INTRODUÇÃO À ANÁLISE DE SENTIMENTOS EM REDES SOCIAIS: O LÉXICO CONSTRUÍDO EM DOIS MOMENTOS DO PROCESSO DE IMPEACHMENT[1]

Jean Medeiros[2]; Patrick Marques Ciarelli[3]; Fabio Malini[4]; Luisa Abdoub[5], Nelson Reis [6], Lígia Venturott[7]

RESUMO

Este artigo se propõe em contribuir para a metodologia perspectivista (Malini, 2016) de análise de redes sociais. Esta contribuição se enquadra na identificação e análise dos sentimentos encontrados nas mensagens postadas nestas redes. Os novos meios de mídia permitiram com que os usuários expressassem seus sentimentos, desejos e preferências em relação a diversos assuntos. As controvérsias, formadas pela carga emocional impulsionada pelos usuários, são visíveis em variadas temáticas, principalmente nas mais polarizadas, como na política.

Palavras-chaves: Análise de Redes Sociais, Twitter, Big Data, Análise de sentimentos

Big Data e Análise de sentimento: considerações introdutórias

O fenômeno Big Data se desencadeou a partir da popularização da internet e do constante e acelerado ritmo de desenvolvimento de novas tecnologias. Com uma enorme parcela da população utilizando diariamente as redes, e consequentemente deixando rastros, e a disponibilidade cada vez maior de tecnologias para reter, agrupar e processar esses dados, aqueles que possuem acesso às mesmas prontamente passaram a utilizar essa poderosa

[1]. Artigo apresentado ao Eixo Temático – Movimentos Sociais / Ciberativismo / Resistência do IX Simpósio Nacional da ABCiber.

[2]. Pesquisador na Ufes. É Mestre em Comunicação e Territorialidades (Ufes) e participa do Grupo de Pesquisa Labic (Ufes). Doutrando na UNICAMP. E-mail: jeanmrmedeiros@gmail.com

[3]. Doutor em Engenharia Elétrica, pela Ufes. Professor na Ufes, onde também coordena o Labic (Laboratório de estudos sobre Imagem e Cibercultura). Email: patrick.ciarelli@ufes.br

[4]. PhD em Comunicação e Cultura, pela URFJ. Professor na Ufes, onde também coordena o Labic (Laboratório de estudos sobre Imagem e Cibercultura). Email: fabiomalini@gmail.com

[5] Iniciação Científica na UFES, graduando em Engenharia de Produção. luaboudib@gmail.com

[6]. Pesquisador na Ufes. É graduando em Comunicação Social (Jornalismo) pela Ufes e participa do Grupo de Pesquisa Labic (Ufes). E-mail: nelsonaloysio@gmail.com

[7] Pesquisadora na Ufes. É graduanda em Engenharia Elétrica pela Ufes e participa do Grupo de Pesquisa Labic (Ufes). E-mail: ligia.venturott@gmail.com

ferramenta 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 possiblidade 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, se deu em meados dos anos 90, 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 Boellstorf (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. Por outro lado, alguns autores defendem que as disciplinas já existentes podem perfeitamente compreender e incorporar o fenômeno, tornando assim redundante e desnecessária uma disciplina que apresente o Big Data como objeto de estudo.

Ainda que a utilização do Big Data e o seu entendimento como disciplina sejam controversos em diversos aspectos, o fato é que o fenômeno é real, e tem abrangido as mais diversas áreas, sendo utilizado não apenas por acadêmicos, mas por corporações, governos, jornalistas e políticos. As aplicações da análise de Big Data vão desde a seleção de jogadores de baseball – como explicitado no filme *"Moneyball"* –, passando pela administração pública, vigilância policial, processos de recrutamento e seleção de empregados em organizações; pela advocacia, pelo mercado financeiro, pela gestão da cadeia de suprimentos, até a orientação de campanhas políticas, como ocorreu na reeleição do presidente dos Estados Unidos Barack Obama.

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 apoio ao dado em relação a um dado específico, como informação sobre localização, engajamento e links (Boellstorff, 2013).

Tufekci (2013) aponta que hoje a grande maioria das pesquisas a partir de social data é baseada no Twitter, embora a maior plataforma de mídia social seja o Facebook. Existem vários fatores que favorecem esse quadro, como a disponibilidade de dados, a disponibilidade e popularidade de ferramentas e a facilidade de análise. O Facebook é menos acessível através de suas APIs e grande parte de seus usuários mantém seus dados e interações privados, uma vez que a plataforma permite uma série de configurações diferentes quanto à privacidade. Por outro lado, o Twitter apenas permite que os perfis sejam completamente públicos ou completamente privados, e a vasta maioria de seus usuários os mantêm públicos. Existem várias ferramentas que facilitam a extração de dados do Twitter, enquanto que a API do Facebook é menos conhecida e tem menos ferramentas já prontas para utilização. O Twitter possui poucas funções básicas e textos curtos, o que facilita a estruturação, o manuseio e a análise dos dados extraídos.

Dessa forma, para Tufekci (2013), o Twitter tem sido uma espécie de organismo modelo para pesquisa, pois apresenta características que facilitam sua análise e é desproporcionalmente mais utilizado para pesquisa em comparação com as outras plataformas de mídia social. No entanto, as características dessa plataforma levam os usuários a certos tipos de comportamento, como a breve divulgação de fatos em tempo real, enquanto inibem outros, como a postagem de textos mais logos. Além disso, o Twitter possui funções características para interação, como os *retweets*, que trazem uma série de significados e intenções específicas àquela plataforma. Isso mostra como a análise social dentro do Twitter muitas vezes não pode ser traduzida para outras plataformas, evidenciando a necessidade de se abranger essas pesquisas também para outras mídias sociais.

Tufekci (2013) faz uma ressalva em relação à metodologia utilizada baseada na criação de *datasets* a partir de uma *hashtag*. Esse tipo de amostra é feita a partir do que se chama "seleção da variável dependente", ou seja, os tweets são selecionados para a amostra a partir de uma característica já apresentada – a presença da *hashtag* – e não aleatoriamente, o que pode trazer desvios na análise quando se busca inferir conclusões em cima daquela controvérsia. A análise desse tipo de *dataset*, embora seja uma maneira interessante de

entender a repercussão em cima de um tema, deve ser cuidadosamente realizada, mantendo-se em vista a sua relação com a população e explicitando ainda a cultura da *hashtag*. Isso se deve ao fato de muitas vezes uma *hashtag* ser utilizada apenas por quem tem um posicionamento específico a respeito do fato, tornando outros posicionamentos invisíveis para a análise. Nesse sentido, cabe aos pesquisadores ampliar o universo semântico que permite disparar a coleta de *tweets*, junto à API. Termos conjugados, nomes de perfis e termos co-associados devem ser coletados junto com a *hashtag*, ampliando o domínio de interpretação dos conteúdos do *dataset*.

Outra questão quanto ao Big Data vindo das redes sociais está relacionada à dificuldade em se compreender os limites da amostra e seus denominadores, ou seja, em se determinar quantos usuários no total visualizaram uma interação. De um modo geral, apenas a quantidade de usuários que escolheram de alguma forma interagir com que foi apresentado está explícita e assim, não sabemos quantos escolheram não interagir de forma alguma, o que limita as pesquisas. Hoje, esse número é frequentemente calculado baseado em estimativas de exposição potencial. Por exemplo, o Facebook apresentou uma média de 35% a quantidade de amigos que visualizam as atualizações de um status, embora a variação desse número seja alta.

A comunicação hoje não está limitada a apenas um meio, mas ela é realizada tanto através da internet, e dentro dela através de inúmeras plataformas diferentes, quanto da mídia irradiada, do celular, pessoalmente, e outros, e se apresenta em uma mistura da velha e nova mídia que se mantêm inseparáveis. Assim, a pesquisa de Big Data baseada em apenas uma plataforma de mídia social deve entender sua posição em meio a esse complexo sistema, e buscar justificar suas afirmações mais abrangentes.

As métricas visíveis nas redes sociais (*likes, shares, comments,* número de *followers*, etc.) têm sido amplamente utilizadas para medição de audiência, influência, reputação, engajamento, entre outros. No entanto, é preciso considerar a limitação da utilização dessas métricas, como aponta Baym (2013). Seus resultados podem ser inclinados ou distorcidos por diversos fatores, como os algoritmos que favorecem a exposição de alguns conteúdos em detrimento de outros, a impossibilidade de se demonstrar impacto negativo, não havendo contagem de "*unlike*" ou de "*unfollows*", a utilização de *bots* ou a compra de engajamento, além da não representatividade da amostra das redes em relação à população em geral. A pesquisa de redes sociais deve considerar que a sua própria existência influencia o comportamento dos usuários, por exemplo, quando se faz uma organização deliberada de "twitaços" com o intuito de levar uma *hashtag* aos *trending topics*.

Além disso, o significado do engajamento, embora possa parecer óbvio, é na realidade ambíguo. Ao mesmo tempo em que um *retweet*pode demonstrar afirmação, apoio e concordância, sua intenção pode ser de denúncia, repúdio ou desgosto. A análise de influência a partir dessas métricas, embora limitada, é amplamente realizada em estudos acadêmicos. Fatores como poder de engajamento e atenção seriam melhores descritos pela quantidade de interações, embora as intenções dessas interações se mostrem variadas e complexas.

O campo da análise de sentimentos passou a ser um desafio para entender o comportamento emocional inscrito nesses megadados. A análise de sentimentos se tornou uma ferramenta muito importante para a compreensão das informações que são transmitidas em redes sociais. Com tais análises é possível fazer levantamentos estatísticos e assim ter uma ideia, por exemplo, do grau de aprovação ou reprovação que indivíduos sentem a respeito de um produto, serviço ou evento. As opiniões expressas pela demonstração de sentimentos são úteis para fazer decisões, e isto não é só verdade para os indivíduos, mas também é verdade para as organizações (Liu, 2010). Tais análises também têm relevância na área de política, de forma que um candidato pode mudar o seu discurso e sua forma de agir conforme a reação dos eleitores.

Sabendo a importância deste tipo de análise, este projeto tem como foco fazer a análise de sentimentos em redes sociais, mais especificamente o Twitter. O caminho metodológico e sua complexidade envolvem o processamento de linguagem natural e a modelagem de palavras, textos e *hashtags*. O método usado consiste em duas abordagens pontuais aos sentimentos: sentimentos genéricos e polaridade. Os sentimentos genéricos correspondem aos sentimentos mais comuns entre os indivíduos de diferentes culturas, raças e etnias, como medo, alegria, surpresa, tristeza, etc. A polaridade, por outro lado, indica se o sentimento expresso é positivo, negativo ou neutro. Em ambas as abordagens as palavras são modeladas de uma forma a serem associadas a um grau, como -1, 0 e 1 ou um grau de positividade, neutralidade ou negatividade em relação ao texto analisado.

Assim, após a junção dessas abordagens, pode-se criar diversas métricas qualitativas e quantitativas de análise. Espera-se com isso obter análises mais refinadas, e até prever com maior assertividade as ações expostas em conjuntos de dados obtidos nas redes sociais.

Há uma necessidade de elaborar um estudo apoiado em teorias sociais que servirá como base estrutural para um processo metodológico, sociológico e comunicacional (Shirky, 2009; Castells, 2009; Latour, 2005). Este trabalho, além das teorias sociais, se preza pelos estudos mais aprofundados no âmbito dos movimentos sociais em rede, como o conceito de

‘Smart Mobs’ de Rheingold (2002), ‘Multidão’ do Negri e Hardt (2005), ‘auto-comunicação de massa’ de Castells (2007), entres outros.

O campo da comunicação social tem se dedicado, sobretudo na área da cibercultura, em estudar fenômenos da “sociedade dos Perfis” como objetos cuja essência se manifesta como relacionais, ou seja, formam processos, como espalhamento de notícias ou mobilizações políticas que, disseminados a partir do desencadeamento de *retweets,shares, replies, comments, matches* geram situações e fatos que vão de grandes manifestações sociais a eventos de cultura de fãs nas redes.

Existe uma grande variedade de trabalhos relacionados ao campo de análise de sentimentos e análise de redes sociais em si (Wang and Cardie, 2014; Gerbaudo, 2012; Toret, 2013). Muitos destes trabalhos servem de base para a implementação e acurácia na criação de um método abrangente. O esforço é aglomerar teorias, casos e técnicas necessárias, além de focar na inovação rumo a uma metodologia de análise de redes sociais em *datasets* de movimentos sociais, ou quaisquer outros casos pertinentes.

Neste trabalho são realizadas análises sobre dois *datasets* coletados do *Twitter* em datas diferentes. As duas principais motivações de se utilizar o *Twitter* são: (a) acessar notícias e se inserir em debates (White et al., 2015); (b) a facilidade com que se pode discutir assuntos, no caso da pesquisa em questão, a situação ambiental, com os demais usuários. O *Twitter* acaba por se revelar como um grande espaço para discussão de assuntos importantes entre indivíduos e instituições do mundo inteiro. Tais discussões podem ser referentes e referenciadas por diversos artigos e publicações, validando suas posições através de hiperlinks, imagens e vídeos, etc. Esta rede social é um dos principais terrenos de disputa ocupados por *slacktivists*.

> O slacktivism é um neologismo formado pelas palavras slacker (preguiçoso) e ativismo e é normalmente usada em um sentido pejorativo para descrever atividades cívicas ou políticas que são realizadas online. Algumas dessas atividades imitam as tradicionais formas de participação offline (por exemplo, a assinatura de um e-petição ou doar para uma causa). Outras atividades evoluíram em simbiose com a tecnologia Web 2.0 e estão intrinsecamente ligadas a certas características das plataformas de mídia social. Exemplos incluem o compartilhamento rápido de conteúdo aprovado através das redes, somente clicando em um botão "Like" ou a cópia de conteúdo para os status de uma rede social, a fim de aumentar a consciência sobre uma questão social ou política. (Breuer e Farroq, 2012, p. 3)

Enviesados entre os termos mais comuns no campo da SNA (*Social Network Analysis*, ou Análise de Redes Sociais) encontram-se os *slacktivists*, os *bots*, os *fakes*, etc. O primeiro é um dos mais significativos, como vemos na definição de Breuer e Farroq (2012), e se enquadra em um paradoxo que é muito discutido por sociólogos e analistas de redes sociais.

Tal paradoxo se sustenta pelo fato do "*slacktivism*" apoiar ou enfraquecer os protestos num quadro em que os usuários são puramente "*slacks*", que por sua vez não vão às ruas e não passam de *likes*, de compartilhamentos, interações virtuais, etc. Esta é uma problematização teórica, e por sua vez, também empírica, que deve ser discutida e analisada de um ponto de vista crítico e imparcial sobre o ativismo online.

Tal discussão se faz pertinente, pois os pontos de vista controversos expostos por autores como Barberá et al. (2015), que em seu texto faz uma importante ressalva: as forças implicadas pelos participantes mais periféricos (os *slackitvists*) somam um importante valor ao movimento, se unindo à carne (Hardt e Negri, 2005) presente nos protestos. Dessa forma, gera atenção pública e uma maior visibilidade, principalmente midiática, para o assunto tratado. Nota-se, assim, que a habilidade de disseminação e o poder convocatório desses participantes devem ser levados em consideração pelos que estão mais ativos no protesto. A camada lógica quando somada à camada física se torna o território das manifestações e dos movimentos sociais.

Outro artifício das redes sociais muito utilizado em movimentos, protestos e assuntos políticos, como eleições, são os *bots*. *Bots* são usuários falsos controlados por computador, ou seja, um algoritmo escreve suas mensagens que na maioria das vezes são mensagens de spam. Por ser uma rede social aberta, o *Twitter* permite que suas mensagens sejam postadas de forma automática e através de softwares terceiros. Sendo assim, existem muitos *bots* capazes de *tweetar* através de programas, ou até mesmo, por código, obtendo acesso via API.

Dessa forma, a detecção desses perfis (os *bots*) e a possibilidade de se analisar o *dataset* sem a presença deles é essencial para viabilizar a real troca de mensagens e as reais interações entre as pessoas envolvidas no social construído pelos dados coletados. Exige-se a aplicação de um estudo para o desenvolvimento de métodos de detecção de *bots*, como em Dickerson, Kagan e Subrahmanian (2014) e suas métricas. Entende-se por pertinente a criação ou utilização de métodos capazes de reconhecer e retirar o máximo de usuários *bots* possível, com intuito de analisar casos, com e sem a presença desses perfis robôs.

Toret (2013), sobre o uso das redes sociais diante dos movimentos sociais, discute o conceito de tecnopolítica. "A tecnopolítica baseia-se na compreensão em massa, intuitiva e profunda da capacidade política de se organizar em rede a partir das tecnologias dispostas" (Toret, 2013, p. 43). Tal conceito se aproxima da prática de ativismo online, porém, não totalmente. Segundo o autor, a tecnopolítica não é feita somente de meros *clicks* ou ativismo de âmbito online. Ela é o uso de ferramentas digitais para ter um efeito tanto fora quanto dentro da internet.

Dentre os diversos vieses sociológicos acoplados às discussões, interações e ao ativismo online, junto às formas de como ele vai ao e de encontro ao ativismo "off-line", necessita-se entender esse social definido por essa sociedade em rede e seus modos de viver conectados. Sendo assim, se faz necessário aplicar um método mais consistente quando se analisa redes sociais. As ferramentas de comunicação têm se tornado cada vez mais complexas e estão apresentando um expoente grau de heterogeneidade. Dessa forma, necessita-se abranger os dados extraídos da melhor forma possível, alcançando as diversas camadas e perspectivas expostas nos *datasets*.

Trabalhos Relacionados

Existem diversos trabalhos que relacionam a identificação de sentimentos em textos com postagens de redes sociais. Dentre esta variedade, destacamos três artigos pertinentes para a discussão e a criação deste artigo. O primeiro, de Robinson (2016), em uma análise de sentimentos genéricos, identificando sentimentos como tristeza, medo, alegria, surpresa, etc. Outro artigo entre os estudados é o de Novak et al. (2015) no qual são trabalhados sentimentos de *emojis* de forma polarizada. O terceiro artigo que abordamos realiza análises de sentimentos dos protestos de junho de 2013 no Brasil (França e Oliveira, 2014).

O primeiro artigo trata de analisar os *tweets* proferidos pelo, na época, candidato à presidência dos EUA, Donald Trump. Primeiramente, percebe-se que as postagens são enviadas por dois dispositivos diferentes, um Android e outro iPhone, e o autor trata de descobrir qual dos dois é utilizado pela equipe de candidatura e qual é utilizado por ele mesmo. Ao analisarem ambos *datasets*, entre as diversas métricas impostas, pode-se perceber um índice maior (40~80% acima) de sentimentos negativos nas postagens provindas do dispositivo Android, como desgosto, tristeza, medo e raiva. Através das análises de palavras e *hashtags* foi possível detectar qual dispositivo era o usado por Trump e concluir que a negatividade foi extensamente mais encontrada nos *tweets* pessoais.

Novak et al. (2015), em seu artigo, elaboraram um léxico de *emojis* e ranquearam eles de acordo com a sua respectiva polaridade entre -1 (negativo), 0 (neutro) e 1 (positivo), ou seja, como abordamos neste artigo. O trabalho é feito em cima de uma classificação humana, a partir de 1,6 milhões de *tweets* de 13 idiomas europeus diferentes. Dentre os resultados obtidos, pode-se perceber na análise do léxico criado que os *emojis* são utilizados mais comumente de forma mais positiva do que negativa, ficando com uma média de pontuação total de +0,3.

A terceira referência, com autores brasileiros, é do artigo feito por França e Oliveira (2014). Este trabalho expõe uma análise de *tweets* coletados a partir das *hashtags* disseminadas pela rede social durante os protestos das jornadas de junho de 2013 no Brasil. As *hashtags* foram: #acordabrasil, #vemprarua, #ForaFifa, #ogiganteacordou, #anonymousbrazil, #MPL, #passelivre, #pec37, #mudabrasil, #ChangeBrazil, #anonymousbrazil, #protesto, #foradilma, #protestorj, #protestabrasil, #primaverabrasileira, #forafeliciano, #ocupa, #copapraquem, #protest, #pec33 e #pec99. Após coletar os tweets, foi feita uma classificação humana utilizando um método conhecido como Naive Bayes para realizar a classificação automática de forma probabilística. A análise de sentimentos foi feita por polaridade (negativo ou positivo) e foi aplicada de diversas maneiras como no período de coleta e na localização da mensagem (por estados).

Os trabalhos relacionados abriram um caminho interessante, porém pertinente para que este artigo fosse pensado. A junção de diversos métodos, podendo assim comparar os resultados deles para cada mensagem analisada se tornou em um objeto de análise. Além disso, a união de uma polaridade de um sentimento específico a demais métodos pode nos retornar uma análise mais ampla e minuciosa do(s) *dataset(s)*.

Metodologia

A análise empírica se aplicará 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, dia em que o processo foi julgado na câmara dos senadores e o *impeachment* da presidenta Dilma Rousseff aconteceu. A metodologia, na qual tal análise é apoiada, consiste em quatro passos: mineração, processamento, visualização e análise dos dados.

A mineração dos dados se faz através das coletas feitas no Twitter nos períodos indicados. A obtenção desses *datasets* foi feita pelos termos: *impeachment, fora dilma* e *fora pt*.

O segundo passo, o processamento dos dados coletados, é o mais complexo e consiste de três etapas: (1) criação do léxico de cada conjunto de dados; (2) validação do léxico formado de forma a abranger o maior número de textos (como tweets) e que retorne resultados mais precisos; (3) após a criação e validação do léxico, utilizá-lo sobre os *datasets* para realizar as análises de sentimentos genéricos e de polaridades.

O terceiro passo também se trata de um processamento. Ele se utiliza dos resultados sobre o *dataset* processado para a criação de visualizações: grafos, gráficos, *word clouds*, etc.

Será a partir do cruzamento das informações providas das visualizações que o quarto passo, a análise, será feita.

Um exemplo estatístico retirado do léxico criado em cima dos dois *datasets* pode ser visto abaixo na Tabela 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 os *tweets*, um por um, em que as palavras foram encontradas.

Pode-se perceber nas tabelas presenças diferenciadas em ambos os dias coletados. Em um primeiro momento, no dia 15 de março de 2015, na Tabela 1, podemos ver uma insurgência forte do movimento contrário ao governo Dilma, no qual quase todas as palavras e hashtags mais recorrentes tem 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, vemos uma tabela mais controversa, porém tendendo para uma polaridade contrária ao impeachment. Ou seja, houve uma atividade maior entre os usuários com ideologias contrárias ao impedimento da presidenta.

Como mencionado anteriormente, neste trabalho foi usada uma abordagem baseada em léxico para análise de sentimentos. Em especial, a análise de sentimentos genéricos e de polaridade foi realizada usando o Emolex, proposto em (Mohammad e Turney, 2013). O Emolex é uma grande base de dados de termos em inglês associados a emoções através de anotações manuais. Ela é focada sobre as emoções de alegria, tristeza, raiva, medo, confiança, desgosto, surpresa e antecipação, sendo defendido por muitos como as emoções básicas do ser humano (Plutchik, 1980). A partir destas emoções foram também elaboradas as polaridades de positivo, negativo e neutro. Os termos escolhidos para o Emolex foram cuidadosamente escolhidos para incluir os termos mais frequentes da língua inglesa, como nomes, verbos, adjetivos e advérbios. A análise das anotações foi extensiva, buscando responder questões como: o quanto que uma palavra pode estar associada a um sentimento e quantas emoções podem estar associadas a um mesmo termo (Mohammad e 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, este dicionário foi traduzido para o português, para assim realizar a análise de sentimentos dos tweets coletados.

Um passo importante antes da utilização do Emolex é 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).

Resultados Obtidos com a Metodologia

Antes de aplicar qualquer método de análise sobre os *datasets*, inicialmente foram realizadas algumas etapas de pré-processamento. A primeira etapa foi a remoção de retweets, por considerar que o uso deste artefato normalmente tem mais caráter informativo e não necessariamente transmite informação sobre os sentimentos dos autores.

Após esta etapa, o *dataset* do dia 15 de março de 2015 possuía 31296 tweets, enquanto que o *dataset* do dia 27 de agosto de 2016 possuía 7500 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 *datasets* para análise de sentimentos genéricos e de polaridade. A Figura 1 ilustra os resultados dos sentimentos genéricos sobre os dois *datasets*, onde cada posição no eixo horizontal é equivalente a um dos oito sentimentos, e no eixo vertical é apresentado o percentual de tweets dos *datasets* que apresentou determinado sentimento. Os valores estão em porcentagem para tornar comparável a análise dos dois *datasets*. Os resultados dos *datasets* de 15 de março e 27 de agosto são apresentados em azul e amarelo, respectivamente.

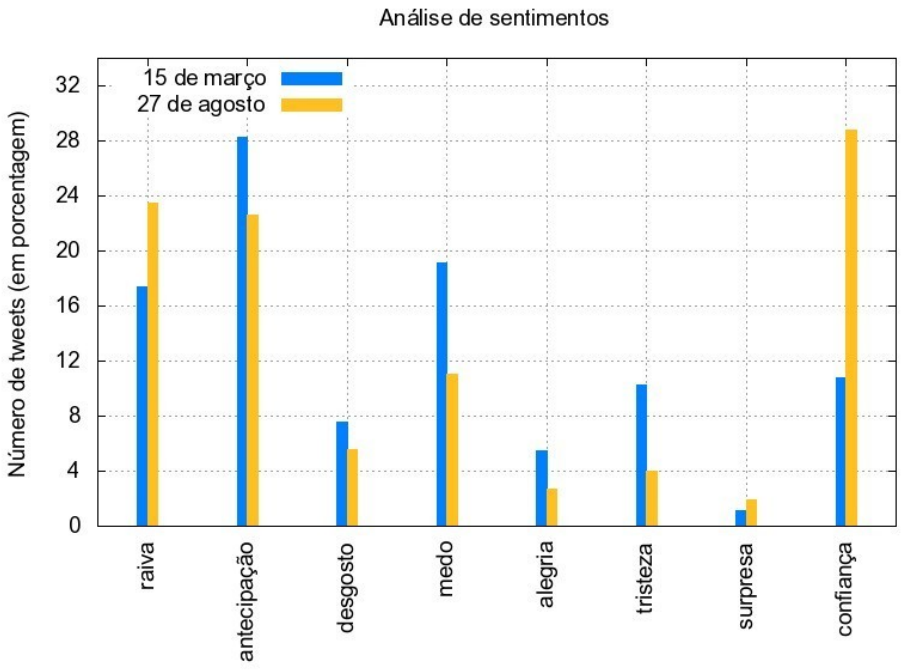
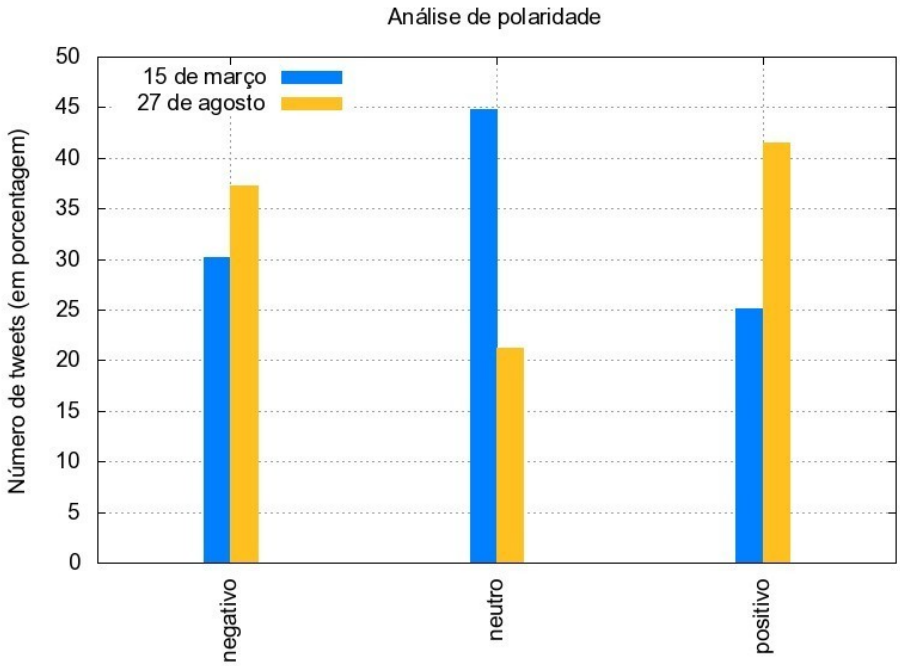


Figura 1 Análise de sentimentos genéricos

Como pode ser observado da Figura 1, os tweets do dia 15 de março apresentaram em maior destaque os sentimentos de antecipação, medo e raiva, enquanto 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.

A Figura 2 ilustra os resultados da análise de polaridade e ela 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 bem equilibradas, com uma ligeira tendência ao negativo. Por outro lado, no *dataset* do dia 27 de agosto existe uma menor quantidade de tweets neutros e uma tendência ligeiramente maior a polaridade positiva.



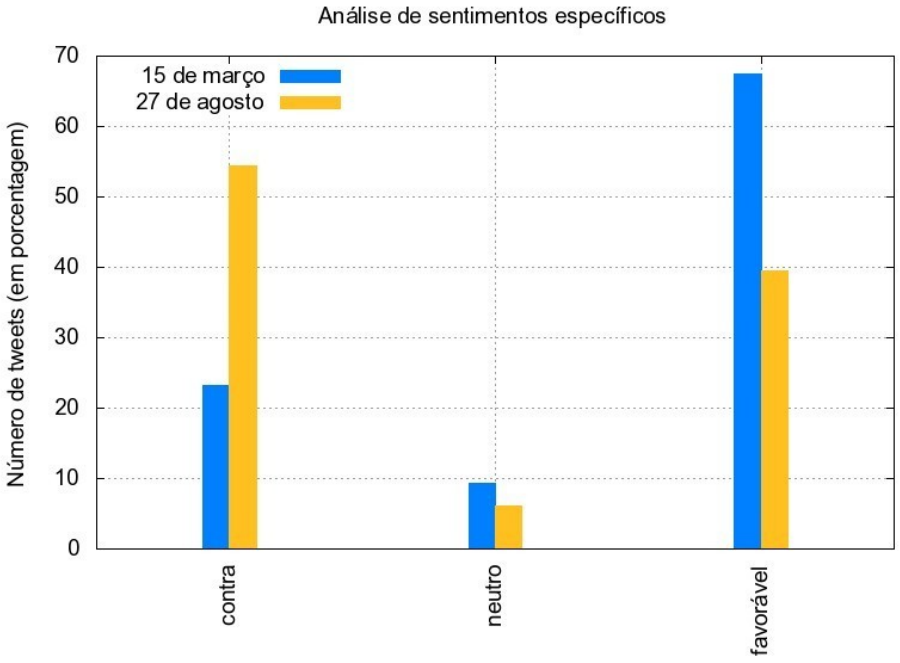


Figura 2 Análise de polaridade

Importante destacar que embora o Emolex apresente um grande potencial para análise de sentimentos em textos, em especial para textos curtos iguais aos de tweets, ele não foi capaz de analisar de forma conclusiva em todos os tweets. Por exemplo, no *dataset* de 15 de março houveram 10284 tweets, cerca de 32% do *dataset*, que ficaram com sentimentos indefinidos, pois não havia palavras nos tweets presentes no dicionário do Emolex. O mesmo foi observado para o *dataset* de 27 de agosto, porém em menor quantidade: 500 tweets (cerca de 6% do *dataset*). Esses resultados não foram apresentados nas Figuras 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 ele ser mais abrangente.

Análise dos Resultados

Vimos que a Figura 1 revelou que o sentimento que predominou, em 2015, quando o *impeachment* se apresentou como uma reivindicação das ruas, foi 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 é 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 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 demostra 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 na Figura 2), de uma certeza, de uma inevitabilidade contida nas estratégias dos perfis que militaram nas ruas e nas redes contra o governo de Dilma.

Conclusão

Nosso objetivo foi demonstrar um trabalho que se remete a testagem de uma metodologia que possa analisar como o ânimo social se revela distinto em diferentes momentos políticos, em função das polaridades expressas nos sentimentos inscritos em mensagens no Twitter. Um trabalho mais minucioso sobre as emoções (alegria, medo, empoderamento, raiva, etc.) será melhor descrito em um futuro artigo a ser apresentado em periódico científico da área.

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 formas diferentes os sentimentos através de textos, e isso pode reduzir a assertividade do dicionário para a língua portuguesa.

Referências bibliográficas

Breuer, A; Farroq, B. **Online Political Participation: Slacktivism or Efficiency Increased Activism? - Evidence from the Brazilian Ficha Limpa Campaign.** 2012 ICA Annual Conference San Francisco. 2012.

Castells, M. **Power Communication.** Oxford: Oxford University Press, 2009.

_____. **Communication, Power and Counter-power in the Network Society.** International Journal of Communication 1 (2007), 238-266, 2007.

Centro de Competência em Software Livre. **CoGrOO: Corretor Gramatical acoplável ao LibreOffice**. Disponível em: <http://cogroo.sourceforge.net/>. Acesso em: 11 fev. 2017.

França, T. C.; Oliveita, J. **Análise de Sentimento de Tweets Relacionados aos Protestos que ocorreram no Brasil entre Junho e Agosto de 2013.** BraSNAM – III Brazilian Workshop on Social Network Analysis and Mining. XXXIV Congresso da Sociedade Brasileira de Computação – CSBC. 2014.

Gerbaudo, P. **Tweets and the streets: Social media and contemporary activism.** Londres: Pluto Press. 2012.

Hardt, M., Negri, A. **Multidão**. Editora Record: Rio de Janeiro, 2005.

Houck, C. R., Joines, J. A., Kay, M. G. **A enetic algorithm for function optimization: a Matlab implementation**. Ncsuie Tr, v. 95, n. 919, p. 1–14. 1995.

Latour, B. **Reassembling the social. An introduction to Actor-Network theory**. Oxford: Oxford University Press, 2005.

Liu, B. **Sentiment analysis and subjectivity.** Handbook of natural language processing, 2010, p. 2:568.

Malini, F. **Um Método Perspectivista De Análise De Redes Sociais: Cartografando Topologias E Temporalidades Em Rede**. 25º Encontro Nacional Compós – Goiânia. 2016.

Mohammad, S. M., Turney, P. D. **Crowdsourcing a word-emotion association lexicon**. Comput Intell 29(3):436-465. 2013

Novak, P. K.; Smailović, J; Sluban, B.; Mozetič, I. **Sentiment of Emojis.** 07 dez, 2015. PLoS ONE 10 (12): e0144296. doi:10.1371/journal.pone.0144296. 2015.

Plutchik, R. **A general psychoevolutionary theory of emotion**. Academic Press, New York, pp 3-33. 1980.

Rheingold, H. **Smart Mobs: The Power of the Mobile Many.** Smart Mobs: The Next Social Revolution. Cambridge, MA: Perseus Publishing. 2002.

Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., Benevenuto, F. **SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods**. Data Science 5(23):1-29. 2016.

Robinson, D. **Text analysis of Trump's tweets confirms he writes only the (angrier) Android half** 09 ago 2016. http://varianceexplained.org/r/trump-tweets/ (Último acesso: 30/01/2017). 2016.

Shirky, C. **Here Comes Everybody: How Change Happens when People Come Together.** London, Penguin Books Ltd. 2009.

Toret, J. **Tecnopolítica: la potencia de las multitudes conectadas. El sistema red 15M, un nuevo paradigma de la política distribuída.** UOC e IN3. Junho, 2013.

Wang, L; Cardie, C. **A Piece of My Mind: A Sentiment Analysis Approach for Online Dispute Detection.** Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 693-699p. Junho, 2014.

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, Vol. 20, N. 1. Janeiro, 2015.