

Abordagens e métodos para o estudo das mídias sociais na comunicação política

Jean Burgess¹

Axel Bruns²

O papel das mídias sociais na comunicação política

Dado o ambiente de mídia distribuído e de multiplataformas em que os meios de comunicação e os cidadãos compartilham e discutem eventos políticos as mídias sociais se fazem sempre presentes na política contemporânea. Conseqüentemente, o estudo do papel das redes sociais na Comunicação Política está agora bem estabelecido nos estudos de Mídia e Comunicação, na Ciência da Informação, tal qual na Ciência Política – como é evidenciado pela produção de volumes multidisciplinares como o *Routledge Companion to Social Media and Politics* (Bruns et al 2016). À medida que pesquisadores em áreas como a comunicação mediada por computador (*computer-mediated communication* – CMC), análises de redes sociais e estudos de internet figuravam entre os primeiros a estudar o uso de mídias sociais (Boyd e Ellison, 2007), os usos políticos e as implicações das mídias sociais passam a ser cada vez mais interessantes e preocupantes em todo o espectro disciplinar em Ciências Sociais. Este interesse generalizado prosseguiu junto à incorporação contínua de plataformas e práticas de mídia social na maquinaria da política e a quase onipresença do uso de mídias sociais entre as populações de muitas nações.

129

¹ Professora de Digital Media e Diretora do Digital Media Research Center na Queensland University of Technology.

² Professor de Digital Media e Diretora do Digital Media Research Center na Queensland University of Technology.

No entanto, as plataformas específicas de redes sociais não são onipresentes, como por exemplo, o *Facebook* e *Twitter*, que têm muita popularidade em muitas partes do mundo, mas são irrelevantes na China onde a atividade de mídia social gira em torno de aplicativos de mensagens instantâneas como *WeChat* e, em menor grau, a plataforma de microblogging *Weibo*.

Mais familiar para os estudiosos da Comunicação Política, as mídias sociais desempenham um papel importante na política formal (eleitoral e estatal). Por exemplo, houve uma grande quantidade de estudos do uso das mídias sociais na campanha Do ex-presidente Barack Obama. Seguindo uma tradição já muito utilizada, os estudiosos de mídia e da comunicação seguem e avaliam o uso dessas novas formas de comunicação política e mediação por candidatos e políticos eleitos, particularmente durante as campanhas eleitorais (para uma visão geral, ver BALDWIN-PHILIPPI, 2017). Como a campanha presidencial dos EUA em 2017 e os eventos políticos no mesmo ano têm mostrado, os modelos de negócios, as operações algorítmicas e as políticas de plataformas tornaram-se cada vez mais problematicamente emaranhadas em torno da política de Estado e da operação de processos democráticos como no caso das eleições.

À medida em que o papel da internet nas campanhas eleitorais junto à democracia deliberativa e o “governo eletrônico” se estabelecia desde a década de 1990, observava-se que uma das características distintivas das mídias sociais residia, por um lado, na forma como ela produz convergências entre a cultura popular e a vida cotidiana e, por outro, nos espaços mais formais da política no que diz respeito a governos e eleições que são mais tradicionalmente entendidos como objetos da “comunicação política”. Esse entendimento leva a novos campos de estudo sobre a “política cotidiana” nas mídias sociais (HIGHFIELD, 2016), incluindo o papel de formas culturais de manifestação vernácula nas políticas públicas como *memes* e *gifs* animados (por exemplo, ver MILNER, 2016).

O papel das mídias sociais em novos movimentos sociais e no ativismo político como um todo tem se mostrado uma área de pesquisa relativamente nova e dinâmica (para uma visão geral ver Poell & VAN DICKJ, 2017). Tem havido uma intensa atenção ao papel das mídias sociais na ecologia midiática em larga escala nos eventos coletivamente e coloquialmente conhecidos como Primavera árabe (BRUNS, BURGESS & HIGHFIELD, 2013); bem como as novas formas de protesto por direitos civis e cultura ativista associada ao movimento *#occupy*

(GERBAUDO, 2017); os elementos de mídia social do movimento dos direitos civis negros organizados em torno da hashtag *#blacklivesmatter* (FREELON et al., 2016) e uma série de campanhas feministas e ambientais.

O Twitter e o crescimento dos “estudos de hashtags”

O Twitter é especialmente popular como um espaço de estudo para a relação entre as mídias sociais e a política. Com efeito, o Twitter tem sido de particular interesse para os pesquisadores devido à sua relativa abertura em termos de acesso a dados e de sua aceitação em vários idiomas e nações. De fato, foi de interesse para nós quando solicitamos financiamento para o estudo relatado nesta coleção de artigos, pois era popular tanto na Austrália quanto no Brasil, e em ambos os países foi usado ao lado do *Facebook* para comunicação política formal e ativismo organizado bem como para o engajamento político informal ou cotidiano.

A relevância do *Twitter* para os estudos de comunicação política deve-se em parte aos recursos comunicativos da plataforma como *tweets* curtos, estruturas de rede que podem ser efêmeras em vez de baseadas em relacionamentos, e seu aspecto em tempo real que se adequa a discussões e debates informais. O *Twitter* também é extremamente popular tanto com jornalistas, atores políticos como ativistas e políticos profissionais. Em grande parte, devido aos recursos de pesquisa da plataforma, especialmente a relativa abertura e acessibilidade de sua Interface de Programação de Aplicações (*Application Programming Interface* – API) que tem sido tão atraente para pesquisadores em Ciências Sociais equipados com a capacidade de realizar análises baseadas em dados usando métodos digitais. No entanto, é importante notar que há desafios significativos para os pesquisadores na área de acesso a dados, pois o modelo de negócios do *Twitter* e as estruturas técnicas de governança se fortaleceram em torno do acesso aberto aos seus dados e esta situação continua a se agravar (BURGESS & BRUNS, 2015; PUSCHMANN & BURGESS, 2014).

Uma grande proporção de estudos do *Twitter* baseia-se em *hashtags*, palavras-chave ou “pacotes” e combinações destes para *tweets* de origem relevantes para um tópico, problema ou evento. Sobretudo em fases anteriores da pesquisa de *Twitter* orientada a dados nos campos da comunicação e da política, essa foi a única maneira confiável de ter um entendimento da interlocução pública

em torno de um determinado evento ou tópico e muitas destas interlocuções foram mediadas por *hashtags*, criando “*públicos de hashtags*” de diferentes tipos e com diferentes propósitos comunicativos (BRUNS & BURGESS, 2015). Contudo, nem sempre foi possível capturar todas ou mesmo a maior parte das interlocuções em qualquer evento em particular já que as *hashtags* “oficiais” para eventos ou tópicos específicos levaram algum tempo a surgir ou várias alternativas permaneceram operando em paralelo. O papel das *hashtags* e sua utilidade para os pesquisadores evoluiu de forma significativa à medida que a plataforma cresceu e há menos consenso entre as comunidades de usuários sobre as *hashtags* ou palavras-chave mais adequadas para se usar, ou se devem usá-las. Além disso, mesmo se as *hashtags* relevantes foram usadas de forma consistente, sempre foi o caso que os tópicos de conversação pudessem se cruzar de um tópico para outro.

Contrariamente, as *hashtags* também podem se tornar menos úteis à medida que os tópicos se tornam mais difundidos. Isso ocorre porque alguns eventos ou temas são tão globais e de tão grande alcance que o uso de *hashtags* para referir-se a eles parece redundante. Por exemplo, quando David Bowie morreu em 2016, um enorme volume de tuítes que estavam preocupados com sua morte não continham uma *hashtag* ou mesmo o nome “Bowie” – essa morte de uma celebridade em particular foi entendida como uma experiência comum entre os usuários em que eles assumiram claramente que não precisavam se referir explicitamente por uma *hashtag* para transmitir o seu significado – uma característica comum de “eventos de mídia social” muito generalizados (BURGESS, MITCHELL & MÜNSCH, 2018). Por conseguinte, é importante que as abordagens orientadas por dados para a comunicação política no *Twitter* se movam para além dos “estudos de *hashtag*” e considerem também outras abordagens mais sistemáticas e abrangentes.

Abordagens baseadas na população para se estudar política no Twitter

Os estudos baseados em *hashtags* são capazes de investigar apenas um subconjunto autosseleccionável de discussões políticas no *Twitter*: eles capturam apenas aqueles *tweets* cujos autores fizeram a escolha consciente de incluir uma *hashtag* específica em seus *tweets*. Na prática, eles geralmente cobrem apenas as *hashtags* que os próprios pesquisadores tomaram conhecimento antes ou durante o evento político em estudo, ignorando outras *hashtags* menores ou alternativas em suas

análises. Dentro dessas limitações, esses estudos são indubitavelmente úteis; no entanto, é igualmente importante explorar abordagens adicionais e alternativas que possam ampliar ou complementar essas análises baseadas em *hashtags*. Uma dessas abordagens é exposta no presente estudo.

Especialmente no contexto de eventos políticos formais que apresentam um conjunto definido de atores centrais – como eleições, referendos, sessões parlamentares, debates televisivos etc. – uma abordagem baseada na população pode gerar uma perspectiva muito diferente sobre os usos políticos do *Twitter*. Nesses contextos, um conjunto menor ou maior de atores oficialmente reconhecidos podem ser facilmente identificados a partir de fontes oficiais – dos dois candidatos restantes na rodada final de um processo de eleição presidencial aos vários membros de um Legislativo nacional; ou dos jornalistas que moderam um debate televisionado ao corpo da imprensa levado a cobrir o parlamento nacional. Para esses atores é possível determinar suas contas oficiais no *Twitter* (e outras presenças de mídia social), na medida em que existam com base em seus perfis oficiais e/ou funções de pesquisa de redes sociais disponíveis. Este processo é auxiliado também pela tendência de que os atores oficiais busquem a verificação de suas contas no *Twitter* a fim de distinguir essas contas oficiais de potenciais contas paródicas que também podem existir.

Este procedimento resulta em uma lista que associa atores políticos por meio de suas identificações pessoais no *Twitter* e que pode ser usada tanto para rastrear os *tweets* públicos dessas contas quanto para todos os *tweets* públicos dos usuários regulares do *Twitter* que mencionam ou retuítam essas contas. Neste caso, esta abordagem baseia-se em uma prática generalizada entre os usuários do *Twitter* – incluindo usuários comuns, jornalistas profissionais e contas institucionais – para mencionar os políticos não apenas por seus nomes (“Donald Trump”), mas sim pelo identificador oficial do Twitter (“@ *RealDonaldTrump*”). Esta prática não é de modo algum universal (um ponto que retornamos abaixo) e os conjuntos de dados resultantes dessa abordagem baseada na população também são autosseleccionados, pois a experiência mostra que o conjunto de dados dos *tweets* assim capturados é substancialmente diferente da abordagem baseada em *hashtags* descrita acima e, portanto, complementa essa abordagem de forma útil, iluminando uma diversidade de práticas de discussão e engajamento político no *Twitter*.

De fato, em comparação com os estudos de *hashtags* é importante notar que as abordagens baseadas na população oferecem uma compreensão consideravelmente mais ampla sobre práticas discursivas no *Twitter*, pois capturam especificamente o compartilhamento de informações (através do *retweeting*) e o debate tópico (através da @replicação). Os conjuntos de dados baseados na população são especialmente valiosos, principalmente nas avaliações da atenção pública referente aos diferentes atores políticos no *Twitter* e do alcance que as mensagens desses atores recebem através da ampliação dos outros usuários do *Twitter* por *retweeting*. Os métodos de análise de rede também podem ser aplicados a esses conjuntos de dados a fim de determinar as estruturas de interação entre os próprios atores políticos bem como entre eles e seu grupo mais amplo de interactantes e pode apontar para a existência de divisões dentro da rede ao longo de linhas ideológicas distintas.

Essas abordagens foram realizadas com êxito em vários estudos até o momento. Grant et al. (2010), por exemplo, rastrearam *tweets* dos políticos e para os políticos em nível local, estadual e federal na Austrália, ao longo de um período de dez meses entre 2009 e 2010; Nuernbergk et al (2016) estudaram *tweets* dos parlamentares e para todos os membros do parlamento alemão antes e durante a campanha eleitoral federal de 2013; Bruns (2017) rastreou *tweets* dos deputados e para todos os deputados e candidatos na campanha eleitoral federal australiana de 2013. Trabalhando com um conjunto mais limitado de grandes contas, Bruns & Highfield (2016) rastrearam os *tweets* dos candidatos e para as contas dos candidatos a presidente e vice-presidente nas eleições presidenciais dos EUA em 2012, assim como as das principais contas das campanhas, enquanto Recuero et al. (2016) fizeram o mesmo para os principais candidatos nas eleições presidenciais brasileiras de 2014. Cada estudo encontrou diferenças nas estratégias de *tweets* das próprias contas dos políticos, o que poderia ser explicado, ao menos em parte, pelo posicionamento ideológico dos candidatos e dos partidos, sua afinidade pessoal e organizacional com as mídias sociais em geral, e o *Twitter* em particular e suas perspectivas para o próximo concurso eleitoral. Outros estudos, entretanto, se concentraram em diferentes conjuntos de atores: Lee et al. (2015), por exemplo, examinaram os *tweets* por uma seleção de jornalistas sul-coreanos, enquanto Bruns et al. (2014) procuraram estabelecer um conjunto abrangente de dados de todos os *tweets* públicos postados por cerca de quatro milhões de contas australianas do *Twitter*, independentemente da demografia pessoal, da profissão ou

dos papéis políticos. Essa pesquisa também encontra uma variação considerável nos padrões de *tweets* entre diferentes subconjuntos da população que está sendo rastreada, o que pode ser explicado de forma variada por fatores demográficos, interesses, aptidão tecnológica, contextos comunicativos e informativos ou desenvolvimentos externos.

Considerações metodológicas

O presente estudo persegue esta abordagem baseada na população para um contexto brasileiro rastreando os *tweets* de um número substancial de atores políticos e de mídia durante mais de um ano. Isso cria uma série de desafios metodológicos consideráveis que devem ser abordados aqui com algum detalhe. Primeiro, a fonte preferida de informações abrangentes sobre as atividades comunicativas públicas *das e para* as contas selecionadas no *Twitter* é, de fato, a Interface de Programação de Aplicações do Twitter (API), que – dentro de limites conhecidos – retorna dados estruturados e bem ordenados sobre atividades de *tweets* mediante solicitação. Para recuperar as atividades de *tweets* de contas específicas e os *tweets* direcionados para essas contas (incluindo @menções e retuítés), dois comandos de API diferentes devem ser realizados: um que rastreia as atividades de um conjunto de contas do *Twitter* identificadas por sua exclusiva identificação de usuário numérica (ID) e um segundo que rastreia as menções dessas contas tratando os nomes das contas como um conjunto de palavras-chave e capturando quaisquer *tweets* que contenham essas palavras-chave.

Nós estabelecemos primeiramente um conjunto de contas brasileiras no *Twitter* para serem rastreadas para então geramos uma lista de seus nomes de conta no *Twitter*, foram selecionados perfis públicos de usuários que de alguma forma se mostram relevantes tanto na esfera pública, quanto na esfera pública interconectada, que trataram do impeachment e das eleições municipais de São Paulo, e também segundo um critério de pluralidade quanto ao tipo de ação e discurso que possuem, assim, criamos os seguintes grupos temáticos: Mídia; Movimentos sociais; Políticos e partidos; Jornalistas; Celebidades; Humor. Utilizamos a ferramenta de linha de comando *t* (MICHAELS-OBBER, 2014) para consultar sistematicamente todas essas contas para suas informações de perfil público completo, incluindo suas IDs de usuário. O comando *t* usa a solicitação da

API de usuários/pesquisa para gerar uma lista de metadados de perfil separada por vírgulas que contém as IDs de usuário. Na medida em que o presente estudo não utilizou nenhuma informação de perfil além dos próprios IDs, esses metadados poderiam ser explorados mais detalhadamente para determinar vários padrões em toda a população da conta estudada aqui, incluindo datas de criação de conta, taxas de *tweets*, comportamentos etc. Hanusch & Bruns (2017) seguem essa abordagem – sem também rastrear e analisar seus *tweets* – para uma população de quase 4.200 contas de jornalistas australianos visando estabelecer padrões comuns de autoapresentação no *Twitter* dentro desse grupo profissional.

Para o presente estudo utilizamos a ferramenta de captura e análise de código aberto do *Twitter* (*Twitter Capture and Analysis Toolkit* – TCAT), desenvolvida pelo grupo de pesquisa em Iniciativa de Métodos Digitais (*Digital Methods Initiative* – DMI) da Universidade de Amsterdã (BORRA & RIEDER, 2014), para acompanhar os *tweets* e a nossa população de contas. Devido à necessidade de capturar ambos *tweets por* e *para* a nossa lista de contas, isso exigiu a criação de pelo menos duas instâncias TCAT: uma em execução no modo “amostra do usuário” (*user sample*) para rastrear a lista de IDs numerados gerados pelo comando *t* e capturar todos os *tweets* postados por essas contas e uma segunda execução no modo “palavra-chave” (*keyword track*) para rastrear a lista de nomes de conta do *Twitter* e capturar todas as @menções e *retweets* dessas contas postadas por outros usuários. Esse procedimento é necessário pois uma única instância do TCAT não pode operar em ambos os modos ao mesmo tempo. (Evidentemente, estudos interessados apenas em postagens ou *tweets* para uma seleção de contas seriam capazes de operar apenas um sistema TCAT no modo de rastreamento apropriado). Neste contexto, também é importante notar que – devido às limitações atuais da API do *Twitter* – uma única instância de “amostra do usuário” do TCAT pode apenas acompanhar até 5.000 IDs de usuários distintos, enquanto uma única instância por “palavras-chave” só pode rastrear até 400 palavras-chave distintas. Portanto, para rastrear uma maior população de contas é também necessário operar TCATs múltiplos que cubram diferentes subconjuntos da população em geral. Contudo, o presente estudo não excedeu nenhum desses limites.

Mesmo o rastreamento de um número limitado de IDs de usuários e de nomes de contas ainda pode resultar em um conjunto de dados de proporções muito significativas se as contas estiverem altamente ativas ou recebendo uma

quantidade substancial de @menções e *retweets*. Isto é bastante provável se, como no nosso caso atual, a população de contas a serem rastreadas contiver uma série de atores cujos perfis atraíam muita atenção (líderes políticos, celebridades, estrelas esportivas etc.), que frequentemente são alvo de @menções ou cujos os *tweets* geralmente são amplamente replicados por seus seguidores. Essas dinâmicas também são consideravelmente mais propensas a estar presentes para atores globalmente reconhecidos ou para contas que operam em países com uma grande base de usuários domésticos do *Twitter*. Um segundo desafio metodológico significativo está relacionado ao processamento e análise dos conjuntos muito amplos de dados usando o TCAT. Nós abordamos esse desafio exportando nossos conjuntos de dados para a potente solução de processamento e armazenamento de dados baseada em nuvem, o *Google BigQuery*.

Além das capacidades analíticas limitadas incorporadas no próprio TCAT, o conjunto de ferramentas oferece uma série de funções de exportação de dados-chave. A função “exportar todos os *tweets* da seleção” gera um arquivo de valores separados por vírgulas ou tabelas que contém os *tweets* capturados e seus metadados associados (como a identificação do *tweet* numérico exclusivo, a data e hora de publicação e a conta de postagem) e serve como base de dados para qualquer análise posterior. Ademais, duas outras funções de exportação fornecem arquivos de valores separados por vírgulas ou tabulações que geram dados adicionais dos próprios textos no *tweet*: a “tabela de *hashtag* de exportação” (*export hashtag table*) resulta em uma lista das marcações associadas a cada *tweet* no conjunto de dados, enquanto a “tabela de menções de exportação” (*export mentions table*) fornece uma lista das contas mencionadas em cada *tweet* e também identifica se cada menção foi um *retweet* ou uma simples menção @. Essas duas tabelas adicionais podem ser anexadas com a tabela principal de *tweets* usando a identificação do *tweet* numérico (ID) como um identificador exclusivo, uma propriedade que utilizamos no *BigQuery*.

Devido ao tamanho do universo de usuários do *Twitter* no Brasil e à visibilidade pública da população em que nos concentramos, nosso rastreamento de contas brasileiras capturou cerca de 25 milhões de *tweets* e resultou em mais de 25GB de dados brutos – bem mais do que poderia ser efetivamente processado e analisado usando soluções de *desktop* convencionais. Isso não é incomum para grandes projetos de pesquisa de mídia social de longo prazo observando processos

comunicativos em larga escala. Esses conjuntos de dados exigem o uso de soluções avançadas de dados baseadas em computação em nuvem que altera o esforço de processamento do cliente para o lado do servidor. Para os nossos propósitos atuais, selecionamos o *Google BigQuery* de uma série de soluções similares oferecidas por provedores concorrentes. O *Big Query* oferece uma interface útil de envio de dados que converte padronizadas exportações de valores separadas por tabelas ou vírgulas em tabelas de banco de dados que podem ser consultadas usando uma versão de consultas padrão SQL (*Structured Query Language*). Isso permite a análise abrangente de conjuntos de dados de redes sociais muito grandes usando consultas manuais ou ferramentas de visualização e análise de dados do lado do cliente.

Para o presente estudo carregamos os três arquivos exportados do TCAT (exportação completa, exportação de *hashtag* e exportação de menções) em três tabelas de banco de dados criadas no *BigQuery*, seguindo o processo descrito em Bruns (2016). Usando a mesma abordagem também criamos uma tabela adicional que não está disponível diretamente do TCAT: aqui, utilizamos um roteiro disponível de Bruns (2016) para identificar todos as URLs incluídas nos textos de tweets que tínhamos capturado e para determinar aquelas URLs encurtadas (*t.co*) para seus destinos finais. Isso resultou em um arquivo de dados adicional que listou os IDs de tweets e os URLs contidos neles que carregamos para uma quarta tabela do *BigQuery*.

Os dados contidos nestas tabelas agora podem ser analisados com consultas SQL manuais iniciadas através da interface do *BigQuery*. A lógica estrutural para tais consultas é sempre uma “junção externa à esquerda da cláusula Join” (*left join*)³ do SQL que conecta a tabela principal de “exportação completa” com as três tabelas adicionais onde uma ID de *tweets* na tabela principal também está presente em uma ou mais das tabelas adicionais. Entretanto, o desenvolvimento e a execução de tais consultas são demorados e propensos a erros, além de não intuitivas para pesquisadores que não estão familiarizados com a sintaxe do SQL.

³ [NOTA DO TRADUTOR]. No servidor SQL, a cláusula JOIN permite que os dados de várias tabelas sejam combinados com base na relação existente entre elas. Por meio dessa cláusula, os dados de uma tabela são usados para selecionar os dados pertencentes à outra tabela. A cláusula LEFT JOIN ou LEFT OUTER JOIN permite obter não apenas os dados relacionados de duas tabelas, mas também os dados não relacionados encontrados na tabela à esquerda da cláusula JOIN. Ver em: <https://www.devmedia.com.br/clausulas-inner-join-left-join-e-right-join-no-sql-server/18930>.

O uso de um gráfico “*front-end*”⁴ que se conecta e gera automaticamente consultas para o *BigQuery* é preferível. Um terceiro componente de nossa configuração metodológica é o uso de tal ferramenta, ou seja, uma sofisticada análise de dados e o emprego do software *Tableau* de visualização.

O software *Tableau* (2017) fornece uma interface de dados padrão para o serviço do *BigQuery*. Depois de se conectar ao *BigQuery* e configurar as relações de junção externa à esquerda da cláusula Join entre a tabela de dados principal de “exportação completa” e as três tabelas adicionais carregadas para o *BigQuery*, os campos de dados contidos nas quatro tabelas ficam disponíveis para uso em sua interface gráfica de usuário (o processo completo para fazê-lo também é descrito em Bruns, 2016). Em comparação com uma consulta manual do banco de dados do *BigQuery*, isso permite uma exploração consideravelmente mais rápida dos dados disponíveis gerando análises e representações visuais de padrões de dados que oferecem informações substanciais sobre a dinâmica das atividades de *tweets* da e para a população de contas que está sendo estudada aqui.

No entanto, deve ser fortemente observado que o resultado final de tal exploração não é necessariamente uma mera representação quantitativa de padrões no conjunto de dados geral dos *tweets*. Preferivelmente, essa análise exploratória pode e também deve ser usada especificamente para identificar subconjuntos particulares de dados que possam beneficiar análises qualitativas ou de métodos mistos de forma muito mais aproximada. Esses subconjuntos podem representar períodos de tempo específicos que exibem padrões de atividade incomuns, grupos notáveis de contas (dentro da população inicial, ou entre os usuários comuns que mencionam as contas selecionadas) ou seleções de *tweets* contendo palavras-chave ou *hashtags* particulares. Esses subconjuntos podem, por sua vez, ser exportados do *Tableau* ou diretamente do *BigQuery* como novos arquivos de dados para serem submetidos a leitura detalhada, codificação manual ou outras formas de análise posterior fora desses pacotes analíticos iniciais.

Nossa abordagem para coletar conjuntos de dados baseados em população em grande escala a partir do *Twitter* conecta três conjuntos de ferramentas principais: reúne dados através da API do *Twitter* usando o *Twitter Capture and Analysis Toolkit* (TCAT); armazena esses dados nos bancos de dados de alto

⁴ [NOTA DO TRADUTOR] - Camada de sintaxe ou de apresentação do lado do cliente que interage diretamente com o usuário.

desempenho fornecidos pelo *Google BigQuery*; e acessa esses bancos de dados para processamento e análise na solução de análise gráfica do *Tableau*. Alternativas para cada uma dessas ferramentas também podem estar disponíveis e podem ser preferíveis em diferentes contextos organizacionais e tecnológicos. A estrutura geral de três passos para a configuração de dados, armazenamento e processamento que apresentamos aqui provavelmente será replicada mesmo que ferramentas específicas sejam trocadas por outras alternativas. A maneira como esta configuração pode ser utilizada em busca de questões de pesquisa específicas é descrita em outro lugar nesta coleção de artigos.

Limitações da abordagem baseada na população

Como este e outros estudos demonstraram, a abordagem baseada na população pode gerar uma série de informações valiosas sobre os padrões de atividade do *Twitter* que complementam e avançam bem além dos estudos de *hashtag* existentes. De forma mais concisa, permite uma análise de padrões relevantes de comunicação pública que não usam as principais *hashtags* relacionadas à questão em discussão. Contudo a abordagem baseada na população não está além de suas próprias limitações. Em primeiro lugar, depende de uma seleção significativa da população de contas a serem estudadas: a omissão de contas que são fundamentais para um determinado tópico ou problema poderia distorcer consideravelmente a análise subsequente. Essas contas ainda podem estar presentes nos dados coletados se elas interagissem diretamente com a população cujas atividades estavam sendo rastreadas (se elas estiveram envolvidas em discussões @replicadas, em contas rastreadas *retwittadas*, ou que foram *retwittadas* por elas). Entretanto, essa presença por associação não capturaria a totalidade das atividades dessas contas. As limitações da API do *Twitter* reproduzidas nas ferramentas de captura do *Twitter*, como o TCAT, significam que normalmente é bastante difícil coletar retrospectivamente dados adicionais que faltam no conjunto de dados primários. Isso torna ainda mais importante começar a partir de uma lista de contas abrangente e muito cuidadosamente construída.

Em segundo lugar, à medida que a abordagem baseada na população captura uma gama muito diferente de atividades em comparação com a abordagem baseada em *hashtags*, também é provável que resulte em um conjunto de dados totalmente exaustivo de tweets relevantes. De forma mais objetiva, como

observado acima, a abordagem baseada na população baseia-se na expectativa de que seja uma prática generalizada para os usuários do *Twitter* se referirem a atores-chave (em política e em outras esferas) por seus nomes de conta no *Twitter*, em vez de simplesmente por seus nomes pessoais – mas esta prática pode não ser igualmente comum em diversos contextos. Por um lado, é possível que o nome da conta oficial de um ator político simplesmente não seja amplamente conhecido, que a sua correta ortografia seja difícil de lembrar, ou que os usuários podem frequentemente confundir outra conta (especialmente por uma conta paródica deliberadamente configurada para imitar o político em questão). Nesses casos, os usuários preferem soletrar o nome do político em seus *tweets* ou podem desviar seus *tweets* para outra conta.

Por outro lado, os usuários podem evitar mencionar diretamente um ator específico pelo nome de sua conta – por exemplo, porque eles não desejam dirigir atenção pública adicional para um político contra quem eles se opõem ou porque temem alguma retaliação, especialmente por declarações críticas. Essa relutância em @mencionar os atores políticos pode ser especialmente pronunciada por contas comparativamente extremistas, por duas razões: os usuários do *Twitter* com visões não extremistas podem querer evitar reconhecer a presença on-line de elementos extremistas e podem estar preocupados com o fato de serem alvo de ataques pelos apoiadores dos extremistas se publicarem críticas públicas de tais atores políticos. Ao mesmo tempo, no entanto, os atores políticos marginais também poderiam atrair muita atividade de seus apoiadores se estes vissem o *Twitter* (e outras plataformas de mídia social) como um meio valioso para promover seus pontos de vista políticos na ausência de uma suficiente e simpaticamente cobertura da mídia convencional, enquanto a atividade de mídia social em torno de atores principais que já estão altamente presentes em mídia impressa e televisiva podem permanecer relativamente silenciosa.

Em qualquer caso, a prevalência destas dinâmicas é provável que seja altamente dependente de uma variedade de fatores contextuais locais, incluindo a situação política atual, a demografia da adoção do *Twitter* em um determinado país e na medida em que os próprios atores políticos estão ativos, usando e encorajando seus seguidores a usar a plataforma para o debate político público. Esses fatores provavelmente diferirão amplamente entre países (e até mesmo entre estados e localidades), bem como ao longo do tempo. O ponto final aqui não

é sugerir uma interpretação padrão dos modelos observáveis em conjuntos de dados populacionais, mas enfatizar que os padrões observados nos dados coletados devem sempre ser interpretados no contexto da abordagem de coleta de dados escolhida – isso é tão verdadeiro para os conjuntos de dados baseados na população que introduzimos aqui como tem sido para os conjuntos de dados baseados em *hashtags* que dominaram grande parte da literatura de pesquisa existente.

Por fim, as limitações básicas da interface de programação de aplicativos do *Twitter*, sobre o qual o TCAT e ferramentas similares de coleta de dados constrói, também podem ser repetidas aqui. Para a coleção total de termos de pesquisa rastreados por uma única instância de TCAT, a API padrão apenas retornará até um por cento do volume global total da atividade de *tweets*, podendo resultar em limitações aos dados que estão sendo capturados. Tais limitações são talvez mais imediatamente sentidas em abordagens de coleta de dados baseadas em *hashtags*: se uma grande *hashtag*, precedida de notícias urgentes difundidas na mídia convencional, estiver presente em dez por cento de todos os tuítes atuais, o TCAT ainda capturará até um por cento de todos os atuais *tweets* e, portanto, apenas um décimo de todos os *tweets* possíveis contendo a respectiva *hashtag*.

Na maioria das vezes é comparativamente muito menos provável que o volume total dos *tweets* que mencionem uma população específica de contas represente mais de um por cento do “*feed*” global de *tweets*, sobretudo se essas contas pertencerem a políticos de baixo escalão em nações relativamente menores. No entanto, as abordagens de rastreamento baseadas na população que incluem líderes mundiais importantes, principalmente durante momentos de grande tensão – @realDonaldTrump e @HillaryClinton na noite de eleições, @dilmabr e @MichelTemer durante o processo de impeachment e protestos subsequentes – podem ocasionalmente superar o limite de um por cento e os conjuntos de dados resultantes estarão incompletos. Isso é inevitável a não ser que sejam usadas ferramentas de coleta de dados comerciais em vez do TCAT ou soluções similares e porque a omissão de *tweets* entregues pela API segue uma seleção essencialmente aleatória não distorcendo indevidamente a análise dos conjuntos de dados incompletos, além de resultar em uma subestimação sistemática do volume total de *tweets* durante esses períodos. Nesse sentido, continua a ser crucial que os pesquisadores identifiquem e observem tais limitações que suas análises apresentam.

Referências

Baldwin-Philippi, Jessica. 'Politics 2.0: Social Media Campaigning.' In **The Sage Handbook of Social Media**, eds. Jean Burgess, Alice Marwick and Thomas Poell. London: Sage, 2017. 527-545.

boyd, danah, and Nicole B. Ellison. 'Social network sites: Definition, History, and Scholarship.' **Journal of Computer Mediated Communication** 13, no. 1(2007), 210-230.

Borra, Erik, and Bernhard Rieder. 'Programmed Method: Developing a Toolset for Capturing and Analyzing Tweets'. **Aslib Journal of Information Management** 66, no. 3 (2014): 262–278. doi:10.1108/AJIM-09-2013-0094.

Bruns, Axel, Gunn Enli, Eli Skogerbo, Anders Olof Larsson, and Christian Christensen (Eds.) **The Routledge Companion to Social Media and Politics**. New York: Routledge, 2016.

Bruns, Axel. 'Tweeting to Save the Furniture: The 2013 Australian Election Campaign on Twitter'. **Media International Australia**, no. 162 (2017): 49–64. doi:10.1177/1329878X16669001.

Bruns, Axel. 'Twitter Analytics Using TCAT and Tableau, via Gawk and BigQuery'. **Mapping Online Publics**, 15 April 2016. <http://mappingonlinepublics.net/2016/04/15/twitter-analytics-using-tcat-and-tableau-via-gawk-and-bigquery/>

Bruns, Axel, and Jean Burgess. 'Twitter hashtags from ad hoc to calculated publics.' In **Hashtag publics: The power and politics of discursive networks**, ed. Nathan Rambukkana. New York: Peter Lang, 2015. 13-28.

Bruns, Axel, Tim Highfield and Jean Burgess. 'The Arab Spring and Social Media Audiences: English and Arabic Twitter Users and their Networks.' **American Behavioral Scientist** 57, no. 7 (2013): 871–898.

Bruns, Axel, and Tim Highfield. 'May the Best Tweeter Win: The Twitter Strategies of Key Campaign Accounts in the 2012 US Election.' In **Die US-Präsidentenschaftswahl 2012: Analysen der Politik- und Kommunikationswissenschaft**, eds. Christoph Bieber and Klaus Kamps. Wiesbaden: Springer Fachmedien, 2016. 425–42.

Bruns, Axel, Burgess, Jean, Banks, John, Tjondronegoro, Dian, Dreiling, Alexander, Hartley, John, Leaver, Tama, Aly, Anne, Highfield, Tim, Wilken, Rowan, Rennie, Ellie, Lusher, Dean, Allen, Matthew, Marshall, David, Demetrious, Kristin, and Sadkowsky, Troy. 'TrISMA: Tracking Infrastructure for Social Media Analysis.' 2016. <http://trisma.org/>.

Burgess, Jean, and Axel Bruns. 'Easy Data, Hard Data: The Politics and Pragmatics of Twitter Research After the Computational Turn.' In **Compromised Data: From Social Media to Big data**, eds. Ganaele Langlois, Joanna Redden and Greg Elmer. London: Bloomsbury, 2015. 93-111.

Burgess, Jean, Peta Mitchell and Felix Münch. 'Social Media Rituals: The Uses of Celebrity Death in Digital Culture.' In **A Networked Self: Birth, Life, Death**, ed. Zizi Papacharissi. New York: Taylor & Francis Group, 2018 (in press).

Freelon, Dean, Charlton D. McIlwain and Meredith D. Clark. 'Beyond the Hashtags: #ferguson, #blacklivesmatter, and the Online Struggle for Offline Justice' **Center for Media & Social Impact, School of Communication, American University** (29 February), 2016. <http://cmsimpact.org/resource/beyond-hashtags-ferguson-blacklivesmatter-online-struggle-offline-justice/>.

Gerbaudo, P. 'Social Media teams as Digital Vanguard: The Question of Leadership in the Management of key Facebook and Twitter Accounts of Occupy Wall Street, Indignados and UK Uncut.' **Information, Communication & Society** 20, no. 2 (2017): 185-202.

Grant, Will J., Brenda Moon, and Janie Busby Grant. 'Digital Dialogue? Australian Politicians' Use of the Social Network Tool Twitter'. **Australian Journal of Political Science** 45, no. 4 (2010): 579-604. doi:10.1080/10361146.2010.517176.

144

Hanusch, Folker, and Axel Bruns. 'Journalistic Branding on Twitter: A Representative Study of Australian Journalists' Profile Descriptions'. **Digital Journalism** 5, no. 1 (2017): 26-43. doi:10.1080/21670811.2016.1152161.

Highfield, Tim. **Social Media and Everyday Politics**. Malden, MA: Polity Press, 2016.

Lee, Na Yeon, Yonghwan Kim, and Jiwon Kim. 'Tweeting Public Affairs or Personal Affairs? Journalists' Tweets, Interactivity, and Ideology'. **Journalism** 17, no. 7 (2015): 845-64. doi:10.1177/1464884915585954.

Michaels-Ober, Erik. 't: A Command-Line Power Tool for Twitter.' **GitHub**, 2014. <http://sferik.github.io/t/>.

Milner, Ryan M. **The World Made Meme: Public Conversations and Participatory Media**. Cambridge, Mass.: MIT Press, 2016.

Nuernbergk, Christian, Jennifer Wladarsch, Julia Neubarth, and Christoph Neuberger. 'Social Media Use in the German Election Campaign 2013.' In **The Routledge Companion to Social Media and Politics**, eds. Axel Bruns, Gunn Enli, Eli Skogerbø, Anders Olof Larsson, and Christian Christensen. New York: Routledge, 2016. 419-433.

Poell, Thomas and José van Dijck. 'Social media and new protest movements.' In **The Sage Handbook of Social Media**, eds. Jean Burgess, Alice Marwick and Thomas Poell. London: Sage, 2017. 546-61.

Puschmann, Cornelius, and Jean Burgess. 'The Politics of Twitter Data.' In **Twitter and Society**, eds. Katrin Weller, Axel Bruns, Jean Burgess, Cornelius Puschmann and Merja Mahrt. New York, Peter Lang, 2014. 43-54.

Recuero, Raquel, Gabriela Zago, and Marco T. Bastos. 'Twitter in Political Campaigns: The Brazilian 2014 Presidential Election.' In **The Routledge Companion to Social Media and Politics**, eds. Axel Bruns, Gunn Enli, Eli Skogerbø, Anders Olof Larsson, and Christian Christensen. New York: Routledge, 2016. 518–530.

Tableau. 2017. <http://www.tableau.com/>

Approaches and methods for the study of social media in political communication

Jean Burgess¹

Axel Bruns²

The role of social media in political communication

Given the distributed, multi-platform media environment across which news outlets and citizens alike share and discuss political events, social media are ever-present in contemporary politics. Consequently, the study of social media's role in political communication is now well-established across media and communication studies, information studies, as well as political science – as is evidenced by the production of multi-disciplinary volumes such as the *Routledge Companion to Social Media and Politics* (Bruns et al 2016). While researchers in areas like computer-mediated communication (CMC), social network analysis and internet studies were among the first to study social media use (boyd & Ellison, 2007), the political uses and implications of social media are increasingly of interest and concern across the spectrum of social science disciplines. This widespread interest has proceeded alongside the ongoing embedding of social media platforms and practices in the machinery of politics, and the near-ubiquity of social media use among the populations of many nations. However, the specific social media *platforms* are not ubiquitous, with Facebook and Twitter popular in many parts of the world but irrelevant in China, where social media activity revolves around instant messaging apps like WeChat and, to a lesser extent, the microblogging platform Weibo.

¹ Professor Digital Media Research Centre. Queensland University of Technology. Brisbane, Australia

² Professor Digital Media Research Centre. Queensland University of Technology. Brisbane, Australia