



O sentimento político em redes sociais: *big data*, algoritmos e as emoções nos tweets sobre o *impeachment* de Dilma Rousseff

Political sentiment in social networks: big data, algorithms and emotions in tweets about the impeachment of Dilma Rousseff

Fabio Malini *

Patrick Ciarelli **

Jean Medeiros***

RESUMO

Este artigo se propõe a ampliar a metodologia perspectivista (MALINI, 2016) de análise de redes sociais, incorporando um procedimento de análise dos sentimentos das mensagens postadas em redes de controvérsias políticas, em particular, em dois momentos distintos da campanha pelo *impeachment* da presidenta Dilma. O primeiro é o período da eclosão das manifestações antipetistas, no dia 15 de março de 2015. O segundo, dia 27 de agosto de 2016, quando a presidenta é deposta do cargo. Realiza uma revisão sobre a análise de sentimentos em megadados do Twitter e constrói uma metodologia que combina classificação humana de textos com aplicação de algoritmos genéticos de análise de textos, no intuito de analisar sentimentos genéricos (baseado na polarização positivo/negativos) e sentimento específicos, baseados nas seguintes emoções: Alegria, Raiva, Medo, Antecipação, Desgosto, Tristeza,

ABSTRACT

This article aims to expand the perspectivist methodology (Malini, 2016) of social networks analysis, incorporating a proceeding of sentiment analysis of the messages posted in networks of political controversies, in particular, in two distinct moments of the campaign for the impeachment of President Dilma. The first is the period of the outbreak of PT protests, on March 15, 2015. The second, on August 27, 2016, when the president is deposed. We will be doing a theoretical review about sentiment analysis in Big Data on Twitter to build a methodology that combines human classification of texts with the application of genetic algorithms of text analysis and to analyze generic sentiments (based on positive / negative polarization) and specific sentiment, based on emotions like Joy, Anger, Fear, Anticipation, Disgust, Sadness, Surprise and Trust. It concludes by demonstrating that pro and anti-Dilma movements are marked by a predominance of anger, fear and anxiety,

* Doutor em Comunicação e Cultura. Professor Associado da Universidade Federal do Espírito Santo. Endereço: UFES, Campus Goiabeiras, Avenida Fernando Ferrari 514, Centro de Artes, Prédio Cemuni I, sala 10, Vitória, CEP 29075-910, Espírito Santo. Telefone: (27) 4009-2752. E-mail: fabiomalini@gmail.com

** Doutor em Engenharia Elétrica. Professor Adjunto da Universidade Federal do Espírito Santo. Endereço: UFES, Campus Goiabeiras, Avenida Fernando Ferrari 514, Centro de Artes, Prédio Cemuni I, sala 10, Vitória, CEP 29075-910, Espírito Santo. Telefone: (27) 4009-2752. E-mail: pmco1@uol.com.br

*** Doutorando em Política Científica e Tecnológica, Unicamp. Pesquisador Associado do Laboratório de estudos sobre Imagem e Cibercultura (LABIC/UFES). Endereço: UFES, Campus Goiabeiras, Avenida Fernando Ferrari 514, Centro de Artes, Prédio Cemuni I, sala 10, Vitória, CEP 29075-910, Espírito Santo. Telefone: (27) 4009-2752. E-mail: jeanmrmedeiros@gmail.com

Surpresa e Confiança. Conclui demonstrando que os movimentos pró e anti-Dilma são marcados pelo predomínio de sentimento de raiva, medo e ansiedade, confirmando a hipótese que a trolagem ofensiva demarca o estilo da indignação propagada em redes políticas no Twitter brasileiro.

Palavras-chave: Análise de Sentimento; Big Data; Redes; Política; Twitter.

confirming the hypothesis that an offensive trolling demarcates the style of indignation propagated by political networks in Brazilian Twitter.

Keywords: Sentiment Analysis; Big Data; Social Network; Politics; Twitter.

BIG DATA E ANÁLISE DE SENTIMENTO NAS REDES SOCIAIS: DEBATE TEÓRICO

Há consenso (embora a crítica ainda seja incipiente): é intensa a transposição de informações pessoais e institucionais para bancos de dados digitais. Com uma enorme parcela da população utilizando diariamente as redes sociais, e conseqüentemente deixando rastros, e a disponibilidade cada vez maior de tecnologias para reter, agrupar e processar esses dados, aqueles que possuem acesso a esse manancial de dados prontamente passaram a utilizar tais ferramentas para apoiar os mais diversos processos de tomada de decisão. Para Tufekci (2014), *big data* não se trata apenas de uma quantidade maior de dados, mas de uma grande mudança na natureza dos dados e sua possibilidade de agregação a outros dados.

De acordo com Diebold (2012), a origem da utilização do termo “*big data*”, ligando-o conscientemente ao fenômeno que hoje descreve, deu-se em meados dos anos 1990, em seminários e até mesmo anúncios publicitários da empresa americana Silicon Graphics (SGI). No entanto, apenas em 2000, identifica-se o primeiro artigo acadêmico sobre *big data*, intitulado “‘Big data’ dynamic factor models for macroeconomic measurement and forecasting”, produzido por ele próprio, no campo da econometria. Para Boellstorff (2013), a consolidação do termo *big data*, porém, se deu apenas por volta de 2008. Embora seja um período curto de tempo, o fenômeno já influencia fortemente os setores da tecnologia, o meio acadêmico, o público, o privado, o militar, entre outros, e movimenta muito dinheiro em torno de si.

Diebold (2012) apresenta “*big data*” não apenas como um termo altamente disseminado, mas como um fenômeno contínuo, e até então inabalável, e uma disciplina que emerge. Para Boyd e Crawford (2011), o valor do *big data* não está em seu tamanho, mas nas relações entre seus dados. A agregação dos dados traz a configuração em rede para a análise, e dois tipos de formação de redes podem surgir a partir dos dados: as “redes articuladas”, resultantes da lista de contatos (amigos, seguidores, etc.) dos usuários; e as “redes de comportamento”, derivadas dos padrões de comunicação (marcações na mesma foto, envio de *e-mail*, presença no mesmo lugar, etc.)

Dentro dos estudos em ciências sociais a partir de *big data*, uma grande vertente é a análise de redes sociais, a partir de fontes denominadas “*social data*”. “*Social data*” são os dados extraídos das mídias sociais num formato legível para computadores, sendo complementado por *metadata*, de forma a fornecer não apenas o conteúdo, mas o contexto do dado. *Metadata* é a inclusão de certos elementos de dado de apoio em relação a um dado específico, como informação sobre localização, engajamento e *links* (BOELLSTORFF, 2013).

Este trabalho busca compreender o caráter emocional inscrito no “social data”, a saber: tweets publicados em dois momentos distintos da jornada em prol do *impeachment* de Dilma Rousseff. O primeiro data da eclosão das manifestações antipetistas, no dia 15 de março de 2015. O segundo, do dia 27 de agosto de 2016, quando a presidenta é deposta do cargo. A partir desse caso, buscamos demonstrar uma experiência metodológica de análise de sentimento no Twitter, a fim de demonstrar como a produção política discursiva dos sujeitos em rede poderá, cada vez mais, ser objeto de “algoritmos emocionais” cujo principal objetivo será identificar os sentimentos que permeiam o debate no espaço público virtualizado.

O campo da análise computacional de sentimento passou a ser um desafio para entender o comportamento emocional coletivo inscrito na cornucópia demensagens –posts, tweets, updates etc. – de perfis *online*. O termo “análise de sentimento” (*sentiment analysis*) possui diferentes sinônimos na literatura científica, tais como *opinion mining*, *opinion extraction*, *sentiment mining*, *subjectivity analysis*. “O papel da análise de sentimento cresceu significativamente com a rápida difusão das redes sociais, *microblogs* e fóruns” (KOLCHYNA et al., 2016, p.2). As opiniões expressas pela demonstração de sentimentos passaram a ser úteis para tomada de decisões, e isso não é só verdade para os indivíduos, mas também é verdade para as organizações (LIU, 2010).

Existem diversas abordagens sobre a forma como os sentimentos devem ser detectados e/ou processados. As quatro maneiras mais viáveis de identificar e analisar sentimentos são: (1) por meio de sentimentos específicos, ou seja, identificação de sentimentos preestabelecidos para o *corpora* em questão, geralmente pensados de acordo com o tema coletado (por exemplo, analisar o medo em contextos de difusão de notícias sobre violência, ou o empoderamento em quadros de ativismo *online*); (2) os sentimentos genéricos, os quais também são preestabelecidos para análises comparativas, porém, se inserem no termo genérico por se enquadrar em qualquer *corpora*, e.g., alegria, raiva, medo, empoderamento, desgosto, surpresa, confiança, etc.; (3) polaridades, nas quais se pode perceber, por meio da soma de pontuações de palavras entre positivo e negativo, qual é a polaridade das frases, mensagens ou perfis analisados; e (4) *emoticons* e/ou *emojis*, os quais se dividem entre positivos e negativos ou entre sentimentos genéricos.

O primeiro visa especificar em que perspectiva os dados são analisados. Por exemplo, se a análise está em um *data set* sobre o movimento Occupy, busca-se compreender, pelo seu conteúdo, o sentimento favorável ou contrário ao movimento, ou se um conjunto de usuário apoia ou recusa as causas. Para tanto, o analista deve organizar um conjunto de palavras que sirvam de filtros para melhor enquadrar os sentimentos num conjunto de termos assertivos. Tecnicamente, esse conjunto de palavras denomina-se dicionário. Assim, o sintagma “Nós somos os 99%” estará no dicionário dos favoráveis ao movimento. E o termo “vandalismo”, no dos contrários.

No segundo caso, o objetivo metodológico é podermos encontrar sentimentos genéricos, como alegria, raiva, ódio, amor, excitação, antecipação, medo, confiança, etc. Para isso, é necessária uma classificação mais profunda das palavras no conjunto ou em um banco de dados pronto, dando mais assertividade emocional às palavras, isto é, correlacionando as possibilidades de sentimentos que elas podem representar. Essa maneira pode ser mais complexa, porém, viável. E feita a partir de uma tarefa manual supervisionada de indexação das palavras às emoções. E pode ser simplificada, por exemplo, ao convertermos as palavras à sua forma canônica, facilitando encontrá-las no banco de dados. De qualquer modo, a presença dos

gatekeepers é importante para qualificar a etiquetagem que, posteriormente, o algoritmo irá executar em todo *data set*.

No terceiro caso, o foco é na frase, mas o cálculo ainda é sobre a palavra. É um método automatizado em que a polaridade (positivo/negativo) de uma frase pode ser calculada por *ranking*: 1 (um) para palavras positivas e -1 (menos um) para palavras negativas, ou pode variar, dependendo do quão positiva/negativa é a palavra. Dessa forma, define-se a polaridade de cada frase a partir da soma das pontuações de polaridade de cada palavra.

Por último, mas não menos importante, não podemos subestimar o uso de *emoticons* e *emojis* em redes sociais. É muito comum para os usuários empregarem essa funcionalidade de texto em mensagens, postagens, perfis, etc., já que é a forma menos complexa para expressar sentimentos. Dessa forma, faz-se necessária a construção de bases de dados ou, até mesmo, encontrar bancos já construídos, para identificar sentimentos representados por um *emoticon* ou *emoji*. A desvantagem é o volume baixo de *emojis* ou *emoticons* em mensagens de redes sociais (ARAÚJO; GONÇALVES; BENEVENUTTO, 2013, p. 2).

Há ainda novos processos metodológicos testados. Por exemplo, Kauer (2016) trata os sentimentos em dois métodos, baseando-os por meio da identificação de aspectos e de atribuição de polaridade. O primeiro método condiz em uma forma que mistura a utilização de ferramentas de processamento de linguagem natural com algoritmos de aprendizagem de máquina. Para a segunda maneira, o autor se utiliza de um motor de busca em que se comparam os termos analisados com textos cujas classes de sentimento já são conhecidas.

A primeira forma é feita mediante a extração da categoria da opinião, extração do alvo da opinião e atribuição da polaridade do sentimento. Esse método divide a frase em uma trinca, que contém: a *categoria*, na qual representa a entidade e os aspectos encontrados na frase; o *alvo da opinião*, e caso não houver, considera-se nulo; e, por fim, o sentimento atribuído à opinião, sendo ele uma polaridade positiva, negativa ou neutra, que condiz com o segundo método levantando pelo autor.

A atribuição de polaridade de sentimento é feita por intermédio de um mecanismo de aprendizagem de máquina: para cada palavra analisada, um banco é construído a partir das associações aplicadas a elas. Se uma ou mais das palavras da afirmação se configurarem em positivas ou negativas, será atribuído às demais palavras da sentença a mesma polaridade encontrada nela(s). Assim, o algoritmo vai aprendendo com a máquina em como classificar as palavras e as frases.

Já no trabalho realizado por Toret (2015) e Oscar Marín Miró, do coletivo Outliers, a análise de sentimentos de *tweets* das mobilizações nas praças espanholas (#15M) é enquadrada a partir de sentimentos genéricos, tais como: empoderamento, indignação, medo e felicidade. Eles extraíram esses sentimentos a partir do que definiram de contexto semântico, isto é, os *tweets* indexados com as *hashtags* #15M no Twitter. Para cada emoção, geraram uma lista de expressões, sendo necessário revisá-las para que não gerassem ruídos em função de um ou de outra expressão não estar dentro do contexto. E assim foram capazes de criar três conceitos: “carga emocional” (fração global de *tuítes* indexado como emocional em relação ao total de *tuítes*) e “viralidade emocional” (fração de *tuítes* que são *retuítes*, indexados como “emocionais” em relação ao total dentro de uma janela temporal). Com isso foram capazes de comparar *tweets* comuns e *tweets* do 15M, descobrindo que o segundo tem uma carga emocional dobrada. Isso prova uma grande força de carga emocional no processo de mobilização social nas ruas.

A combinação entre alta e ampla ativação emocional de milhares de pessoas (nós) e uma alta coesão de vocabulário (sintonia de discurso e pensamento a respeito do que ocorre) ativa uma inteligência comum e uma criatividade expandida que permite a produção acelerada de linguagem comum, de um corpo conectado (TORET, 2016, p. 81).

Um outro trabalho importante de se ressaltar é o dos autores Bollen, Mao e Pepe (2011). Eles articularam a ideia de sentimentos à de humor (*mood*), a partir da psicometria definida pela escala de classificação *Profile of mood states*, que define os seguintes estados de ânimo: tensão, depressão, fadiga, vigor, raiva e confusão. Os dados para a análise foram todos os tweets postados no segundo semestre de 2008. A fase empírica final consistiu em comparar os tweets coletados com aqueles associados a grandes eventos da mídia tradicional e demais culturas populares. O objetivo consistiu em identificar uma relação entre as emoções expressas pelos usuários em mensagens do seu cotidiano e as comparar com aquelas mensagens que circulam sobre grandes eventos sociais e econômicos.

Em seus resultados, Bollen, Mao e Pepe (2011) puderam concluir que o período coletado mostrou-se tumultuado, com variações bruscas de emoções entre eventos. E podem-se notar picos de emoções, como tensão no dia das eleições presidenciais dos EUA e no Dia de Ação de Graças, em que o vigor teve um aumento exponencial.

Nós concluímos que eventos sociais, políticos, culturais e econômicos estão correlacionados significativamente, mesmo se há atrasos nas flutuações de níveis de humor público ao longo de uma gama de diferentes dimensões do humor. Para concluir, nós trazemos a seguinte contribuição metodológica: defendemos que a análise de sentimentos de um *corpora* de texto-minuto (tais como os tweets) é eficiente obtida através de uma abordagem sintática, baseado no termo que não requer nenhum treinamento ou aprendizagem de máquina (BOLLEN: MAO; PEPE, 2011, p. 4)

Os autores aplicam uma crítica ao método de *machine learning* no âmbito do seu método enraizado em uma pesquisa empírica psicométrica já fundamentada. Eles admitem que o método de aprendizagem de máquina tem um rendimento ótimo para uma grande quantidade de dados. Porém, as mensagens do *microblog* (como o Twitter) em questão pode trazer desafios específicos para esse tipo de abordagem, argumentando-se que a análise de sentimento em redes sociais flutua em função de instantes emocionais (*trending topics*, por exemplo), sendo mais eficiente compreender tais flutuações a partir de técnica de análise sintáticas, permitindo a análise do sentimento dentro da esfera social que cada um dos perfis se encontra.

ALGORITMOS E MODOS DE CARTOGRAFAR SENTIMENTOS

A análise empírica neste trabalho foi aplicada em dois momentos do processo de *impeachment*: 15 de março de 2015, dia do primeiro grande ato *pró-impeachment*, e no dia 27 de agosto de 2016, quando o processo é finalizado no plenário do Senado, varrendo a análise tweets (*data set*) contendo os termos “*impeachment*”, “*Fora Dilma*” e “*Fora PT*”. Esses dados foram extraídos diretamente da API do Twitter.

O segundo passo, o processamento dos dados coletados, foi o mais complexo e consistiu em três etapas: (1) criação do léxico de cada conjunto de dados; (2) validação do léxico formado de forma a abranger o maior número de tweets e que retorne resultados mais precisos; (3) após a criação e validação do léxico, utilizá-lo

sobre os *data sets* para realizar as análises de sentimentos genéricos, sentimentos específicos e polaridades.

Um exemplo estatístico retirado do léxico criado a partir dos dois *data sets* pode ser visto abaixo nas Tabelas 1 e 2. Os graus de cada palavra e *hashtag* variam entre -1 e 1, sendo -1 contra o *impeachment* e 1 a favor do *impeachment*. Tais graus foram levantados de forma humana, ou seja, classificando-se os tweets, um por um, em que as palavras foram encontradas.

Tabela 1: As palavras e *hashtags* mais recorrentes do dia 15 de março de 2015.

Ordem	Palavras (15/03/2015)	Grau	Hashtags (15/03/2015)	Grau
1	dilma	1.00	#impeachmentdilma	0.94
2	impeachment	0.81	#foradilma	0.98
3	brasil	1.00	#forapt	0.95
4	pt	1.00	#impeachment	0.96
5	rua	0.88	#vemprarua15demarco	1.00
6	manifestação	1.00	#vemprarua	1.00
7	pedindo	1.00	#vaiadilma	1.00
8	brasília	0.86	#vemprarua15demarço	1.00
9	melhor	1.00	#imaginaseadilma	1.00
10	país	1.00	#tchaudilma	1.00

Fonte: Elaborado pelos autores.

Tabela 2. As palavras e *hashtags* mais recorrentes do dia 27 de agosto de 2016.

Ordem	Palavras (27/08/2015)	Grau	Hashtags (27/08/2015)	Grau
1	impeachment	-0.04	#impeachment	0.29
2	dilma	0.43	#golpe	-1.00
3	venezuela	-0.24	#senadovotenão	-1.00
4	janaína	-0.37	#geraldoprado	-1.00
5	juízo	-0.09	#jn	1.00
6	processo	-0.14	#foratemer	-1.00
7	golpe	-0.40	#jornalnacional	1.00
8	pt	0.38	#juizamentodedilma	-0.50
9	janaina	-0.07	#eutoposdvcommafia	0.00
10	brasil	-0.13	#brasilreprovaadilma	1.00

Fonte: Elaborado pelos autores.

Pode-se perceber nas tabelas valores diferenciados em ambos os dias coletados. Em um primeiro momento, no dia 15 de março de 2015, na Tabela 1, podemos ver uma

prevalência forte do movimento contrário ao governo Dilma, no qual quase todas as palavras e *hashtags* mais recorrentes têm um grau bastante elevado, senão máximo, a favor do *impeachment*. Já no dia 27 de agosto de 2016, dia em que se consolidou o processo de impedimento, vemos uma tabela mais controversa, porém, tendendo para uma polaridade contrária ao *impeachment*.

Neste trabalho, as análises de sentimentos genéricos e de polaridade foram realizadas usando o Emolex, proposto em Mohammad e Turney (2013). O Emolex é uma grande base de dados de termos em inglês associados a emoções por meio de anotações manuais. Ela é focada sobre as emoções de alegria, tristeza, raiva, medo, confiança, desgosto, surpresa e antecipação, defendidas por muitos como as emoções básicas do ser humano (PLUTCHIK, 1980). A partir dessas emoções, foram também elaboradas as polaridades de positivo, negativo e neutro. Os termos para o Emolex foram cuidadosamente escolhidos para incluir as palavras mais frequentes da língua inglesa, como substantivos, verbos, adjetivos e advérbios. A análise das anotações foi extensiva, buscando responder questões como: o quanto uma palavra pode estar associada a um sentimento e quantas emoções podem estar associadas a um mesmo termo (MOHAMMAD; TURNEY, 2013). Por exemplo, na base de dados, a palavra “*unhappy*” (infeliz, em português) está associada aos sentimentos de raiva, desgosto e tristeza (Ribeiro et al., 2016).

Neste trabalho foi usada a ferramenta desenvolvida em Ribeiro et al. (2016), que utiliza o dicionário Emolex para a análise de sentimentos. Porém, esse dicionário foi traduzido para o português, para assim realizar a análise de sentimentos dos *tweets* coletados. Um passo importante antes da utilização do Emolex foi reduzir cada palavra presente nos comentários dos *tweets* para a sua forma canônica, ou seja, a forma como ela aparece no dicionário. Para isso foi utilizado o *software* Cogroo (CENTRO DE COMPETÊNCIA EM SOFTWARE LIVRE, 2017).

De modo complementar, uma segunda etapa da nossa metodologia foi verificar se os comentários dos *tweets* eram favoráveis ou contrários ao *impeachment*, o que chamamos aqui de análise de sentimentos específicos. Para isso, um novo dicionário de palavras foi elaborado usando um algoritmo de otimização muito utilizado em ciência da computação e engenharias, chamado de algoritmo genético (HOUCK; JOINES; KAY, 1995).

O algoritmo genético busca encontrar de forma heurística a melhor solução para um problema, sendo muito utilizado para resolver problemas cuja resolução seja difícil ou até mesmo impossível por meios analíticos. O algoritmo genético parte do princípio de que existe uma população inicial de indivíduos, e esses indivíduos são representados pelos seus cromossomos genéticos, em que cada indivíduo corresponde a uma possível solução do problema a ser resolvido. Em um passo seguinte, ocorre a operação de avaliação, que mede de alguma forma a capacidade do indivíduo de resolver o problema (HOUCK; JOINES; KAY, 1995).

Um exemplo simples de como algoritmo genético funciona é mostrado a seguir. Imagine que seja desejado encontrar o valor máximo da função $f(x_1, x_2) = -x_1^2 - x_2^2 + 1$, onde os valores de x_1 e x_2 devem ser determinados. Por ser um problema simples, sabe-se que o valor máximo da função é 1. O algoritmo inicialmente vai gerar aleatoriamente a primeira geração de indivíduos, como mostrado nas duas colunas mais à esquerda da Tabela 3. Cada linha é um indivíduo, composto por dois valores, x_1 e x_2 , onde cada um corresponde ao gene do indivíduo. A terceira coluna mostra a avaliação de cada indivíduo, onde o quarto indivíduo foi o melhor avaliado. Baseado

nesse resultado, a seleção dos indivíduos é feita: por exemplo, o segundo, quarto e o quinto indivíduos.

As três colunas mais à direita na Tabela 3 mostram os indivíduos gerados a partir dos indivíduos selecionados na primeira geração. Por exemplo, o primeiro e segundo indivíduos foram obtidos por meio do cruzamento de genes do segundo e quinto indivíduos da primeira geração. O quinto indivíduo foi obtido do cruzamento entre o segundo e quarto indivíduos da primeira geração. No entanto, o segundo gene dele foi alterado por mutação de -1 para -0,5. Como pode ser observado na avaliação da segunda geração, o segundo indivíduo encontrou o valor máximo para a função. Portanto, a solução do problema é $x_1 = 0$ e $x_2 = 0$.

Tabela 3. Exemplo de funcionamento de algoritmo genético.

1º Geração			2º Geração		
x_1	x_2	Avaliação	x_1	x_2	Avaliação
1	1	-1	1	-1	-1
0	-1	0	0	0	1
-1	-1	-1	1	0,5	-0,25
0,5	0,5	0,5	0,5	0	0,75
1	0	0	0,5	-0,5	0,5
1	0,5	-0,25	0	0,5	0,75

Fonte: elaborado pelos autores.

Explicado o procedimento básico de funcionamento do algoritmo genético, a seguinte metodologia foi aplicada para a criação do dicionário de palavras para análise de sentimentos específicos:

- a) Inicialmente, um conjunto de tweets foi selecionado de cada *data set* e foi realizada anotação manual sobre o sentimento expresso em cada comentário.
- b) A partir das anotações, as palavras presentes em cada tweet foram analisadas, buscando encontrar as palavras mais associadas a cada sentimento. Para este fim, foi utilizado um algoritmo genético proposto em Houck, Joines e Kay (1995). Ele foi aplicado para ponderar cada termo de acordo com a sua relevância para cada sentimento. Os valores associados para cada termo estão dentro do intervalo $[-1, 1]$, em intervalos de 0,1, sendo os de valor negativo associados a palavras contra o *impeachment* e os de valor positivo a palavras favoráveis. O valor -1 indica alta associação com o posicionamento contrário, enquanto 1 indica alta associação com favorável. O valor zero indica que a palavra não é relevante para nenhum dos dois posicionamentos.
- c) Para avaliar a qualidade da solução de cada indivíduo, foi realizado o seguinte procedimento. Para cada tweet anotado manualmente, é realizada a soma das palavras de acordo com os valores associados pelo algoritmo genético. Por exemplo, na frase “*impeachment* já #foradilma”, se o algoritmo forneceu os valores *impeachment* = 0,5, já = 0 e #foradilma = -1, o resultado é -0,5,

sendo associado a opinião contrária ao *impeachment*. Porém, se os valores forem *impeachment* = 0,3, já = 0 e #foradilma = 1, o resultado é 1,3, sendo a frase associada a favorável ao *impeachment* e, portanto, mais apropriado ao contexto da frase.

- d) A melhor solução encontrada pelo algoritmo genético sobre os dados anotados é utilizada para analisar o resto dos *tweets* nos *data sets*.

Eventualmente pode aparecer a condição neutra na análise, o que pode significar uma de duas situações: (1) nenhum termo do dicionário foi encontrado no comentário do *tweet*; (2) os termos contrários e favoráveis acabaram se anulando, tornando difícil identificar qual foi o posicionamento expresso no comentário. Em ambos os casos, assume-se a situação neutra, pois não é possível identificar claramente o posicionamento.

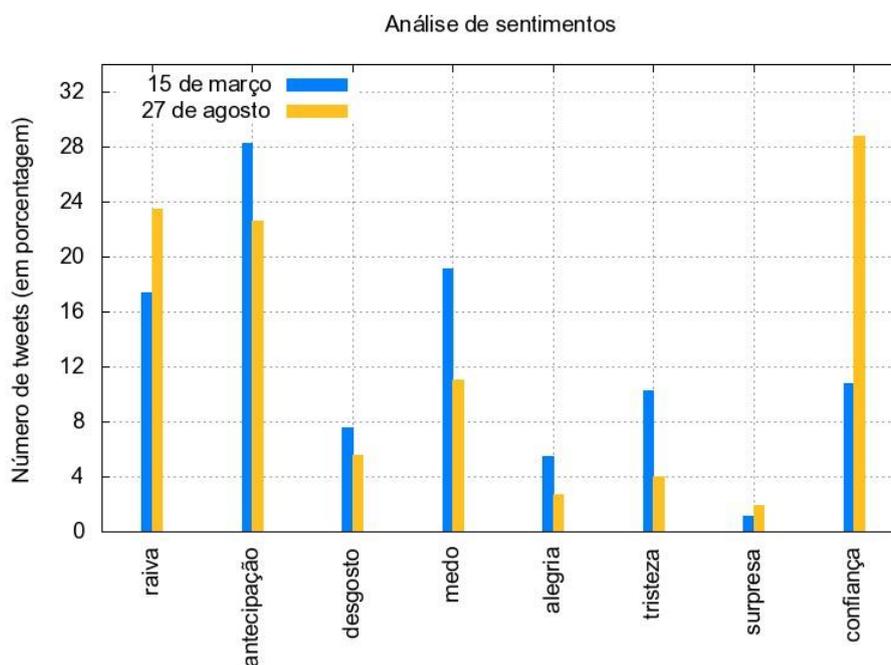
O IMPEACHMENT DE DILMA ROUSSEFF: MEDIÇÃO DE SENTIMENTO EM ANÁLISES TEXTUAIS

Antes de aplicar qualquer método de análise sobre os *data sets*, inicialmente foi realizado algumas etapas de pré-processamento. A primeira etapa foi a remoção de *retweets*, por considerar que o uso desse artefato normalmente pode ser contaminado pela atuação de robôs e militantes, inflacionando a popularidade de uma determinada mensagem.

Após essa etapa, o *data set* do dia 15 de março de 2015 possuía 31.296 *tweets*, enquanto que o *data set* do dia 27 de agosto de 2016 possuía 7.500 *tweets*. A etapa seguinte foi aplicar o Cogroo para obter a forma canônica das palavras e assim poder aplicar a metodologia proposta neste trabalho.

Inicialmente foi aplicado o Emolex sobre os dois *data sets* para análise de sentimentos genéricos e de polaridade. O Gráfico 1 ilustra os resultados dos sentimentos genéricos sobre os dois *data sets*, em que cada posição no eixo horizontal é equivalente a um dos oito sentimentos, e no eixo vertical é apresentado o percentual de *tweets* dos *data sets* que apresentaram determinado sentimento. Os valores estão em porcentagem para tornar comparável a análise dos dois *data sets*. Os resultados dos *data sets* de 15 de março e 27 de agosto são apresentados em azul e laranja, respectivamente.

Gráfico 1. Análise de sentimentos genéricos.

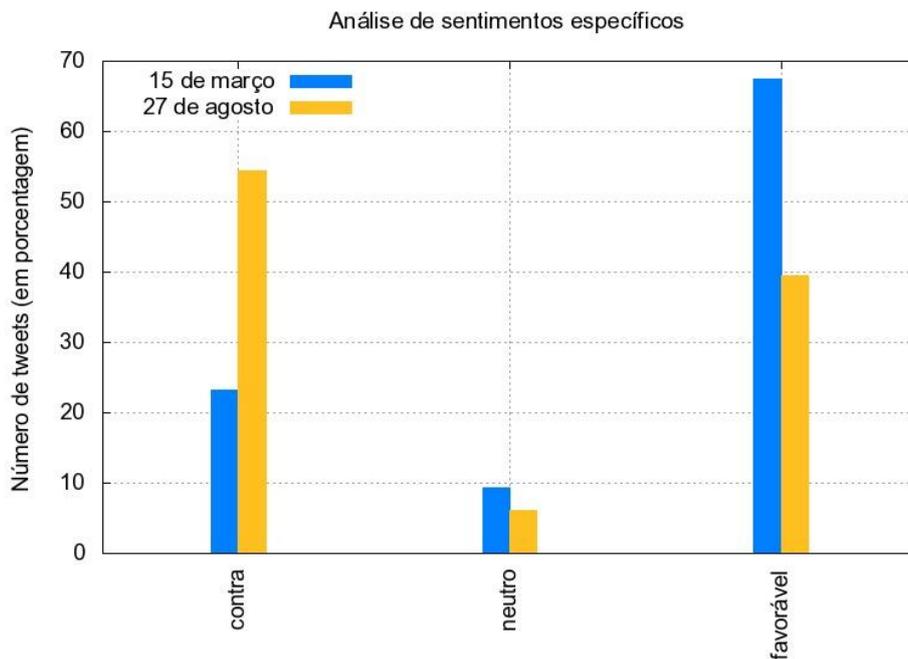
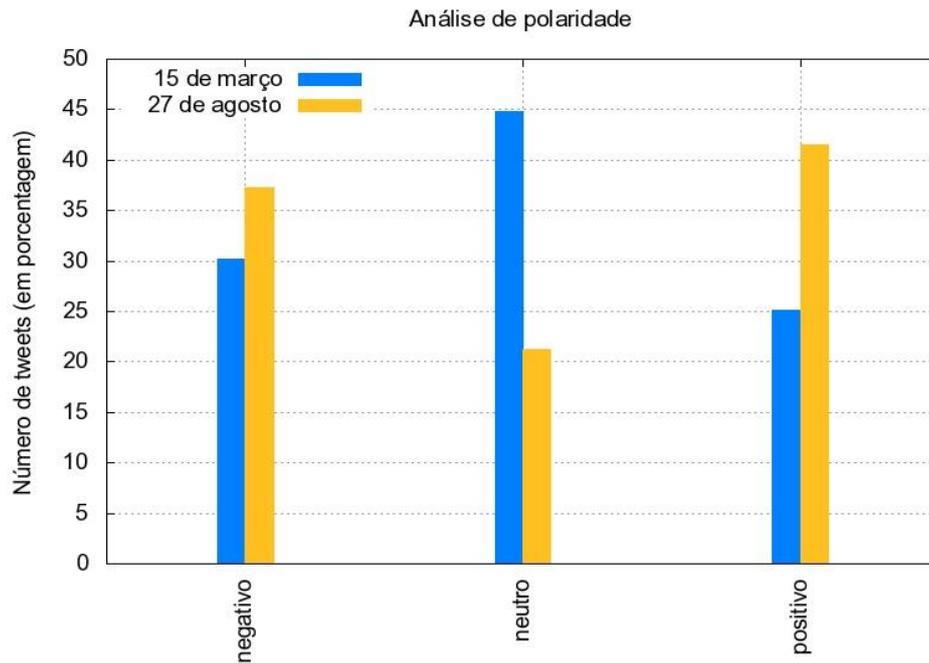


Fonte: elaborado pelos autores.

Como pode ser observado no Gráfico 1, os tweets do dia 15 de março apresentaram em maior destaque os sentimentos de antecipação, medo e raiva, ao passo que nos tweets do dia 27 de agosto há uma maior predominância dos sentimentos de confiança, antecipação e raiva. Observa-se que certos sentimentos são poucos expressivos nas bases de dados, como surpresa e alegria. Em 2015, o tom maior da emoção no Twitter expressa um sentimento de ansiedade, um desejo de antecipação do futuro amparado na vontade de fazer cair o governo Dilma, revelando um padrão emocional que articula a mensagem de rua a uma pressão por qualquer tipo de mudança social. Por mais que muitos discursos apontem o ódio como um *modus operandi* dos protestos anti-Dilma, na verdade, a predominância foi a ânsia pela deposição da presidenta, o que pode revelar um padrão emocional de conflito quando mobilizações ocorrem contra instituições. Já, em 2016, a confiança na derrota da presidenta, num clima de "crônica da morte anunciada", estimulou uma certeza no impeachment ao mesmo tempo que uma raiva acumulada se expressava ali, num jogo ambíguo em que a raiva e a comemoração confiante pela derrota co-habitavam a conversação virtual.

O Gráfico 2 ilustra os resultados da análise de polaridade e possui estrutura similar ao da Figura 1. Pode-se observar que no *dataset* do dia 15 de março há uma maior predominância de neutralidade, enquanto que as polaridades negativa e positiva estão equilibradas. Por outro lado, no *dataset* do dia 27 de agosto existe uma menor quantidade de tweets neutros.

Gráfico 2. Análise de polaridade.



Fonte: elaborado pelos autores.

Importante destacar que embora o Emolex apresente um grande potencial para a análise de sentimentos em textos, em especial para textos curtos iguais aos de tweets, ele não foi capaz de analisar de forma conclusiva todos os tweets. Por exemplo, no *data set* de 15 de março, houve 10.284 tweets, cerca de 32% dos *data sets*, que ficaram com sentimentos indefinidos, pois não havia palavras nos tweets

presentes no dicionário do Emolex. O mesmo foi observado para o *data set* de 27 de agosto, porém em menor quantidade: 500 tweets (cerca de 6% do *data set*). Esses resultados não foram apresentados nos Gráficos 1 e 2 para não distorcer a análise dos resultados. Trabalhos futuros podem ser guiados no intuito de enriquecer o dicionário de palavras do Emolex, de forma a torná-lo mais abrangente.

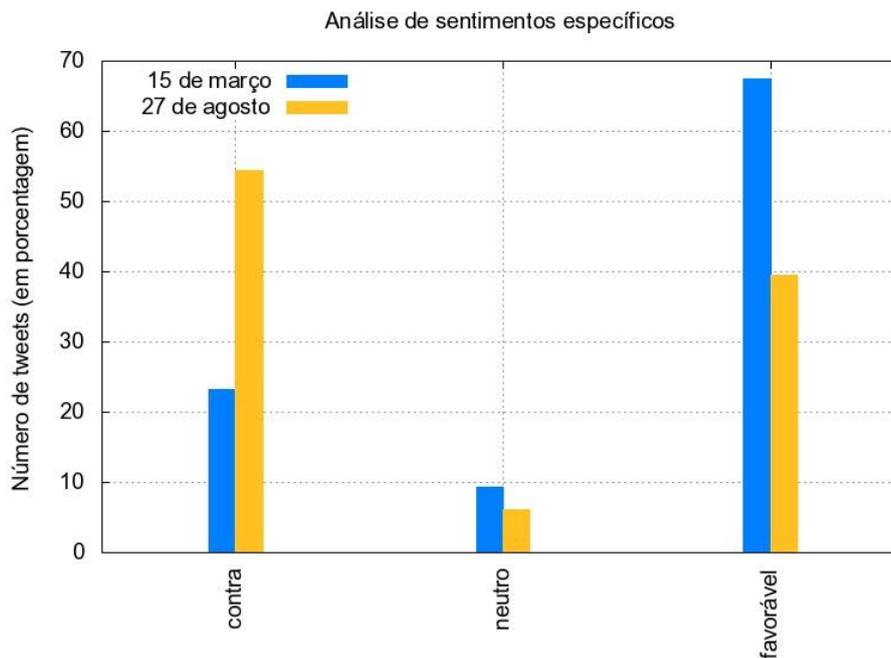
Vimos que o Gráfico 1 revelou que o sentimento que predomina, em 2015, quando o *impeachment* se apresenta como uma reivindicação das ruas, é o da antecipação. Por antecipação, é preciso entender o comportamento emocional da audiência em imaginar uma previsibilidade conclusiva de um fato. Nesse sentido, o desejo de se fazer inevitável a queda da Dilma foi mecanismo mais forte desse comportamento naquele período. Essa análise traz a possibilidade de fundar uma hipótese para futuros trabalhos: movimentos políticos de rua forjam a antecipação de uma irredutibilidade de sua causa, como um modo de fazer com que ela seja encarada, no espaço público, como uma inevitabilidade. Assim, no caso das manifestações do dia 15 de março, no Twitter, o inevitável seria a saída da presidenta Dilma. O modo de exclamar essa inevitabilidade era propagar continuamente a antecipação de enunciados endereçados à ideia que o governo Dilma estava em seus últimos momentos.

Chama também a atenção que os sentimentos de raiva e medo sejam as emoções mais propagadas depois da antecipação. Esse é um dado importante para compreender como os discursos dos “*haters*” inflam a dinâmica política de conversação, criando uma tendência para a comunicação política, a multiplicação de postagens cujo valor esteja na divisão da opinião. Mesmo sendo um movimento vitorioso nas ruas e no Parlamento, a intensidade de emoções mais negativas demonstra o papel fundador do estilo dos “*hater*”, em pauta a dinâmica institucional que marcou o *impeachment* de Dilma, que acabou sendo, no dia 27 de setembro, o objeto muito maior de uma confiança (como observado no Gráfico 2), de uma certeza, de uma inevitabilidade contida nas estratégias dos perfis que militaram nas ruas e nas redes contra o governo de Dilma.

A análise a seguir foi realizada a partir do dicionário formado pela metodologia deste trabalho. O Gráfico 3 exibe os resultados obtidos e possui estrutura similar aos dos gráficos anteriores. Observa-se que a maior parte dos tweets do dia 15 de março (cerca de 67%) são favoráveis ao *impeachment*, e uma margem bem reduzida é contrária. Por outro lado, no dia 27 de agosto há uma predominância de tweets contrários ao *impeachment*, embora exista um equilíbrio maior entre os tweets contrários e favoráveis. O resultado obtido sobre os *data sets* é compatível com os resultados obtidos nas anotações manuais, já que nas anotações foi obtido uma grande concentração de tweets favoráveis no dia 27 de agosto e um cenário mais equilibrado do que o dia 15 de março. Em ambos os *data sets*, o percentual de neutralidade foi baixo. Essa mudança no clima das redes revelou-se como uma dinâmica importante para próximos estudos, já que o domínio do termo *#impeachment* foi modificado. Se ele era uma palavra hegemônica pelos movimentos conservadores de rua desde 2015, diante da certeza do impedimento ser uma operação política institucional para neutralizar a ação da presidenta Dilma em dar prosseguimentos às investigações contra os partidos envolvidos em escândalos de corrupção, um conjunto de manifestações da sociedade passou a contextualizar o *impeachment* como um golpe parlamentar. Isso desde o 17 de abril, quando a Câmara vota pelo afastamento da presidenta, por meio de articulação política do então presidente da Câmara, Eduardo Cunha (posteriormente preso por corrupção).

As redes políticas, antes dominadas por uma polarização entre os favoráveis e os contrários ao impedimento da presidenta Dilma, foram rompidas por uma saraivada de críticas ao sistema político brasileiro no Twitter. Na época, a intensidade de “ao vivo” na cobertura televisiva gerou um efeito reverso à intenção dos executivos das TVs. No lugar de gerar ondas massivas de mobilização social pró-*impeachment* nas ruas, acabou por gerar um esvaziamento delas e também uma alta indignação nas redes contra o apodrecimento da representação política da Câmara dos Deputados. Alimentou um imaginário crítico, ajudando a estimular posts que denunciavam como uma maioria parlamentar completamente desqualificada e despreparada votava um assunto tão importante num regime democrático, um afastamento presidencial. A “onda” contra a Câmara apenas começou: escrachos contra deputados, visualizações de processos de corrupção, alto impacto na imprensa internacional (em alguns casos, chamando-os de palhaços e corruptos), campanhas e correntes em redes sociais e mobilizações de rua. Isso não vai parar. Com a TV aberta, o imaginário contra o impedimento alcançou setores mais populares, que, longe da certeza sobre o golpe, reforçam o mantra do “é tudo corrupto contra uma presidenta idônea, mas com um governo fraco”. Assim, ao chegar no seu clímax, quando Senado dá o tiro de misericórdia, afastando a presidenta, a onda anti-*impeachment*, pelo menos no Twitter, expressava-se em rota de colisão com a decisão do Congresso, que, blindado contra qualquer manifestação popular, daria a continuidade ao pacto para “estancar a sangria da Lava Jato”, como manifestou o senador Romero Jucá.

Gráfico 3. Análise de sentimentos específicos.

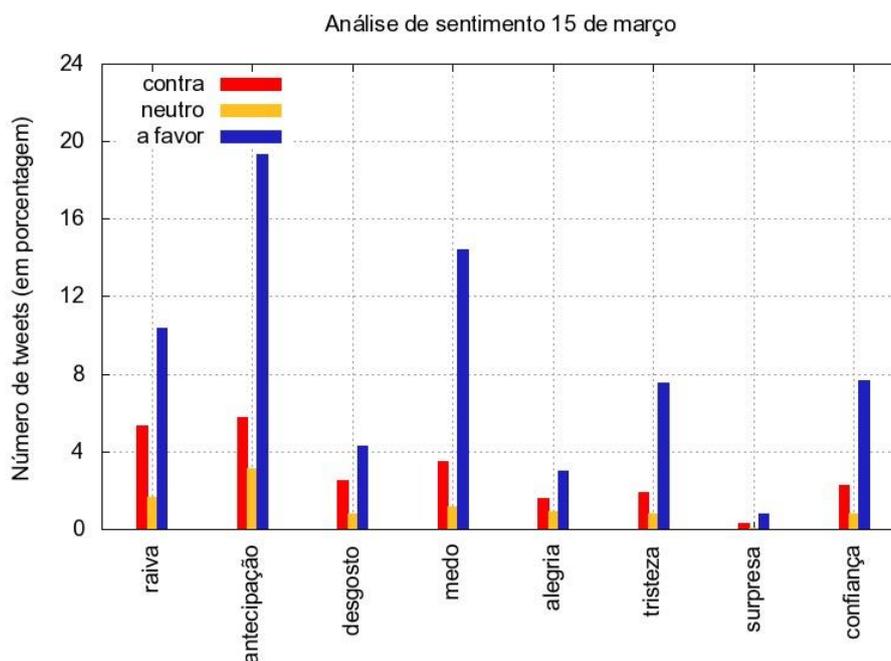


Fonte: elaborado pelos autores.

Os Gráficos 4 e 5 cruzam as informações ilustradas anteriormente nas Gráficos 1 e 3. Em cada gráfico, é mostrada para cada sentimento a divisão obtida por sentimentos específicos de contrário (vermelho), neutro (laranja) e a favor (azul). Os resultados obtidos para os dias 15 de março de 2015 e 27 de agosto de 2016 são ilustrados nos Gráficos 4 e 5, respectivamente. As tabelas são importantes para se analisar, mais

minuciosamente, a complexidade de emoções em cada período de estudo. No dia 15 de março (Figura 4), como já salientamos, o comportamento dos perfis foi mais intensamente relacionado ao sentimento de antecipação, medo e raiva, no que tange aos tweets a favor do *impeachment*. O mesmo comportamento notou-se entre os tweets contrários ao *impeachment*. Esse indicador demonstra que a polarização política refletia, naquele momento, uma relação de ódio mútuo entre os atores políticos, engajando os atores dentro de campos de indignação distintos. Para a oposição a Dilma, a indignação passava por um contexto pós-eleitoral, quando Dilma vence Aécio por uma diferença de 2% dos votos. As manifestações populares de março eram formadas por 80% de eleitores de Aécio, ou seja, uma base oposicionista que não aceitava os rumos do país, já em processo de crise financeira e ajuste fiscal instalados. De outro lado, a base social governista se indignava contra as ruas, por elas não conterem pautas sociais explícitas, mantendo toda a energia na derrubada da presidenta. Por isso ridicularizavam as manifestações, vistas como resultante de um comportamento autoritário de não aceitação das derrotas nas urnas e pela defesa irrestrita da agenda de uma elite empresarial, também alvo de investigações de corrupção na Lava Jato. Enquanto a manifestação oposicionista era taxada de “Marcha dos Coxinhas”, “CarnaCoxinhas”, “Patos Verde-Amarelos”; os protestos de defesa do governo Dilma eram nomeados de “Passeata do Pão com Mortadela”, “Marcha dos Petralhas” etc. O clima hostil já contaminava o quadro político brasileiro, mergulhado no domínio das pautas-bomba, aceleradas por Eduardo Cunha.

Gráfico 4. Análise de sentimentos genéricos e específicos para o dia 15 de março.

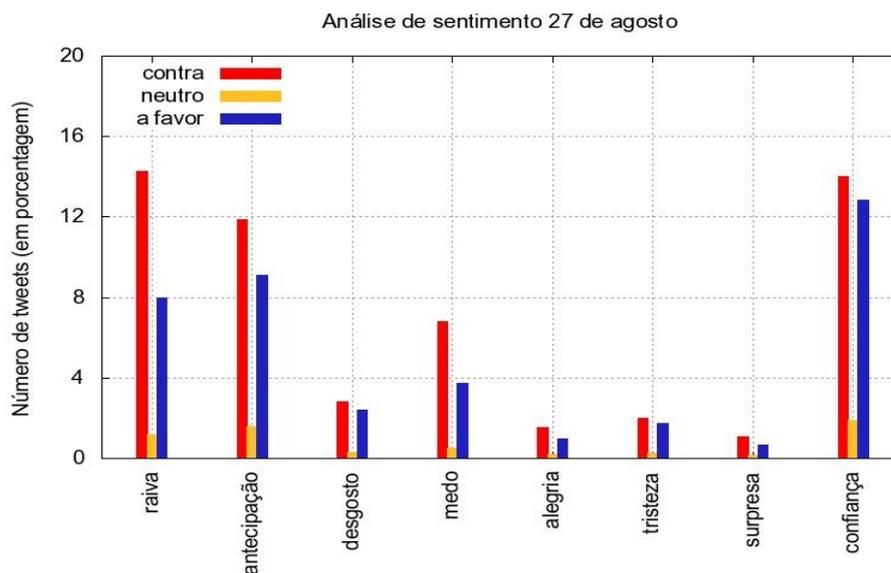


Fonte: elaborado pelos autores.

Quando do impedimento de Dilma, ambos os lados já produziam textos cujo comportamento emocional tendia para a confiança na derrocada do governo petista. Mas são os contrários ao *impeachment* que hegemonizam o tom emocional do período. Como demonstra o Gráfico 5 (reforçando o método empregado na Gráfico 1), os favoráveis ao *impeachment* explicitavam suas certezas da morte anunciada do

governo do PT. No teatro do golpe parlamentar, toda uma dinâmica farsesca de ritos e ditos contrastava com o ato final: a inocência de Dilma aprovada pelos senadores, que concluíram que houve crime de responsabilidade fiscal, mas não havia criminoso a se condenar. O excesso de confiança de golpistas e golpeados é, no polo das audiências participativas do Twitter, retrato também da apatia de segmentos da sociedade brasileira diante de um Congresso cujos parlamentares lutavam pela sua própria sobrevivência jurídica, afinal, em sua maioria, estavam (e ainda estão) envolvidos em maracutaias e atos de corrupção. É claro que entre os contrários ao impeachment, o sentimento de medo passou a se explicitar ainda mais, dado que o governo que o apoiava – e o projeto político que ele simbolizava – estava para acabar, dando vazão um regime de governabilidade baseado na destruição de direitos sociais recém-adquiridos. É interessante como o sentimento de neutralidade praticamente desaparece nos textos desse período, demonstrando que os lados são devidamente formados, que é algo comum em pautas legislativas, quando não há mediação possível senão o contra e o favor a um projeto de lei.

Gráfico 5. Análise de sentimentos genéricos e específicos para o dia 27 de agosto.



Fonte: elaborado pelos autores.

CONCLUSÃO

Nosso objetivo foi demonstrar um trabalho que se remete à testagem de uma metodologia que possa analisar como, em momentos políticos distintos, o ânimo social se revela distinto, em função das polaridades dos sentimentos inscritos em mensagens no Twitter. E como diferentes técnicas algorítmicas de análise de sentimento aponta para a modulação das emoções da atenção *online*, implicando no desenvolvimento de estratégias de comunicação e *marketing* para dar ênfase a determinados comportamentos sociais, com o objetivo de possibilitar que decisões possam ser tomadas, ações realizadas e opiniões consumadas.

Um caminho futuro contempla enriquecer o dicionário do Emolex, de forma a torná-lo mais abrangente e mais voltado a textos escritos em português brasileiro, pois embora o dicionário tenha sido traduzido do inglês para o português, sabe-se que

diferentes povos podem expressar de forma diferente os sentimentos através de textos, e isso pode reduzir a assertividade do dicionário para a língua portuguesa.

O comportamento político contaminado pelo sentimento de raiva e medo confirmou nossa hipótese de que o êxito das manifestações ligados ao *impeachment* de Dilma estava relacionado à eficácia de trolagens e ofensas compartilhadas em torno do alvo dos protestos (no caso, a própria presidenta). A política transformada em ação de *haters/lovers* será a tônica dos movimentos seguintes, focados mais em destruir a reputação de políticos do que em enaltecer suas conquistas. A trolagem ofensiva como mecanismo de pressão política se conjugará com as demonstrações de certezas, já que a percepção predictiva (sentimento de antecipação) se consolidou como principal emoção dessa época. Isso implica dizer que o futuro, nesse período, estava já previsto, sem grandes embaraços, apesar de nele haver também o medo de como vai se suceder – o que chamamos de ansiedade. Isso também implica questionar se os modos de mediação são possíveis em processos políticos baseados no empoderamento popular a partir do comportamento de raiva e medo. É óbvio que há uma certa visão moralista também em torno do ódio, uma vez que ele é mais comumente lido como uma emoção desmedida, quase irracional. Contudo, o comportamento de indignação passa por esse sentimento. E, em certa medida, a métrica do êxito de movimentos de massas precisa levar em consideração se é o ódio o motor das dinâmicas da indignação ou a crítica, a ironia, o deboche, o escárnio e a franqueza. São fronteiras difíceis de cartografar, mas que vale estar em trabalhos que lidam com a explosão de raiva em cenários políticos.

Este artigo demonstra que se fará necessário, nos anos que seguem, uma articulação cada vez mais intensa entre ciências sociais e ciências da computação – o que está se convencendo chamar de ciências sociais computacionais –, à medida que o trato da hiperinformação em plataformas de redes de relacionamento e mídias sociais produz uma “flodagem” incontrolável, que leva o pesquisador a um limite cognitivo, ou seja, à impossibilidade de ver a olho nu todas as conversações geradas nas plataformas, ao mesmo tempo que a presença de um conjunto de dados repetidos em pesados data sets permite-nos averiguar que essa informação carrega uma variedade de sentimentos que podem ser modelados e classificados para fins de entendimento da opinião pública distribuída online. Não há análise social por meio de big data que não possa se deslocar para o small data, à medida que, mesmo num aluvião de conversas, as plataformas estimulam métricas que garantem que o comportamento possa ser reproduzido, por intermédio de RTs, compartilhamentos, matches, likes, favoritação, subscrições, enfim, um conjunto de rastros relacionais que permitem às ciências sociais continuar seu fluxo de entendimento sobre os padrões sociais que estabelecemos, agora digitalmente.

Artigo recebido em 19/07/2017 e aprovado em 05/10/2017.

REFERÊNCIAS

ANDERSON, Chris. The end of theory: the data deluge makes the scientific method obsolete. *Wired Magazine*, 23 June 2008

ARAUJO, Matheus; GONÇALVES, Pollyana; BENEVENUTTO, Fabrício. Comparing and combining sentiment analysis methods. In: ACM CONFERENCE ON ONLINE SOCIAL NETWORKS, 1., Boston, 2013. *Proceedings...* Boston: ACM New York, 2013. Disponível em: <<https://dl.acm.org/citation.cfm?id=2512951>>.

- ASQUER, A. The governance of big data: perspectives and issues. 2013. Disponível em: <<http://ssrn.com/abstract=2272608>> ou <<http://dx.doi.org/10.2139/ssrn.2272608>>.
- BAROCAS, Solon; SELBST, Andrew D. Big data's disparate impact, *California Law Review*, v. 104, 2016. Disponível em: <<http://ssrn.com/abstract=2477899>>.
- BAYM, Nancy K. Data not seen: the uses and shortcomings of social media metrics. *First Monday*, Sep. 2013. ISSN 13960466. Disponível em: <<http://firstmonday.org/ojs/index.php/fm/article/view/4873>>. Acesso em: 1 maio 2015. doi:10.5210/fm.v18i10.4873.
- BOELLSTORFF, Tom. Making big data, in theory. *First Monday*, Sep. 2013. Disponível em: <<http://firstmonday.org/ojs/index.php/fm/article/view/4869>>. Acesso em: 1 maio 2015. doi:10.5210/fm.v18i10.4869.
- BOLLEN, Johan; PEPE, Alberto; HUINA, Mao. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In: INTERNATIONAL AAAI CONFERENCE ON BLOGS AND SOCIAL MEDIA, 5., Barcelona, 2011. *Proceedings...* Barcelona: Association for the Advancement of Artificial Intelligence, 2011. Disponível em: <<https://arxiv.org/abs/0911.1583>>.
- BOYD, Danah; CRAWFORD, Kate. *Six provocations for big data: a decade in internet Time*. 2011. Disponível em: <<http://ssrn.com/abstract=1926431>>; <<http://dx.doi.org/10.2139/ssrn.1926431>>.
- BREUER, A.; FARROQ, B. Online political participation: slacktivism or efficiency increased activism? Evidence from the Brazilian Ficha Limpa Campaign. In: ICA ANNUAL CONFERENCE, 62., San Francisco, 2012. *Proceedings...* Washington, DC: ICA, 2012.
- BRUNO, F. Rastros digitais sob a perspectiva da teoria ator-rede. *Revista Famecos: mídia, cultura e tecnologia*, v. 19, n. 3, p. 681-704, set./dez. 2012.
- BRUNS, Axel. Faster than the speed of print: reconciling “big data” social media analysis and academic scholarship. *First Monday*, Oct. 2013. Disponível em: <<http://firstmonday.org/ojs/index.php/fm/article/view/4879>>. Acesso: 1 maio 2015. doi:10.5210/fm.v18i10.4879.
- BRYANT, Antony; RAJA, Uzma. In the realm of big data ... *First Monday*, Jan. 2014. Disponível em: <<http://firstmonday.org/ojs/index.php/fm/article/view/4991>>. Acesso em: 1 maio 2015. doi:10.5210/fm.v19i2.4991.
- CENTRO DE COMPETÊNCIA EM SOFTWARE LIVRE. CoGrOO: Corretor gramatical acoplável ao LibreOffice. Disponível em: <<http://cogroo.sourceforge.net/>>. Acesso em: 11 fev. 2017.
- DIEBOLD, Francis X. *A personal perspective on the origin(s) and development of 'big data': the phenomenon, the term, and the discipline*. Second version. 2012. (PIER Working Paper, No. 13-003). Disponível em: <<http://ssrn.com/abstract=2202843>> ou <<http://dx.doi.org/10.2139/ssrn.2202843>>.
- _____. *On the origin(s) and development of the term 'big data'*. 2012. (PIER Working Paper, No. 12-037). Disponível em: <<http://ssrn.com/abstract=2152421>; <<http://dx.doi.org/10.2139/ssrn.2152421>>.
- FERGUNSON Andrew Guthrie. Big data and predictive reasonable suspicion. *University of Pennsylvania Law Review*, v. 163, n. 327. Disponível em: <<http://ssrn.com/abstract=2394683>> ou <<http://dx.doi.org/10.2139/ssrn.2394683>>.

FRANÇA, T. C.; OLIVEITA, J. Análise de sentimento de tweets relacionados aos protestos que ocorreram no Brasil entre junho e agosto de 2013. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 3., CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO (CBC), 34., Brasília, 2014. *Anais...* Brasília: CBC, 2014.

GRAY, David C.; CITRON, Danielle Keats. The right to quantitative privacy. *Minnesota Law Review*, v. 98, 2013; *University of Maryland Legal Studies Research Paper*, 2013-23. Disponível em: <<http://ssrn.com/abstract=2228919>>.

HELLES, Rasmus. The big head and the long tail: An illustration of explanatory strategies for big data Internet studies. *First Monday*, Sep. 2013. Disponível em: <<http://firstmonday.org/ojs/index.php/fm/article/view/4874>>. Acesso em: 1 maio 2015. doi: 10.5210/fm.v18i10.4874.

HOUCK, C. R.; JOINES, J. A.; KAY, M. G. A genetic algorithm for function optimization: a Matlab implementation. *NCSU ITR*, v. 95, n. 919, p. 1-14. 1995.

JOH, Elizabeth E. Policing by numbers: big data and the Fourth Amendment. *Washington Law Review*, v. 89, n. 35, 2014. Disponível em: <<http://ssrn.com/abstract=2403028>>.

KAUER, Anderson Uilian. *Análise de sentimentos baseada em aspectos e atribuições de polaridade*. Porto Alegre, 2016. Dissertação Mestrado. Programa de Pós-Graduação em Computação, 2016.

KOLCHYNA, Olga et al. Twitter sentiment analysis: lexicon method, machine learning method and their combination. In: MITRA, G.; YU, X. KOLCHYNA (Ed.). Nova York: Albury Books, 2016. Disponível em: <<https://arxiv.org/abs/1507.00955>>.

LATOURET, B. *Reassembling the social: an introduction to actor-network-theory*. Oxford, UK: Oxford University Press. 2007.

LIU, B. Sentiment analysis and subjectivity. In: INDURKHAYA, Nitin; DAMERAU, Fred J. *Handbook of natural language processing*. Boca Raton, FL: CRC Press, 2010.

MALINI, Fabio. Um método perspectivista de análise de redes sociais: cartografando popologias e temporalidades em rede. In: ENCONTRO ANUAL DA COMPÓS, 25., Goiânia, 2016. *Anais...* Goiânia: Associação Nacional dos Programas de Pós-Graduação em Comunicação, 2016.

MANOVICH, L. *Trending: the promises and the challenges of big social data*. Disponível em: <<http://manovich.net/index.php/projects/trending-the-promises-and-the-challenges-of-big-social-data>>. Acesso em: 1 maio 2015.

MARKHAM, Annette N. Undermining “data”: a critical examination of a core term in scientific inquiry. *First Monday*, Sep. 2013. Disponível em: <<http://firstmonday.org/ojs/index.php/fm/article/view/4868>>. Acesso em: 1 maio 2015. doi:10.5210/fm.v18i10.4868.

MOHAMMAD, S. M.; TURNEY, P. D. Crowdsourcing a word-emotion association lexicon. *Computation Intelligence*, v. 29, n. 3, p. 436-465, 2013

NOVAK, P. K.; SMAILOVIĆ, J.; SLUBAN, B.; MOZETIČ, I. Sentiment of emojis. *PLOS ONE*, v. 10, n.12, 7 dez. 2015. doi:10.1371/journal.pone.0144296.

OBOLER, Andre; WELSH, Kristopher; CRUZ, Lito. The danger of big data: social media as computational social science. *First Monday*, Jun. 2012. Disponível em:

<<http://firstmonday.org/ojs/index.php/fm/article/view/3993/3269>>. Acesso em: 1 maio 2015. doi:10.5210/fm.v17i7.3993.

PELED, Alon. The politics of big data: a three-level analysis. In: EUROPEAN CONSORTIUM OF POLITICAL RESEARCH (ECPR), GENERAL CONFERENCE, Bordeaux, 2013. *Anais...* Bordeaux: [s. n.], 2013. Disponível em: <<http://ssrn.com/abstract=2315891>> ou <<http://dx.doi.org/10.2139/ssrn.2315891>>.

PLUTCHIK, R. A general psychoevolutionary theory of emotion. In: PLUTCHIK, R.; KELLERMAN, Henry. *Emotion: theory, research, and experience*, v.1: Theories of emotion. New York: Academic Press, 1980. p. 3-33.

RIBEIRO, F. N. et al. SentiBench: a benchmark comparison of state-of-the-practice sentiment analysis methods. *Data Science*, v. 5, n. 23, p. 1-29, 2016.

RICHARDS, Neil M. and King, Jonathan H., Three paradoxes of big data *Stanford Law Review Online*, v. 66, n. 41, 2013. Disponível em: <<http://ssrn.com/abstract=2325537>>.

ROBINSON, D. Text analysis of Trump's tweets confirms he writes only the (angrier) Android half. 2016. Disponível em: <<http://varianceexplained.org/r/trump-tweets/>>. Acesso em: 30 jan. 2017.

STEVENSON, Drury D.; WAGONER, Nicholas J. Bargaining in the shadow of big data. *Florida Law Review*, v. 66, n. 5, 2014. Disponível em: <<http://ssrn.com/abstract=2325137>> ou <<http://dx.doi.org/10.2139/ssrn.2325137>>.

TORET, J et al. (Coord.). *Tecnopolítica: la potencia de las multitudes conectadas – el sistema red 15M, um nuevo paradigma de la política distribuída*. Barcelona: Internet Interdisciplinary Institute (IN3): Universitat Oberta de Catalunya (UOC), 2013. (IN3 Working Paper Series).

TUFECKI, Zeynep, *Big data: pitfalls, methods and concepts for an emergent field*. 2013. Disponível em: <<http://ssrn.com/abstract=2229952>> ou <<http://dx.doi.org/10.2139/ssrn.2229952>>.

_____. Engineering the public: big data, surveillance and computational politics. *First Monday*, [S. l.], Jul. 2014. Disponível em: <<http://firstmonday.org/ojs/index.php/fm/article/view/4901/4097>>. Acesso em: 1 maio 2015. doi:10.5210/fm.v19i7.4901.

VENTURINI, T. Building on faults: how to represent controversies with digital methods. *SAGE Journals*, December 5, 201

VIS, Farida. A critical reflection on big data: considering APIs, researchers and tools as data makers. *First Monday*, Oct. 2013. Disponível em: <<http://firstmonday.org/ojs/index.php/fm/article/view/4878>>. Acesso em: 1 maio 2015. doi:10.5210/fm.v18i10.4878.

WALLER Matthew A.; FAWCETT, Stanley E. Click here for a data scientist: big data, predictive analytics, and theory development in the era of a maker movement supply chain. *Journal of Business Logistics*, v. 34, n. 4, Dez. 2013. Disponível em: <<http://ssrn.com/abstract=2339972>>.

WANG, L.; Cardie, C. A piece of my mind: A sentiment analysis approach for online dispute detection. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 52., Maryland, 2014. *Proceedings...* Maryland: The Association for Computing Linguistics, 2014. v 2: Short Papers. p. 693-699.

WHITE, B.; CASTLEDEN, H.; GRUZD, A. Talking to Twitter users: motivations behind Twitter use on the Alberta oil sands and the Northern Gateway Pipeline. *First Monday*, v. 20, n. 1, Jan. 2015.

WU, Kesheng et al. A big data approach to analyzing market volatility. *Algorithmic Finance*, v. 2, n. 3-4, p. 241-267. Disponível em: <<http://ssrn.com/abstract=2274991>> ou <<http://dx.doi.org/10.2139/ssrn.2274991>>.