro escolhido aleatoriamente de acordo com a distribuição out-degree.
 - e) Δ_{in} : Tendência para escolher nós de distribuição in-degree.
 - f) Δ_{out} : Tendência para escolher nós de distribuição out-degree.
- 3) Gerar rede Livre de Escala com os parâmetros definidos acima.
- 4) Dinâmica:
 - a) Escolher um nó da rede aleatoriamente. Caso tenha atingido o limite, ir para (4.c).
 - b) Decidir com probabilidade p_{postar} se ele vai realizar uma postagem
 - i) Se não, voltar ao passo (4.a) para escolher outro nó.
 - ii) Se sim, percorrer a lista de seguidores e com probabilidade p_{retuitar} decidir se seguidor vai retuitar. Se sim, realizar retuite e continuar a lista de seguidores. Ao final, voltar para (4.a).
 - c) Imprimir quantidade de postagem e seus respectivos retuites

A geração da rede na simulação foi realizada através do uso da função “scale_free_graph” da biblioteca networkx. Esta função recebe parâmetros de característica e tamanho da rede e retorna uma rede Livre de Escala direcionada³. A base conceitual para esta função está no trabalho escrito por BOLLOBÁS et al. (2003).

³ Os parâmetros de geração da rede podem ser vistos no item acima 4.1 no parágrafo 2.

O código fonte em Python para a construção desta simulação pode ser encontrado no Apêndice F.

4.2 RESULTADOS DA SIMULAÇÃO

Na Figura 18 abaixo apresentamos a distribuição de nós da rede empírica na escala logxlog. Podemos observar que a distribuição de graus in-degree e out-degree tem inclinações diferentes. A curva in-degree é mais inclinada que a out-degree. Isso significa que existe uma assimetria entre número de seguidos e seguidores.

Figura 18 - Distribuição de graus dos dados da rede empírica. Fonte: SILVA, Rogério Olímpio

Figura 19 - Distribuição de graus dos dados da Rede Simulada. Fonte: SILVA, Rogério Olímpio

Na Figura 19 apresentamos os dados da rede simulada para comparação com os dados empíricos. Podemos observar a assimetria nas curvas in-degree e out-degree, como apresentado na rede empírica (Figura 18). Isso reforça nossa confiança no modelo gerador de rede, que foi devidamente calibrado.

Figura 20 - Comparação entre dados empíricos e simulados considerando quantidade de postagens e quantidades de usuários. Fonte: SILVA, Rogério Olímpio da

Na Figura 20 temos a comparação entre dados empíricos (acima) e dados simulados (abaixo), no lado esquerdo a análise de quantidade de postagens por usuários, e no lado direito, na escala log x log, a quantidade de usuários por quantidade de postagens.

Podemos observar as similaridades de distribuição entre os dados empíricos e os dados simulados, e guardadas as devidas proporções, podemos inclusive identificar o comportamento de ligação preferencial com a preferência clara de seguidores optando por seguir os usuários com maior grau e número de seguidores. Também pode ser vista similaridade nos gráficos de usuários por quantidade de postagens, tendo tanto no de dados empíricos quanto no de dados simulados curvas de potência semelhantes.

4.3 DISCUSSÃO

Ao analisar as referências nos trabalhos de LIU et al. (2011) e DOERR et al. (2012) em relação aos nossos resultados, identificamos questões similares no que se refere à persistência das mensagens e da assimetria de seguidores e seguidos, nos remetendo aos seguintes pontos de discussão:

LIU et al. (2011) ao analisar o modelo clássico de rumor dizem que não consideram o grau de origem do gerador do rumor, estendendo este modelo considerando a importância das conexões entre as pessoas. Pelas características do Twitter de ser unidirecional e que se utiliza de ligação preferencial, o Twitter teria características de ser mais "um canal para expressar" do que "um canal para ouvir", isso talvez explicaria porque as celebridades de redes sociais seguem poucas pessoas. Da mesma forma, em nosso trabalho identificamos não ter importância definitiva o número de seguidores, e observamos que as características de ligação preferencial e o fato de ser unidirecional sem necessidade de reciprocidade serem aspectos principais e marcantes da rede de usuários do Twitter.

Também foi dito por LIU et al. (2011) que uma evolução importante seria testar o comportamento do quanto uma pessoa acreditaria em uma postagem de outra pessoa observando os dados reais do Twitter. O que parcialmente já estamos realizando em nosso trabalho, ao comparar o engajamento de alguns usuários seguidores e o quanto a rede fica 'aquecida' por receber influências externas. Vale frisar que nossa simulação foi calibrada com dados reais empíricos.

DOERR, et al. (2012) mencionam que as redes sociais foram estruturadas para que seu crescimento orgânico facilitasse a passagem de notícias e rumores. Ainda que não haja organização central ou desenho prévio da rede. Talvez essa seja a razão porque, em nosso trabalho, constatamos que é fácil conseguir algumas dezenas de seguidores quando se é novo na rede, mas difícil conseguir mais ao atingir os milhares. E que manchetes, independente da veracidade da informação, recebem mais atenção e retuites do que postagens que não venham de agências de notícia.

Na simulação de DOERR, et al. (2012), foi observado que pequenos *nós* repassam informações de rumor assim que as recebem, que os grandes hubs se comunicam com outros grandes de forma facilitada e mantém a informação acessível a um grande número de pessoas, e por fim, usuários médios são importantes para transmitir aos seus vizinhos. Em nosso trabalho identificamos que poucos usuários têm muitos retuites por seguidores, poucos possuem alta taxa de retuites por serem "seguidores fiéis", e quanto mais retuites uma postagem tem maior é o engajamento de outros usuários com aquela postagem.

5 CONSIDERAÇÕES FINAIS

Analizamos a propagação de notícias no Twitter, utilizando o termo “Dilma” no período de 01 de março de 2015 a 16 de março de 2015, observando o comportamento das postagens e as características da rede de usuários. Inclusive com simulação da rede e da dinâmica de postagens.

Ao analisar as postagens de usuários no Twitter, quando uma postagem é retuitada por um usuário com mais de 10 mil seguidores de alcance, ela é disponibilizada com outros agrupamentos de usuários independentemente da quantidade de seguidores do usuário originador da mensagem.

Quando um usuário ganha notoriedade ele atrai seguidores formando um agrupamento ao seu entorno, o que se verifica ao analisarmos a quantidade de postagens retuitadas destes usuários, ou mensagens com o conteúdo similar reaproveitando o conteúdo das mensagens iniciais. Assim as mensagens persistem ativas na rede por horas e até dias. Podemos assim observar o impacto inter-relacionado do alcance dos usuários, com a quantidade de postagens publicadas, e com o quanto uma notícia, manchete ou rumor serão possivelmente compartilhados pela audiência desse usuário.

O poder de propagação de uma postagem por uma celebridade, independente da veracidade ou ambiguidade da informação, é muito forte, principalmente porque após o primeiro retuite é muito provável que a postagem atinja um alto nível de alcance até fora da rede da celebridade. Independente se é uma notícia confirmada, uma manchete dúbia ou um rumor, o alcance do usuário se torna importante tanto para ser assunto entre os que concordam com ele quanto com os que não concordam, mas que ao referenciá-lo vão levar indiretamente a mensagem dele para outras pessoas (AARTS, OLAV, et al., 2012), (vide Figura 16).

As pessoas têm capacidade finita de processamento de postagens. Isso limita o tempo em que uma informação estará disponível para ser vista por um usuário (GONÇALVES, et al., 2011). Assim uma menor quantidade de postagens proporciona que as mesmas sejam consideradas mais importantes pelos usuários que as recebem, facilitando inclusive a permanência das mesmas nas telas dos usuários. (CHA, MEEYOUNG, et al., 2012).

Através de modelagem por agentes, pudemos reproduzir com simulação um comportamento similar aos dados empíricos. Um dos resultados observados, é que o número de seguidores de um usuário não é o único fator decisivo para garantir que uma mensagem perdure na rede. Por outro lado, um diferencial foi o fato dos usuários com menor número de

seguidores e que retuitam a postagem para outros agrupamentos, causando o acesso à postagem possível para outras audiências (AARTS, OLAV, et al., 2012).

Conforme também observado na análise das Figuras, existem diferenças nas durações das postagens originadas por usuários geradores de notícias em relação às postagens originadas por usuários não geradores de notícias. As principais diferenças estão nos fatos de que as notícias, em média, duram mais do que as postagens normais, e a cauda (decrecente após o ápice) das notícias é mais persistente no tempo. Como já dito no decorrer desta dissertação, a rede social Twitter tem todas as características de uma rede Livre de Escala, mas o principal aspecto é que as pessoas têm a tendência de seguir usuários com muitos seguidores, sem se preocupar com reciprocidade no relacionamento.

Um aspecto importante a se observar é o fato de que o ciclo de vida de uma notícia pode causar um ruído nas redes sociais durante todo o tempo em que a notícia persiste vigente, e sua vigência pode acabar quando surge uma outra notícia ou naturalmente ao cair no esquecimento.

Conforme podemos observar na análise dos dados empíricos, postagens de um mesmo usuário passam por diferentes seguidores. Logo uma vez postada uma mensagem por usuários com alto alcance, independente da veracidade ou não das informações, ela atinge um largo alcance e em alguns casos, por meio de retuites, perdura por um longo período de tempo. Algumas postagens até receberam novo período de evidência ao serem retuitadas depois de alguns dias, reiniciando o ciclo de vida da postagem em um processo de cauda longa como já dito anteriormente na análise das Figuras.

Sobre o aspecto da autopromoção de usuários que retuitam suas postagens e de usuários que compartilham mais de uma vez conteúdos que querem disseminar, multiplicam na rede o alcance de algumas postagens, mantendo-as persistentes na rede por muito mais tempo. Desta forma, milhões de seguidores terão acesso à publicação após alguns retuites. Quando uma agência de notícias faz este novo compartilhamento, multiplica o alcance de sua própria postagem, possibilitando que sua rede de seguidores tenham uma nova chance de ver o mesmo conteúdo em suas linhas do tempo e assim abrindo novas possibilidades de leituras e publicações.

Ainda sobre os dados empíricos, no caso de rumores, manchetes dúbias ou notícias não confirmadas, os usuários irão ler e retuitar, levando para novas audiências. Caso o usuário de origem da postagem pioneira publique uma retratação, correção ou desminta a informação dita, é praticamente impossível que toda audiência da informação inicial seja alcançada pela informação atualizada. Portanto, os rumores na rede social têm a tendência de se espalhar

fortemente sem mecanismo de correção ou atualização, principalmente pelo fato de que uma nova postagem de retratação não alcançaria o mesmo conjunto de seguidores da postagem inicial.

Através da identificação do modelo de simulação pudemos entender o comportamento das postagens e repostagens, de forma que identificamos o valor do retuite dentro da estrutura da rede do Twitter. Fica claro que a cada rodada de retuite a postagem tem nova sobrevida e possível acesso a outros usuários. Pela característica de ligação preferencial, inclusive na simulação, o agrupamento de usuários ao redor de um usuário com alto grau possibilita que a publicação tenha um bom alcance logo após a postagem, porém ficou evidente o valor dos usuários com menor alcance que fazem a ligação entre diferentes agrupamentos, os quais quando retuitam uma postagem, possibilitam que ela seja acessível a outros usuários.

Portanto para que uma postagem perdure por muito tempo na rede, ação rara pela baixa probabilidade de retuites e conforme reproduzido na simulação, é necessária uma combinação de postagem e retuites de usuários com alto alcance e de usuários de baixo alcance que façam parte de outros agrupamentos e clusters.

Possíveis extensões deste trabalho poderiam abordar dinâmicas de opiniões na rede Twitter, desenvolver uma máquina de aprendizagem, que lê postagens online e aprende informações dos parâmetros que influenciam na propagação, e a utilização de aprendizagem profunda (*Deep Learning*).

REFERÊNCIAS

- AARTS, Olav, et al. **"Online social behavior in twitter: A literature review."** 2012 IEEE 12th International Conference on Data Mining Workshops. IEEE, 2012.
- AIRSTRIPO.CO. Disponível em: <<http://airstrip.com.br>> Acesso em: julho de 2016.
- BARABÁSI, Albert-László, and Réka Albert. **"Emergence of scaling in random networks."** science 286.5439 (1999): 509-512.
- BARABÁSI, Albert-László. **"The architecture of complexity."** Control Systems, IEEE 27.4 (2007): 33-42.
- BARABÁSI, Albert-László; ALBERT, Réka; JEONG, Hawoong. **Scale-free characteristics of random networks: the topology of the world-wide web.** Physica A: Statistical Mechanics and its Applications, v. 281, n. 1, p. 69-77, 2000.
- BARABÁSI, Albert-László; BONABEAU, Eric. **Scale-free.** Scientific American, 2003.
- BARABÁSI, Albert-László; LIU, Yang-Yu; SLOTTINE, Jean-Jacques. **Control centrality and hierarchical structure in complex networks.** Plos one, v. 7, n. 9, p. e44459, 2012.
- BASSETT e Bullmore, 2006; disponível em: <https://openi.nlm.nih.gov/detailedresult.php>. Acesso em: julho de 2016.
- BERTELLE, Cyrille, Gérard HE Duchamp, and Hakima Kadri-Dahmani, eds. **Complex systems and self-organization modelling.** Springer Science & Business Media, 2008.
- BOCCARA, Nino. **Modeling Complex Systems.** 2nd ed. New York: Springer Science and Business Media, 2010. 489 p.
- BOLLOBÁS, B., Riordan, O., Spencer, J., and Tusnády, G. **The degree sequence of a scale-free random graph process.** Random Structures & Algorithms 18, 3 (2001), 279–290.
- BOLLOBÁS, Béla et al. Directed scale-free graphs. In: **Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms.** Society for Industrial and Applied Mathematics, 2003. p. 132-139.
- BOYD, Danah, Scott Golder, and Gilad Lotan. **"Tweet, tweet, retweet: Conversational aspects of retweeting on twitter."** System Sciences (HICSS), 2010 43rd Hawaii International Conference on. IEEE, 2010.
- CANTRIL, Hadley. **The invasion from Mars: A study in the psychology of panic.** Transaction Publishers, 1940.
- CHA, Meeyoung, et al. **"The world of connections and information flow in twitter."** Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 42.4 (2012): 991-998.

comScore, Inc. Disponível em: <<https://www.comscore.com/por/Insights/Press-Releases/2014/5/Estudo-da-comScore-Brazil-Digital-Future-in-Focus-2014-esta-disponivel>> Acesso em: julho de 2016.

DOERR, Benjamin, Mahmoud Fouz, and Tobias Friedrich. "Why rumors spread so quickly in social networks." *Communications of the ACM* 55.6 (2012): 70-75.

DUNBAR, R. I. "**The social brain hypothesis.**" *brain* 9.10 (1998): 178-190.

DUNBAR, Robin IM. "**Neocortex size as a constraint on group size in primates.**" *Journal of Human Evolution* 22.6 (1992): 469-493.

ÉRDI, P. **Complexity explained.** Springer Science & Business Media, 2007.

FILOTAS, E., et al. **Viewing forests through the lens of complex systems science.** *Ecosphere*, v. 5, n. 1, p. art1, 2014.

GOLDENBERG, Jacob, Barak Libai, and Eitan Muller. "**Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata.**" *Academy of Marketing Science Review* 2001 (2001): 1.

GONÇALVES, Bruno, Nicola Perra, and Alessandro Vespignani. "**Modeling users' activity on twitter networks: Validation of dunbar's number.**" *Plos One* 6.8 (2011): e22656.

GRABOWICZ, Przemyslaw A., et al. "**Social features of online networks: The strength of intermediary ties in online social media.**" *PloS one* 7.1 (2012): e29358.

GRANOVETTER, Mark. "**Threshold models of collective behavior.**" *American journal of sociology* (1978): 1420-1443.

HUBERMAN, Bernardo A., Daniel M. Romero, and Fang Wu. "**Social networks that matter: Twitter under the microscope.**" Available at SSRN 1313405 (2008).

J. BORONDO, A. J. Morales, J. C. Losada, and R. M. Benito, "**Characterizing and modeling an electoral campaign in the context of Twitter: 2011 Spanish Presidential election as a case study**", *Chaos*, vol. 22, n. 2, p. 023138, Jun 2012.

JANSEN, Bernard J., et al. "**Micro-blogging as online word of mouth branding.**" *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2009.

KATZ, Elihu, and Paul Felix Lazarsfeld. **Personal Influence, the part played by people in the flow of mass communications.** Transaction Publishers, 1966.

KEMPE, David, Jon Kleinberg, and Éva Tardos. "**Influential nodes in a diffusion model for social networks.**" *International Colloquium on Automata, Languages, and Programming*. Springer Berlin Heidelberg, 2005.

KEMPE, David, Jon Kleinberg, and Éva Tardos. **"Maximizing the spread of influence through a social network."** Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2003. the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2003.

KOSTKA, Jan, Yvonne Anne Oswald, and Roger Wattenhofer. **"Word of mouth: Rumor dissemination in social networks."** International Colloquium on Structural Information and Communication Complexity. Springer Berlin Heidelberg, 2008.

KWAK, Haewoon, et al. **"What is Twitter, a social network or a news media?"** Proceedings of the 19th international conference on World wide web. ACM, 2010.

KWON, Sejeong, et al. **"Aspects of rumor spreading on a microblog network."** International Conference on Social Informatics. Springer International Publishing, 2013.

KWON, Sejeong, et al. **"Prominent features of rumor propagation in online social media."** 2013 IEEE 13th International Conference on Data Mining. IEEE, 2013.

L. FESTINGER, D. Cartwright, K. Barber, J. Fleischi, J. Gottsdanker, A. Keysen, and G. Leavitt. **A study of a rumor: its origin and spread.** Human Relations, 1948.

LEITE, M. S. A. **Proposta de uma modelagem de referência para representar sistemas complexos.** Tese de doutorado. Universidade Federal de Santa Catarina. 2004.

LERMAN, Kristina, and Rumi Ghosh. **"Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks."** ICWSM 10 (2010): 90-97.

LESKOVEC, Jure, Lars Backstrom, and Jon Kleinberg. **"Meme-tracking and the dynamics of the news cycle."** Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009.

LIU, Dechun, and Xi Chen. **"Rumor Propagation in Online Social Networks Like Twitter-A Simulation Study."** 2011 Third International Conference on Multimedia Information Networking and Security. IEEE, 2011.

MENDES, J. **"Física de redes complexas."** Gazeta de Física 28.4 (2005): 10-16.

MILLER, John H.; PAGE, Scott E. **Complex adaptive systems: an introduction to computational models of social life:** Princeton: Princeton university press, 2009. 271 p.

MITCHELL, Melanie. **Complexity: A Guided Tour.** New York: Oxford university press Inc, 2009.

MORALES, A. J., J. C. Losada, and R. M. Benito. **"Users structure and behavior on an online social network during a political protest."** Physica A: Statistical Mechanics and its Applications 391.21 (2012): 5244-5253.

NEWMAN, Mark EJ. **"The structure and function of complex networks."** SIAM review 45.2 (2003): 167-256.

PETERSON, Warren A., and Noel P. Gist. **"Rumor and public opinion."** American Journal of Sociology (1951): 159-167.

PFEIFFER III, Joseph J., and Jennifer Neville. **"Methods to determine node centrality and clustering in graphs with uncertain structure."** arXiv preprint arXiv:1104.0319 (2011).

ROGERS, Everett M. **Diffusion of innovations.** Simon and Schuster, 2010.

ROSENMAN, Evan TR. **Retweets—but not just retweets: Quantifying and predicting influence on twitter.** Diss. Bachelor's thesis, applied mathematics. Harvard College, Cambridge, 2012.

ROSENMAN, Evan TR. Retweets—but not just retweets: **Quantifying and predicting influence on twitter. Diss. Bachelor's thesis, applied mathematics.** Harvard College, Cambridge, 2012.

SCHELLING, Thomas C. **Micromotives and macrobehavior.** WW Norton & Company, 2006.

SEO et al (2013). Disponível em <<https://www.spandidos-publications.com/ijo/43/6/1737>>. Acesso em: julho de 2016.

SUH, Bongwon, et al. **"Want to be retweeted? large scale analytics on factors impacting retweet in twitter network."** Social computing (socialcom), 2010 ieee second international conference on. IEEE, 2010.

Tecmundo "História das Redes Sociais - Como Tudo Começou". Disponível em: <<http://www.tecmundo.com.br/redes-sociais/33036-a-historia-das-redes-sociais-como-tudo-comecou.ht>> Acesso em: julho de 2016.

Tecmundo "História das Redes Sociais". Disponível em: <<http://www.techtodo.com.br/artigos/noticia/2012/07/historia-das-redes-sociais.html>> Acesso em: julho de 2016.

Tibco Spotfire. Disponível em <<http://spotfire.tibco.com>> Acesso em: julho de 2016.

Twitter About. Disponível em: <<https://about.twitter.com/company>> Acesso em: julho de 2016.

Twitter Blog. Disponível em: <<https://blog.twitter.com>> Acesso em: julho de 2016.

Twitter Support. Disponível em: <<https://support.twitter.com>> Acesso em: julho de 2016.

VON FERBER, C., Yu Holovatch, and V. Palchykov. **"Scaling in public transport networks."** arXiv preprint cond-mat/0501296 (2005).

WATTS, Duncan J., and Peter Sheridan Dodds. **"Influentials, networks, and public opinion formation."** Journal of consumer research 34.4 (2007): 441-458.

WATTS, Duncan J., and Steven H. Strogatz. **"Collective dynamics of 'small-world' networks."** nature 393.6684 (1998): 440-442.

Y. MATSUBARA, Y. Sakurai, B. Prakash, L. Li, and C. Faloutsos. **Rise and fall patterns of information diffusion: Model and implications.** In KDD, 2012.

YANG, Jaewon, and Jure Leskovec. **"Patterns of temporal variation in online media."** Proceedings of the fourth ACM international conference on Web search and data mining. ACM, 2011.

ZHAO, Dejin, and Mary Beth Rosson. **"How and why people Twitter: the role that micro-blogging plays in informal communication at work."** Proceedings of the ACM 2009 international conference on Supporting group work. ACM, 2009.

APÊNDICE A – REDES COMPLEXAS

Há duas definições simplificadas mais utilizadas para descrever redes complexas, a de BARABÁSI e ALBERT-LÁSZLÓ em 2007 em que dizem que uma rede complexa é uma topologia que apresenta uma estrutura não-trivial de conexão, composta por um conjunto elevado de nós conectados entre si. E a definição de MITCHELL em 2009 que diz que uma rede é uma coleção de nós conectados por links, e os *nós* correspondem aos indivíduos e os links correspondem às conexões entre eles. O estudo de redes e mais especificamente de redes complexas vem sendo aplicada em diversas áreas, se tornando uma forma comum de apresentação de comunicação de pontos dentro de um sistema, seja tanto na natureza quanto em construções modernas como ciência e tecnologia.

BARABÁSI et al. (2012) ainda observam que na estrutura de todo sistema complexo há uma rede complexa que codifica as interações entre seus componentes. Fazendo parte de nosso cotidiano, em serviços de conectividade como o Google, Facebook, Waze e Twitter, as redes sociais apresentam todas as características de redes complexas.

Ao longo deste trabalho será utilizada a convenção da Ciência de Redes, em que os usuários serão chamados de nós ou vértices e as conexões entre eles de links ou arestas. Por ser uma área da ciência crescente em novas pesquisas e descobertas, de acordo com NEWMAN (2003), há ainda muitos espaços que serão preenchidos por novos conhecimentos que afetarão em muito os sistemas de redes.

TIPOS DE REDE

Os dois tipos de redes mais próximas da que estamos estudando neste projeto são as redes Mundo Pequeno (Small World) e Livre de Escala (Scale-Free). Em uma rede Mundo Pequeno, cada nó é ligado inicialmente a um conjunto de nós vizinhos alcançando qualquer outro nó por poucas arestas, ainda que não seja seu vizinho imediato. Essa forma de organização foi proposta por WATTS e STROGATZ (1998) como um modelo matemático capaz de gerar uma rede aleatória com características de Mundo Pequeno, formando uma rede com estrutura de anel (lattice). Nas redes Mundo Pequeno, os *nós* têm características similares, não há celebridades (hubs) com maior quantidade de conexões se destacando da média, e não há muita diferença dos graus de centralidade. Outra característica marcante é que o caminho médio é baixo entre os nós.

Em uma rede aleatória composta exclusivamente por conexões locais com seus vizinhos, é necessário um grande número de passos para que a informação trafegue e chegue ao destino. Já na rede Mundo Pequeno, algumas dessas conexões locais são removidas aleatoriamente e substituídas por conexões para qualquer nó no sistema, tendendo a ser conectados localmente, possibilitando o acesso entre dois nós de forma rápida. Logo, apesar de existirem conexões de longa distância, agem como atalhos, de forma que a informação pode ser transmitida entre quaisquer dois nós por um pequeno número de conexões (MILLER e PAGE, 2009). O modelo parte de uma rede em forma de anel, respeitando o número de nós N , de forma que cada nó é interligado aos nós vizinhos mais próximos por meio de arestas K . Cada uma das arestas criadas sofre uma reconexão com probabilidade P , ligando o nó de origem a um novo destino.

Figura 21 - Forma como a probabilidade P afeta a existência ou não da rede mundo pequeno. Fonte: BASSETT e BULLMORE (2006)

Quanto mais o valor de probabilidade P se aproxima de 1, mais próxima a rede fica de uma rede aleatória. É uma crítica a este modelo indicado, que ele não captura ou descreve comportamentos semelhantes às redes sociais. Para contemplar as características de redes sociais foram estudados outros modelos, dos quais o mais celebrado é o chamado Livre de Escala. Que vamos discutir na próxima sessão.

LIVRE DE ESCALA (SCALE-FREE)

Um dos principais modelos é o Livre de Escala (Scale-Free), em que podem ser observados comportamentos Livre de Escala $P(k)$ com distribuição dos nós em lei de potência em diferentes ordens de grandeza (FERBER, et al., 2005) (BARABÁSI, et al., 2003). Nessa

topologia temos a característica de distribuição de graus com ligação preferencial (descrito com maiores detalhes no item 2.3 deste trabalho).

Normalmente não são observadas estruturas combinadas ou de distribuição uniforme dos relacionamentos entre os nós, uma propriedade comum é a forte dependência de agrupamentos (*cluster*) com distribuição exponencial em alguns casos (BARABÁSI, et al., 2014).

Figura 22 - Podemos verificar a diferença de topologia de uma rede Livre de Escala e uma rede aleatória.

Fonte: SEO et al (2013).

MÉTRICAS DE REDES

Para analisar os usuários do experimento e suas conexões se faz necessário a composição de uma rede, em que podemos aplicar métricas que nos permitem entender com maior profundidade as características de conectividade e as interações entre os usuários.

A rede que é composta pelos usuários, ou agentes do experimento, pode ser representada por $G = (V, E)$, onde V é o conjunto de $|V|$ nós, e E é o conjunto de arestas que conectam dois elementos de V . Uma aresta específica $(v_i, v_j) \in E$ representa uma parceria estabelecida entre dois agentes representados pelos nós v_i e v_j ($v_i, v_j \in V$) para a transmissão de dados de v_i para v_j .

Grau: O grau g de um vértice $v_i(t)$ no momento t é definido pela soma do grau de entrada e saída de $v_i(t)$. O grau de entrada de v_i é o número de v para os quais ele aponta, enquanto o grau de saída é o número de v que apontam para v_i . O grau de um vértice é uma das propriedades estruturais mais básicas que avalia o número de arestas adjacentes. (LERMAN et al., 2010).

Grau Máximo: Por definição o grau máximo de uma rede G no momento t é dado pelo maior grau de um vértice $v_i(t)$.

Diâmetro: A maior distância m definida pelo número de arestas entre dois vértices (v_i, v_j) , considerando todos os pares de vértices na rede G do momento t .

Caminho Mínimo Médio: O caminho mínimo médio de uma rede G no momento t é definido pelo menor caminho l de um vértice v_i para todos os outros vértices da rede utilizando informações de distância para o cálculo de l .

Coefficiente de Agrupamento: Para uma rede G no momento t é dado pela média do coeficiente de agrupamento de todos os vértices C_v . Dados três vértices (v_j, v_w) com arestas definidas entre (v_i, v_j) e (v_j, v_w) , o coeficiente de agrupamento de um vértice C_v é definido pela probabilidade de existir uma aresta entre (v_i, v_w) .

Distância geodésica média e máxima entre os nós da rede: Representa a "distância" do nó que fez o retuite em relação ao autor do retuite original. Quanto mais distante maior foi a distância percorrida pela informação. A distância geodésica máxima mostra a maior distância atingida na difusão de informação e a média, sendo aquela em que a maior parte dos retuites chegou.

Betweenness: Para um vértice v_i é a razão de $\sigma_{st}(v_i, t)$ por σ_{sw} , em que $\sigma_{st}(v_i, t)$ representa o número de caminhos mínimos de vs a vw que passam por $v_i(t)$, e σ_{sw} o número de caminhos mínimos entre vs e vw , sendo a probabilidade dele estar no caminho mínimo entre dois outros vértices quaisquer. Essa métrica identifica os caminhos mínimos entre os vértices.

Closeness: Dado um vértice $v_i(t)$ no instante t é usado para definir o caminho mínimo médio, definido por l , entre o vértice v_i e todo vértice vw alcançável a partir de v_i tal que $(v_i, vw) \in V$. Com essa métrica é possível identificar o quão próximo um vértice está dos demais vértices da rede.

APÊNDICE B – GLOSSÁRIO DE TERMOS

Assuntos do Momento (*Trending Topics*): Assuntos que estão sendo muito comentados e estão em tendência de continuar sendo falados.

Canal: Meio de comunicação pelo qual informações são apresentadas ao público.

Celebridades de Redes Sociais: Usuários que possuem um alto número de seguidores, ou seja, alto alcance para suas postagens.

Citação (@): Postagem de um usuário na rede social Twitter, em que esse usuário faz referência a um outro usuário, normalmente representado pelo uso do caractere '@' antes do nome do usuário na postagem. Cada vez que um usuário é citado a postagem fica disponível no perfil tanto do usuário que postou quanto do usuário que foi citado. Assim como na visualização de postagens dos seguidores de um dos usuários envolvidos.

Cluster: Mais comumente traduzido nesse trabalho por Agrupamento. Grupo de usuários que compartilham interesses similares, se agrupando em torno dos interesses comuns.

Hashtags: Quando o usuário escreve o tuite fazendo uso do operador '#'. Quando dentro de um tuite o usuário usa o operador *hashtag* seguido de uma palavra, essa palavra se torna uma palavra especial, ligada aos outros tuites em que o mesmo *hashtag* é utilizado.

Hub Social: Mais comumente citado nesse trabalho como “usuário com alto nível de conectividade”. Dentro de uma rede, são os nós com muito mais conexões do que os outros, possuindo centralidade muito acima da média, possuindo forte influência e relevância para uma audiência qualificada.

In-Degree: Número de conexões que possuem relacionamento direcional em direção a este nó, no caso do Twitter é o número de seguidores.

Ligação Preferencial (*Preferential Attachment*): Característica em que os nós que estão na rede há mais tempo, ou que possuem maiores números de conexões, recebem maior probabilidade de receber novas conexões.

Menção (*Mention*): Quando um usuário cita outro usuário em seu tuite através do uso do operador '@', ao escrever '@' seguido do nome de um usuário válido, a ferramenta faz o vínculo com o usuário citado, aparecendo na linha do tempo do mesmo.

Mensagem Direta (*Direct Message*): Mensagem direta para um usuário sem visibilidade dos demais, como um e-mail interno.

Microblogging: Blog que permite apenas postagens curtas, normalmente com limite de caracteres permitidos.

Notícia: Postagem feita por um usuário oficial gerador de notícias.

Out-Degree: Número de conexões que possuem relacionamento direcional vindas deste nó, no caso do Twitter refere-se ao número de usuários seguidos.

Perfil de usuário: Cada usuário em uma rede social possui um usuário único, chamado de perfil de usuário.

Popularidade ou Sucesso de uma Postagem (Buzz): Quando uma postagem atinge um grande número de usuários.

Postagem: Informação postada de atualização de status por usuário em uma rede social que reflete uma atualização de status ou compartilhamento de opinião.

Postagem Pioneira: Postagem inicial que deu origem a outras postagens, seja por retuite, por repetição de conteúdo ou por novos usuários que irão usar a mesma interface de internet que gerou a postagem.

Postagem Comum: Postagem feita por um usuário que não é um gerador de notícias.

Push: Mensagens de notificação instantâneas de celular, que aparecem para o usuário toda vez que uma nova notificação é criada no ambiente monitorado.

Rede Social: Estruturas de comunicação social que possuem conexões por um ou mais tipos de relações, em que os indivíduos e/ou organizações interagem entre si e compartilham informações.

Resposta (Reply): Quando um usuário menciona (@) outro, e o segundo usuário faz menção do primeiro citando-o de volta de forma pública.

Retuite ou RT (Retweet): Tradução em português comumente utilizada para o termo *Retweet*. Ação na qual um usuário realiza uma repostagem da postagem de outra pessoa, compartilhando desta forma a postagem com sua rede de seguidores, amplificando assim o acesso da postagem do usuário inicial.

Seguido ou sendo seguido por: Referência aos usuários seguidos pelo usuário em questão.

Seguidor (Follower): Referência aos usuários que seguem as postagens do usuário em questão.

Seguir um Usuário: Quando um usuário opta por seguir um outro usuário, o coloca na rede de seguidores do segundo usuário. Recebendo em sua página de visualização de postagens, as atualizações do usuário seguido.

Sucesso de uma Postagem (Buzz): Vide Popularidade.

Tuite (Tweet): Postagem por mensagem na Rede Social Twitter, com até 140 caracteres na página do usuário, sendo enviado conseqüentemente para todos os seus seguidores que assinaram para acompanhar suas publicações.

Twitter: Rede social de *microblogging* unidirecional. Pode ser encontrado em: www.twitter.com.

Twitter Moments: Iniciativa do Twitter de montar uma seleção das principais postagens e notícias atuais, segundo critérios do próprio Twitter.

Usuário Agrupador: Usuário *Hub* ou com alto grau de centralidade.

Visualização de postagens: Na página do usuário, são apresentadas suas próprias postagens e as postagens dos usuários a quem ele segue, na ordem em que foram postadas.

APÊNDICE C – MÉTODO DE PESQUISA

Para o preparo dos dados empíricos que foram utilizados na pesquisa, foi realizado um processo de captura e tratamento dos dados. Conforme pode ser visto na Figura de processo abaixo:

Figura 23 - Processo de preparo dos dados empíricos. Fonte: SILVA, Rogério Olímpio da

A sequência de atividades realizadas foi a seguinte:

- 1) Os dados foram extraídos pela API do Twitter por uma empresa que realiza este tipo de monitoramento e nos forneceu os dados como incentivo acadêmico. O nome da empresa é Airstrip e pode ser localizada no endereço <https://www.airstrip.com.br>.
- 2) Os dados em formato de arquivo CSV foram tratados através de procedimentos de BigData, sem utilização de ferramenta específica, resultando nas tabelas de postagens iniciais e usuários iniciais.
- 3) Através de atividades de mineração de dados, utilizando os bancos de dados SQL Server, MySQL e SQLite, foram extraídos os dados estatísticos das postagens. Neste momento foi formada a rede de usuários seguidores e seguidos.
- 4) Utilizando funções de análise de rede em Python e R, foram geradas tabelas auxiliares que possibilitaram a análise dos KPIs das postagens e da rede. Merece ênfase o trabalho de pesquisa de identificação das postagens pioneiras e da quantificação do número de reaparições das postagens após a primeira incidência.
- 5) Foram gerados relatórios em Tibco Spotfire e em Python que nos possibilitaram analisar as informações estatísticas, bem como gerar os gráficos existentes neste documento.

- 6) Com base nos indicadores de performance e nos dados dos gráficos foram gerados parâmetros de calibragem para a simulação, permitindo que os dados de alfa, beta e gama da rede simulada tivesse características próximas da rede empírica.
- 7) Por fim, foram gerados gráficos da simulação nos mesmos formatos dos gráficos dos dados empíricos para que o comportamento do modelo fosse descrito e entendido.

APÊNDICE D - TABELAS AUXILIARES

Tabelas auxiliares geradas durante o trabalho de mineração de dados:

Tabela 7 - Tabela de Dados 'Usuario'

Tabela:	Usuario
Atributo	Dicionário
Perfil_usuario	Usuário público que é visível para ser seguido e referenciado
Nome_Usuario	Nome de cadastro do usuário
Data_Hora	Data e Hora que a postagem foi publicada no Twitter
Posts_Totais	Quantidade de Postagens totais do usuário no Twitter no momento desta postagem
Seguidores	Quantidade de seguidores do usuário
Contador	Número de sequência de contagem do número da linha
Qtos_Posts	Quantidade de Postagens do usuário neste experimento
Gera_noticia	É um usuário gerador de notícias
qtde_@_outros	Quantidade de postagens em que o usuário cita outros usuários
qtde_outros_@	Quantidade de postagens em que o usuário é citado por outros
max_mesmo_conteudo	Número máximo de uma postagem do usuário que teve outros com o mesmo conteúdo, quando esta postagem foi a primeira
horas_posts_durou_mais	Quantidade de horas da postagem deste usuário que durou mais tempo
qtde_posts_RT	Quantas postagens deste usuário foram retuitadas
qtde_hash	Quantas postagens deste usuário tem hashtag

Tabela 8 - Tabela de Dados 'Analitico'

Tabela:	Analitico
Atributo	Dicionário
ID_Distinct	Identificador do registro único do texto postado na tabela Distinct.
Qtde_MesmoConteudo	Quantidade de outras postagens com o mesmo conteúdo desta.
ID_Primeiro	Identificador Único (Distinct) da postagem que teve o primeiro conteúdo desta, caso essa seja a primeira, o valor será o mesmo do ID_Distinct
ID_Ultimo	Identificador Distinct da última postagem que teve o mesmo conteúdo desta.
Data_Inicial	Data da primeira postagem com o mesmo conteúdo, caso essa seja a primeira.
Hora_Inicial	Hora da primeira postagem com o mesmo conteúdo, caso essa seja a primeira.
Data_Final	Data da última postagem com o mesmo conteúdo, caso essa seja a primeira.
Hora_Final	Hora da última postagem com o mesmo conteúdo, caso essa seja a primeira.

Qtde_Horas	Quantidade de horas que esta postagem durou.
ID_Usuario	Identificador do usuário
Eh_noticia	Esta postagem é uma notícia

Tabela 9 - Tabela de Dados 'Chamados'

Tabela:	Chamados
Atributo	Dicionário
Usuario_Origem	Usuário que citou outro usuário em uma postagem
Usuario_Chamado	Usuário citado pelo outro na postagem

Tabela 10 - Tabela de Dados 'Hashtags'

Tabela:	Hashtags
Atributo	Dicionário
ID_Hashtag	Número sequencial identificador
Usuario	Usuário que postou a Hashtag
Hashtag	Texto postado como Hashtag

APÊNDICE E – CRIAÇÃO DAS TABELAS DE DADOS

Abaixo seguem os comandos SQL (versão MySQL Workbench 6.1) para a criação das tabelas utilizadas no experimento.

Tabela “analitico”:

```
CREATE TABLE `analitico` (
  `ID_Distinct` int(11) DEFAULT '0',
  `Qtde_MesmoConteudo` int(11) DEFAULT '0',
  `ID_Primeiro` int(11) DEFAULT '0',
  `ID_Ultimo` int(11) DEFAULT '0',
  `Data_Inicial` text,
  `Hora_Inicial` text,
  `Data_Final` text,
  `Hora_Final` text,
  `Qtde_Horas` int(1) NOT NULL DEFAULT '0',
  `ID_Usuario` varchar(16) DEFAULT NULL,
  KEY `idx_a3_ID_Prim` (`ID_Primeiro`),
  KEY `idx_a3_ID_Ult` (`ID_Ultimo`),
  KEY `idx_a_ID_Usu` (`ID_Usuario`),
  KEY `idx_a_ID_Dist` (`ID_Distinct`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;7
```

Tabela “chamados”:

```
CREATE TABLE `chamados` (
  `Usuario_Origem` text,
  `Usuario_Chamado` text
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
```

Tabela “dados”:

```

CREATE TABLE `dados` (
  `Data_Post` text,
  `Hora` text,
  `Contador` bigint(20) DEFAULT NULL,
  `Nome_usuario` varchar(23) DEFAULT NULL,
  `Perfil_usuario` text,
  `Polaridade` text,
  `Categoria` varchar(33) DEFAULT NULL,
  `Likes_Favs` text,
  `Shares_RTs` text,
  `Posts_tot_usuario` text,
  `Seguidores` text,
  `Genero` text,
  `Regiao` text,
  `Dispositivo` text,
  `ID_post` text,
  `ID_usuario` text,
  `Tamanho_Conteudo` text,
  `PeriodoDia` varchar(10) DEFAULT NULL,
  `QtdeDepois` int(11) DEFAULT NULL,
  `DiaSemana` varchar(10) DEFAULT NULL,
  UNIQUE KEY `idx_dados_Contador` (`Contador`),
  FULLTEXT KEY `idx_dados_Perfil_usuario` (`Perfil_usuario`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;

```

Tabela “distintos”:

```

CREATE TABLE `distintos` (
  `ID_Distinto` int(11) DEFAULT '0',
  `Texto_Distinto` varchar(61) DEFAULT NULL,
  UNIQUE KEY `idx_ID_Distinto` (`ID_Distinto`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;

```

Tabela “hashtags”:

```

CREATE TABLE `hashtags` (
  `ID_Hashtag` int(11) DEFAULT '0',
  `Usuario` varchar(45) DEFAULT NULL,
  `Hashtag` text
) ENGINE=InnoDB DEFAULT CHARSET=utf8;

```

Tabela “posts”:

```

CREATE TABLE `posts` (
  `Contador` int(7) DEFAULT NULL,
  `Perfil_usuario` char(15) DEFAULT NULL,
  `Data_Hora` datetime DEFAULT NULL,
  `Posts_Totais` int(6) DEFAULT '0',
  `Seguidores` int(6) DEFAULT '0',
  `PeriodoDia` varchar(9) DEFAULT NULL,
  `DiaSemana` varchar(7) DEFAULT NULL,
  `Tamanho_Conteudo` int(11) DEFAULT '0',
  `QtdeDepois` int(11) DEFAULT '0',
  `Tem_RT` char(1) DEFAULT NULL,
  `Eh_Primeiro` char(1) DEFAULT NULL,
  `Eh_Ultimo` char(1) DEFAULT NULL,
  `Pri_QtdeHoras` int(11) DEFAULT '0',
  `Pri_Distinto` int(11) DEFAULT '0',
  `Qtde_MesmoConteudo` int(11) DEFAULT '0',
  KEY `idx_tmp_cont` (`Contador`),
  KEY `idx_tmp_usu` (`Perfil_usuario`),
  KEY `idx_tmp_dh` (`Data_Hora`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;

```

Tabela “usuarios”:

```

CREATE TABLE `usuarios` (
  `Perfil_usuario` char(15) DEFAULT NULL,
  `Data_Hora` datetime DEFAULT NULL,
  `Posts_Totais` int(6) DEFAULT '0',
  `Seguidores` int(6) DEFAULT '0',
  `Contador` int(7) DEFAULT NULL,
  `Qtos_Posts` int(11) DEFAULT '0',
  KEY `idx_usu_perf` (`Perfil_usuario`),
  KEY `idx_usu_dh` (`Data_Hora`),
  KEY `idx_usu_cont` (`Contador`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;

```

APÊNDICE F – SIMULAÇÃO POR MODELAGEM BASEADA EM AGENTES

Abaixo segue o código fonte da programação da simulação em Python:

```
# coding=utf-8
# C:\Users\Rogério\Google Drive\Rogério\SCX_Rog\00. 2017\1. Defesa\2. Networkx
# mklink /d c:\networkx "C:\Users\Rogério\Google Drive\Rogério\SCX_Rog\00. 2017\1. Defesa\2. Networkx"

# importação das bibliotecas
import networkx as nx
import random
import matplotlib.pyplot as plt
from datetime import datetime
import numpy as np
import sys
from collections import Counter

pinf = float('+inf')
ninf = float('-inf')
fnan = float('nan')

caminho_destino = "C://networkx//03.Redre_Simulada//Resultados//"
#C:\networkx\00.Simulacao\Resultados
arquivo = 'simulacao'

def limpar_lista(lista):
    for n,i in enumerate(lista):
        if (i==0 or i==pinf or i==ninf or i==fnan):
            lista[n]=1
    return lista

# documentando data e hora atual para arquivos de dados que serão gerados
agora = str(datetime.now()).replace("-", "").replace(":", "").replace(" ", "_")[:15]

# definição do tamanho da rede em nós (agentes)
# definição da distribuição da classificação dos alcances dos nós dentro da rede simulada

delta_in  = 0.50 # 0.2
delta_out = 0.01 # 0
alpha     = 0.41 # 41
beta      = 0.54 # 54
gamma     = 0.05 # 05

G = nx.scale_free_graph(50000, alpha=alpha, beta=beta, gamma=gamma, delta_in =
delta_in, delta_out = delta_out)
# scale_free_graph(n, alpha=0.41, beta=0.54, gamma=0.05, delta_in=0.2, delta_out=0,
create_using=None, seed=None)
# https://networkx.github.io/documentation/networkx-
1.10/reference/generated/networkx.generators.directed.scale_free_graph.html
```

```

# variável node_list recebe a estrutura de rede com os nós e suas arestas
node_list = G.nodes()

# define posição dos pontos no gráfico a ser plotado, plota gráfico e salva em arquivo
#pos = nx.circular_layout(G)
#nx.draw_networkx(G, pos=pos, with_labels=False, node_size=1, edge_size=1)
#plt.savefig(caminho_destino+"1-Figura_"+agora+".png", dpi=1000)

# gera arquivo com dados das conexões
print ('1) Rede Gerada!')
nx.write_edgelist(G, caminho_destino+"2-Dados_"+agora+".txt", data=False)

# Definição das Probabilidades de postagem, período online do dia, de leitura e de retuitar
p_postar = (0.62) # probabilidade de postar = 32% CALIBRADO

p_online = (0.25) # probabilidade de estar online = 25%
p_ler = (0.35) # probabilidade de ler a postagem = 35%
p_retuitar = (0.05) # probabilidade de retuitar = 5%

# Definição de um percentual geral de chance para retuite, onde todas as probabilidades são
colocadas juntas
p_geral_retuite = 0.0016 #p_online * p_ler * p_retuitar = 16% CALIBRADO

# Inicialização das variáveis da rede
p_random = 0 # variável aleatória de sorteio da probabilidade para cada rotação
i = 0 # variável de iteração para cada nó postando
postagem = 0 # variável de numeração do nome da postagem
ultima_post = 0 # variável de memória da última postagem
nos_postando = [] # lista com todos os nós que estão postando
postagens = [] # lista com as postagens realizadas
seguidores = [] # lista com os seguidores do nó postando
lista_seguidores = [] # lista de usuários e seguidores para calcular a distribuição

# Iteração que passa por todos os nós e sorteia, de acordo com as probabilidades, se o
usuário/nó vai postar ou não
# Caso o nó vá postar, ele é incluído na lista de nós que efetivamente vão postar
for i in range(len(node_list)-1):

    # sorteio da probabilidade de comparação para essa rodada
    p_random = random.random()

    # verifica se a probabilidade sorteada está dentro da faixa probabilidades desse usuário
    if (p_random <= p_postar):

        # usuários que vão postar nesta sessão de simulação, sendo acrescido na lista: usuário,
        número de postagem
        nos_postando.append([i,0])

# gerando arquivo externo com lista de nós que vão postar nessa sessão
print ('2) Lista de Nós-Postando Geradas!')
np.savetxt(caminho_destino+'3-Nos_Postando_'+agora+'.txt', nos_postando, delimiter=',',
fmt='%s')

```

```

qtos_postando = len(nos_postando)-1

# variáveis que atuam nas iterações de postagem e repostagem
usuario_alcance = []      # lista com os usuários que estão postando e a quantidade de
seguidores que eles possuem
i = 0                    # variável que identifica o número da rodada
z = 0                    # variável acumula a quantidade de rodadas
x_rodadas = 12           # quantidade de dias de postagens
qp = qtos_postando       # controle auxiliar da quantidade de nós que estão postando neste
dia

# Iteração que realiza todas as rodadas de postagens da lista de nós que postariam
# Inclusive os retuites também, uma vez que cada retuite é acrescentado na mesma lista como
nova postagem futura
while i <= qtos_postando:

    # escape para caso a simulação desestabilize não ultrapasse 7 casas decimais
    if (qtos_postando >= (1*10**6)):
        print ('Processo interrompido porque passou de 7 casas decimais!')
        break

    # Nesse ponto é verificado se a quantidade de rodadas chegou ao limite de rodadas
    # Iniciando-se assim um novo "dia", zerando o contador de rodadas e voltando para o
primeiro usuário da lista (i==0)
    if (i==qtos_postando):
        z = z+1
        if (z<x_rodadas):
            i=0
    # Neste cenário é a primeira postagem desse dia, iniciando com o primeiro usuário (i==0)
    if (i == 0):
        # como é a primeira postagem, nova postagem terá o código fixado em "1"
        ultima_post = 1 # se nao tiver uma postagem definida (caso de retuite), criar nova
postagem
        # simulação da postagem, adiciona na lista de postagens, armazenando: usuário e
número da postagem
        postagens.append([ nos_postando[i][0], ultima_post ])
        # armazena o número da postagem atual como referência numérica
        postagem_ref = ultima_post

    # Caso não seja o primeiro desse dia, e não seja retuite, armazena novo código de
postagem
    elif (i > 0 and nos_postando[i][1] == 0):
        # não é a primeira postagem, então adiciona uma unidade ao código das postagens
        ultima_post = ultima_post + 1
        # simulação da postagem, adiciona na lista de postagens, armazenando: usuário e
número da postagem
        postagens.append([ nos_postando[i][0], ultima_post ])
        # armazena o número da postagem atual como referência numérica
        postagem_ref = ultima_post

    # Caso não seja o primeiro desse dia, e seja um retuite, armazena o mesmo código de
postagem
    elif (i > 0 and nos_postando[i][1] > 0):

```



```

    # simulação da postagem, adiciona na lista de postagens, armazenando: usuário que
    # está postando e número da postagem de retuite
    postagens.append([ nos_postando[i][0], nos_postando[i][1] ])
    # armazena o número da postagem de retuite como referência numérica
    postagem_ref = nos_postando[i][1]

    # caso seja um nó não encontrado na lista, é gerada uma mensagem para análise do
    # problema
    else:
        print ("ATENCAO")
        print ("i -> ", i, " - nos_postando[0] -> ", nos_postando[0], " - nos_postando[1] -> ",
        nos_postando[1])
        break

    # a lista de seguidores é zerada antes da análise dos seguidores do usuário que acabou
    # de postar
    seguidores = []

    try:
        # a lista de seguidores recebe todos os seguidores do usuário que acabou de postar,
        # armazenando: o seguidor
        seguidores = G.predecessors(nos_postando[i][0])
        # lista com os usuários que estão postando armazena: usuário que postou, e quantidade
        # de seguidores deste usuário
        usuario_alcance.append([nos_postando[i][0],len(seguidores)])

    # caso seja gerada uma excessão, ela será exibida em tela para análise do problema
    except:
        # para tratamento da excessão, imprime o número de usuário e o momento da iteração
        print ('exception ->', i)
        print ('tamanho exception -> ', qtos_postando)
        # então soma a iteração para o próximo usuário e continua o looping
        i = i+1
        continue

    # após a postagem, serão simulados os retuites de dentro dos seguidores do usuário que
    # postou
    # iteração na lista de seguidores que receberam
    for j in seguidores:
        lista_seguidores.append([i,j]) # Gerando lista de seguidores na marra
        p_random = random.random()

        # verifica se a probabilidade aleatória está dentro da probabilidades de retuite deste nó
        if (p_random <= p_geral_retuite):
            # coloca o retuite na lista das próximas postagens, armazenando usuário e número da
            # postagem atual (referência numérica)
            nos_postando.append([j,postagem_ref])
            # adiciona mais uma unidade na variável que controla quantidade de postagens sendo
            # realizada nesse dia
            qtos_postando = qtos_postando+1

    # soma a iteração para o próximo usuário e continua o looping
    i = i+1

```

```

# Gera lista das postagens realizadas, onde: [i][0] = usuario ; [i][1] = postagem
print ('3) Lista de Postagens Gerada!')
np.savetxt(caminho_destino+'4-Postagens'+agora+'.txt', postagens, delimiter=',', fmt='%s')
# Imprime lista de postagens e quantidade de postagens realizadas
print ('4)len(postagens) -> ', len((postagens)))

# Variáveis são inicializadas para gerar lista com quantidade de postagens acumuladas e
usuários que postaram
i = 0          # número da iteração na lista de postagens realizadas
j = 0          # número da iteração na sublista que
qtde_postadas = [] # lista com quantidades das postagens, que será exportada
qtde_simples = []  # quantidades simples com usuário para relatório

# Percorre lista de postagens realizadas
for i in range(len((postagens))-1):
    # Define nome do usuário com uma letra aleatória de "A K" e o número do usuário atual da
    lista sendo percorrida
    usuario = random.choice('ASDFGHJKL')+str(postagens[i][0])
    # Define como postagem o número da postagem da lista sendo percorrida
    postagem = postagens[i][1]

    # Zera flag para verificar se encontrou a postagem
    achou = False

    # Percorre lista de quantidades de postagens para saber se a postagem já faz parte desta
    lista
    for j in range(len(qtde_postadas)):
        # Verifica se a postagem atual é igual ao registro atual percorrido
        if (qtde_postadas[j][0] == postagem):
            # Se for igual marca o flag e interrompe a iteração de procura
            achou = True
            break

    if achou:
        # Se encontrou a postagem na lista soma um no número de postagens encontradas
        anteriormente
        qtde_postadas[j][2] = str(int(qtde_postadas[j][2]) + 1)
        qtde_simples[j][1] = str(int(qtde_postadas[j][1]) + 1)
    else:
        # Se não encontrou, inicializa a quantidade de postagens na lista,
        # armazenando: postagem; código do usuário; quantidade de postagens (inicialmente ==
        1); nome do usuário
        qtde_postadas.append([ postagens[i][1], postagens[i][0], 1, usuario])
        qtde_simples.append([usuario,1])

# Extrai e imprime informações estatísticas sobre a rede
nx_info = []
nx_info = nx.info(G)

# grava arquivos com as listas geradas
np.savetxt(caminho_destino+'5-Nos_Postando_'+agora+'.txt',  nos_postando,  delimiter=',',
fmt='%s')

```

```

np.savetxt(caminho_destino+'6-Qtde_Postadas'+agora+'.txt', qtde_postadas, delimiter=',',
fmt='%s')
np.savetxt(caminho_destino+'7-Alcance'+agora+'.txt', usuario_alcance, delimiter=',', fmt='%s')
#np.savetxt(caminho_destino+'8-Tipo_Alcance'+agora+'.txt', tipo_alcance, delimiter=',',
fmt='%s')
np.savetxt(caminho_destino+'9-Lista de Seguidores'+agora+'.csv', lista_seguidores,
delimiter=',', fmt='%s')
np.savetxt(caminho_destino+'1-Qtde_simples'+agora+'.csv', qtde_simples, delimiter=',',
fmt='%s')

# imprime mensagem sobre finalização da simulação
print ("\nPostagens Simuladas!")

# Início do cálculo da distribuição de grau
deltain = 1.0984
deltaout = 1.7781

rede = nx.read_edgelist(caminho_destino+'9-Lista de Seguidores'+agora+'.csv', delimiter=',',
create_using=nx.DiGraph())

N,K = rede.order(), rede.size()
avg_deg = float(K)/N

# variável node_list recebe a estrutura de rede com os nós e suas arestas
node_list = rede.nodes()

in_degrees = rede.in_degree()
in_values = sorted(set(in_degrees.values()))
in_hist = [list(in_degrees.values()).count(x) for x in in_values]

out_degrees = rede.out_degree()
out_values = sorted(set(out_degrees.values()))
out_hist = [list(out_degrees.values()).count(x) for x in out_values]

plt.figure()
plt.loglog(out_values, out_hist, 'bv-') # blue color with marker 'v'
plt.loglog(in_values, in_hist, 'ro-') # red color with marker 'o'

#sys.exit()

plt.legend(['In-degree', 'Out-degree'])
plt.xlabel('Grau')
plt.ylabel('Número de nós')
#plt.title('Simulada'+ N ->'+str(int(N))+ K ->'+str(int(K)))
plt.title('Simulada'+ N ->'+str(int(N))+ K ->'+str(int(K))+ d_in ->'+str(delta_in)+ d_out -
->'+str(delta_out))
plt.grid(True)

plt.xlim([0, 2 * 10**2])
plt.savefig(caminho_destino+"20.Sim_Rede_loglog_"+agora+'.pdf')
plt.close()

```

```
out_values = limpar_lista(out_values)
out_hist = limpar_lista(out_hist)
in_values = limpar_lista(in_values)
in_hist = limpar_lista(in_hist)

slope1, intercept1 = np.polyfit(np.log(out_values), np.log(out_hist), 1)
slope2, intercept2 = np.polyfit(np.log(in_values), np.log(in_hist), 1)

arq = open(caminho_destino+'Slope_'+arquivo+agora+'.txt', "w")

arq.write("Arestas: "+str(K))
arq.write("\n")
arq.write("Grau médio: "+str(avg_deg))
arq.write("\n\n")
arq.write("Informações da rede: "+str(nx_info))
arq.write("\n\n")
arq.write("slope outdegree: "+str(slope1))
arq.write("\n")
arq.write("slope indegree: "+str(slope2))
arq.close

# Considerar 7 para qtde_usu x seguidores
# Considerar 11 para qtde_usu x postagens ou RTs

# Final do trecho novo
print ('Trecho novo executado')
```