

ROBÔS, REDES SOCIAIS E POLÍTICA NO BRASIL

**Estudo sobre interferências ilegítimas no debate público
na web, riscos à democracia e processo eleitoral de 2018**

20 de agosto de 2017

Robôs, redes sociais e política no Brasil [recurso eletrônico]: estudo sobre interferências ilegítimas no debate público na web, riscos à democracia e processo eleitoral de 2018 / Coordenação Marco Aurélio Ruediger. – Rio de Janeiro : FGV, DAPP, 2017.

Dados eletrônicos

Inclui bibliografia.

ISBN: 978-85-68823-41-5

1. Políticas públicas. 2. Eleições. 3. Redes sociais on-line. 4. Boatos (Opinião pública). 5. Internet. 6. Robôs. 7. Computação humana. I. Ruediger, Marco Aurélio, 1959- . II. Fundação Getúlio Vargas. Diretoria de Análise de Políticas Públicas.

CDD – 351

ÍNDICE

5	Robôs nas redes sociais
9	O que são e o que fazem?
10	Como podem afetar nossas vidas?
12	Como funcionam?
13	Como podem ser identificados?
16	Como a DAPP está trabalhando para identificá-los?
		CASOS:
17	Eleições 2014
20	Impeachment 2015
21	Eleições Municipais de São Paulo de 2016
22	Greve geral de 28 de Abril de 2017
23	Votação da Reforma Trabalhista no Senado em 11 de julho de 2017
24	Verificação da análise
25	Conclusão
27	Bibliografia

SUMÁRIO-EXECUTIVO

- As redes sociais possuem um evidente mérito de proporcionar debates e amplificar vozes em um espaço que permite grande repercussão.
- Diversos estudos mostram como Twitter, Facebook, entre outras plataformas, ao possibilitarem trocas e estimularem discussões, tornaram-se instrumentos importantes da democracia.
- No entanto, assim como no debate público fora do mundo virtual, as redes também passaram a ser usadas como um fértil espaço para disseminação de informações falsas.
- Contas automatizadas que permitem a massificação de postagens se converteram em uma potencial ferramenta para a manipulação de debates nas redes sociais, em especial em momentos de relevância política.
- Na greve geral de abril de 2017, por exemplo, mais de 20% das interações ocorridas no Twitter entre os usuários a favor da greve foram provocadas por esse tipo de conta. Durante as eleições presidenciais de 2014, os robôs também chegaram a gerar mais de 10% do debate.
- Com isso, o mundo virtual tem permitido a adaptação de velhas estratégias políticas de difamação e manipulação de debates públicos, agora em maior escala.
- Identificar a presença destes robôs e os debates que criam é de fundamental importância para diferenciar quais situações são reais e quais são manipuladas no ambiente virtual. Apenas assim é possível ter compreensão efetiva dos processos sociais originados nas redes.
- O esforço de pesquisa da FGV/DAPP emite um alerta de que não estamos imunes, e que devemos buscar entender, filtrar e denunciar o uso e a disseminação de informações falsas ou manipulativas por meio desse tipo de estratégia e tecnologia. É importante ter atenção e proteger os espaços democráticos inclusive nas redes sociais.
- Às vésperas de início do "ano eleitoral" que definirá o próximo presidente brasileiro, cujas campanhas se anunciam de extremo acirramento, torna-se essencial mapear os padrões de uso desses mecanismos, a fim de evitar intervenções ilegítimas no debate como já verificado em outros países.

ROBÔS NAS REDES SOCIAIS

Adormecidas desde as manifestações à época do impeachment do presidente Fernando Collor de Mello (1992), as ruas voltaram a fazer parte da política brasileira quando o aumento das tarifas de ônibus fez irromper uma onda de protestos no Brasil em junho de 2013. Naquele momento, a estratégia de evitar a apropriação do movimento por parte de partidos políticos - principais alvos da revolta das ruas - e a desconfiança na cobertura das manifestações pela imprensa tradicional **transformaram as redes sociais em um reconhecido espaço de expressão da democracia**, uma vez que assumiram protagonismo como lugar de organização e de difusão de informações.

“

O crescimento da ação concertada de robôs representa, portanto, uma ameaça real para o debate público, representando riscos, no limite, à democracia (...)

”

“

Ao identificarmos robôs operando para um campo, porém não queremos dizer que os atores políticos e públicos ali situados sejam responsáveis diretos pelos robôs a seu favor.

”

De 2013 para cá, não apenas as ruas foram ocupadas por movimentos partidários como as **redes foram inundadas por velhas estratégias políticas de difamação e manipulação de debates públicos**. No entanto, agora elas ocorrem em um meio que permite uma veloz massificação dos discursos de tal forma que coloca em risco a credibilidade do espaço e da informação que nele circula. A tradicional panfletagem partidária, por exemplo, ocupa a mesma timeline de notícias difundidas pela imprensa, assim como boatos e detrações propagados por atores políticos de todo o espectro partidário.

“

Esse esforço de pesquisa aqui apresentado busca emitir um alerta de que não estamos imunes e que devemos nos preocupar em buscar entender, filtrar e denunciar o uso e a disseminação de informações falsas (...)

”

A internet e as redes sociais se converteram em campo importantíssimo, crescente, e dinâmico do debate público e da disputa de narrativas, que levam à busca de hegemonias na política. Essa realidade abre espaço para discussões legítimas e factuais, mas também para discursos corsários, não legítimos e não factuais (fake news).

Somado a este ambiente fértil para a disseminação opiniões, a automatização de ferramentas de publicação possibilitou o surgimento e a propagação de **robôs – contas controladas por softwares se fazendo passar por seres humanos que já dominam parte da vida nas redes sociais e participam ativamente das discussões em momentos políticos de grande repercussão.**

O estudo feito pela FGV/DAPP aponta que esse tipo de conta chegou a ser responsável por mais de 10% das interações no Twitter nas eleições presidenciais de 2014. Durante protestos pelo Impeachment, essas interações provocadas por robôs representaram mais de 20% do debate entre apoiadores de Dilma Rousseff, que usavam significativamente esse tipo de mecanismo. Um outro exemplo analisado mostra que quase 20% das interações no debate entre os usuários favoráveis a Aécio Neves no segundo turno das eleições de 2014 foi motivado por robôs.

Nas discussões políticas, os robôs têm sido usados por todo o espectro partidário não apenas para conquistar seguidores, mas também para conduzir ataques a opositores e forjar discussões artificiais. Eles manipulam debates, criam e disseminam notícias falsas e influenciam a opinião pública postando e replicando mensagens em larga escala. Comumente, por exemplo, eles promovem hashtags que ganham destaque com a massificação de postagens automatizadas de forma a sufocar algum debate espontâneo sobre algum tema.

“

Apontamos com esse esforço dois compromissos da DAPP. O primeiro, é vinculado ao acompanhamento do debate nas redes e atenção pela democracia. O segundo, de esforço contínuo em desenvolver e aprimorar tecnologias de detecção e compreensão desse fenômeno.

”

Ao identificarmos robôs operando para um campo, porém não queremos dizer que os atores políticos e públicos ali situados sejam responsáveis diretos pelos robôs a seu favor. Diversos grupos de interesse podem estar fazendo uso desse tipo de recurso de disseminação de informações. Na verdade, lato sensu, há robôs até operando do exterior. Isso inclusive enseja a reflexão de manipulação não só interna, mas também para além dos campos políticos nacionais, sugerindo a hipótese da possibilidade de até mesmo outros atores, estranhos ao quadro nacional, operarem nas redes esses mecanismos.

O crescimento da ação concertada de robôs representa, portanto, **uma ameaça real para o debate público, representando riscos, no limite, à democracia**, ao manipular o processo de formação de consensos na esfera pública e de seleção de representantes e agendas de governo que podem definir o futuro do país.

Identificar esses robôs torna-se assim um desafio da maior importância, uma vez que sua operação é cada vez mais refinada e capaz de replicar o padrão humano com mais precisão. Diferenciar o dado real do manipulado, na análise dos processos sociais e políticos em pleno curso, é definidor tanto para o governo – cujo processo de tomada de decisão deve estar ancorado em informações qualificadas – como para a sociedade civil, que reverbera a agenda produzida nas redes em debates e ações fora delas.

Por isso a **FGV/DAPP desenvolveu um sistema apurado de identificação de contas suspeitas que atuam como robôs**, gerando conteúdo algoritmicamente, e cujos resultados demonstram o importante papel jogado pelos robôs em momentos-chave da política brasileira recente. Trata-se, pois, de um risco real e comprovado ao debate público legítimo na democracia brasileira.

Mas, apesar de os robôs operarem a favor de agendas específicas, isso não quer dizer que dominem completamente a rede nem que a percepção final da maior parte das pessoas será resultante direta da influência desses dispositivos. O que constatamos, no entanto, é que eles existem, já operam no debate brasileiro, obedecem padrões e buscam influenciar. Sobretudo, esse esforço de pesquisa aqui apresentado busca emitir um alerta de que não estamos imunes e que devemos nos preocupar em buscar entender, filtrar e denunciar o uso e a disseminação de informações falsas ou manipulativas por meio desse tipo de estratégia e tecnologia. Deve-se ter atenção e proteger os espaços democráticos inclusive nas redes sociais.

Considerando que as próximas eleições terão importância central para o país, e supondo que nosso caso não será muito diferente de outras democracias em períodos eleitorais recentes, nos quais ocorreram claras tentativas de manipulação (França, Estados Unidos etc), apontamos com esse esforço dois compromissos da DAPP. O primeiro, vinculado ao acompanhamento do debate nas redes e atenção pela democracia. O segundo, de esforço contínuo em desenvolver e aprimorar tecnologias de detecção e compreensão desse fenômeno.

A primeira fase deste estudo, aqui apresentada, concentrou-se em momentos políticos de alta repercussão nas redes nos últimos três anos: ¹as eleições de 2014, ²o impeachment de Dilma Rousseff, ³as eleições municipais de 2016 e ⁴a greve geral de 2017. **A análise levou em consideração múltiplas características e metadados que denunciam a presença de contas suspeitas.**

O estudo do uso de robôs no período analisado já demonstra de forma clara o potencial danoso dessa prática para a disputa política e o debate público. Uma das conclusões mais evidentes nesse sentido é a concentração dessas ações em polos políticos localizados no extremo do espectro político, promovendo artificialmente uma radicalização do debate e, conseqüentemente, minando possíveis pontes de diálogo entre os diferentes campos políticos constituídos. Outro elemento flagrante é o “inchamento” de movimentos políticos que são, na realidade, de dimensão bastante inferior. Somados, **esses riscos e outros representados pelos robôs, são mais do que o suficiente para jogar luz sobre uma ameaça real à qualidade do debate público no Brasil e, conseqüentemente, do processo político e social definidor dos próximos anos.**

Marco Aurélio Ruediger
Diretor da FGV/DAPP

O QUE SÃO & O QUE FAZEM?

Importante meio de comunicação, informação e construção de conexões, as redes sociais são parte cada vez mais significativa do dia a dia das pessoas. Estudos realizados pelo Pew Research Center mostram, por exemplo, que a maioria dos adultos nos Estados Unidos (62%) se informa por meio das redes sociais. No entanto, 64% afirmam que as notícias falsas que circulam nas redes causam “confusão” sobre fatos e acontecimentos diários. No Brasil, a Pesquisa Brasileira de Mídia 2016, realizada pela Secretaria Especial de Comunicação Social (Secom) da Presidência da República, revela que 49% das pessoas já se informam pela internet, uma fatia em rápido crescimento. É nesse ambiente de “confiança”, mas de alta circulação de informações duvidosas que os robôs se proliferam.

A princípio, as contas automatizadas podem até contribuir positivamente em alguns aspectos da vida nas redes sociais. Os chatbots (chats operados por robôs), por exemplo, agilizam o atendimento a clientes de empresas e, em alguns casos, até auxiliam refugiados a processarem seus pedidos de visto. **Porém, o número crescente de robôs atua na verdade com fins maliciosos. Os robôs sociais (social bots) são contas controladas por software que geram artificialmente conteúdo e estabelecem interações com não robôs. Eles buscam imitar o comportamento humano e se passar como tal de maneira a interferir em debates espontâneos e criar discussões forçadas.**

Com este tipo de manipulação, os robôs criam a falsa sensação de amplo apoio político a certa proposta, ideia ou figura pública, modificam o rumo de políticas públicas, interferem no mercado de ações, disseminam rumores, notícias falsas e teorias conspiratórias, geram desinformação e poluição de conteúdo, além de atrair usuários para links maliciosos que roubam dados pessoais, entre outros riscos.

Ao mesmo tempo, as redes sociais se tornaram parte integral não apenas da vida pessoal dos cidadãos, mas da sua atividade política e da atuação dos seus representantes. Os partidos e demais movimentos de representação social também aproveitam o espaço para engajar eleitores assim como para atacar opositores e pautar debates em torno de seus interesses. Nesse caso, é comum observar **o uso orquestrado de redes de robôs (botnets) para gerar um movimento em determinado momento, manipulando trending topics e o debate em geral.**

Estas ações têm sido identificadas em grandes eventos da política internacional, como as eleições americanas de 2010, a eleição de Donald Trump em 2016 e o plebiscito de saída do Reino Unido da União Europeia, o Brexit. No Brasil, o cenário não é diferente: ações orquestradas de robôs ocorreram em momentos-chave da política nacional, como a aprovação da Reforma Trabalhista, a greve geral de 2017, as eleições de 2014, o debate sobre o Impeachment e a eleição municipal de São Paulo em 2016, como mostraremos à frente.

COMO PODEM AFETAR NOSSAS VIDAS?

Ao interferir em debates em desenvolvimento nas redes sociais, robôs estão atingindo diretamente os processos políticos e democráticos através da influência da opinião pública. Sua ação pode, por exemplo, **produzir uma opinião artificial, ou dimensão irreal de determinada opinião ou figura pública**, ao compartilhar versões de determinado tema, que se espalham na rede como se houvesse, dentre a parcela da sociedade ali representada, uma opinião muito forte sobre determinado assunto (Davis et al., 2016). Isso acontece com o compartilhamento coordenado de certa opinião, dando a ela um volume irreal e, conseqüentemente, influenciando os usuários indecisos sobre o tema e fortalecendo os usuários mais radicais no debate orgânico, dada a localização mais frequentes dos robôs nos polos do debate político.

Os perfis automatizados também promovem a desinformação com a **propagação de notícias falsas e campanhas de poluição da rede**. Robôs frequentemente usam as redes sociais para reproduzir notícias falsas com o objetivo de influenciar determinada opinião sobre uma pessoa ou tema, ou poluir o debate com informações reais, porém irrelevantes para

a discussão em questão. Esta ação, que conta com o compartilhamento de links como principal mecanismo de propagação, tenta evitar ou diminuir o peso do debate sobre determinado assunto. Para isso, os robôs geram um número enorme de informações, que chegam até os usuários simultaneamente às informações reais e relevantes, que acabam tendo seu impacto diminuído. Assim, a atuação de robôs não apenas dissemina notícias falsas, que podem ter efeitos nocivos para a sociedade, mas também busca ativamente impedir que os usuários se informem de maneira adequada.

Outra estratégia comum dos perfis automatizados é o compartilhamento de links maliciosos, que tem como fim o roubo de dados ou informações pessoais. Essas informações podem ser usadas para a criação de novos perfis robôs que tenham características que auxiliem estes robôs a iniciarem conexões nas redes com usuários reais, como fotos de perfil. Uma ação comum e que costuma gerar suspeita sobre a atuação de robôs é a marcação por parte de um usuário desconhecido

em um link reduzido, sem identificação clara do seu conteúdo. Esses links podem também, além de roubar informações pessoais para uso na própria rede social, direcionar o usuário para notícias falsas ou sites que usarão o número de acessos para ampliar sua influência na rede (Wang, 2010).

Também já foram detectadas ações de robôs com o objetivo de manipular o mercado de ações. Isso ocorre quando redes de robôs são colocadas em funcionamento para gerar conversas que envolvam de maneira positiva determinada empresa ou tema, manipulando, assim, os sistemas das corretoras de acompanhamento do debate nas redes. Desta forma, ações em questão podem ser valorizadas com base em um otimismo forjado pela ação de robôs.

Um caso recente deste tipo de ação envolveu um debate gerado por robôs nas redes com relação a uma empresa de tecnologia chamada Cynk. Os algoritmos automáticos de compra e venda de ações identificaram esse debate e começaram a fazer transações com ações da empresa, cujo valor de mercado aumentou 200 vezes, chegando aos 5 bilhões de

dólares. Quando corretores de ações identificaram que era uma ação orquestrada e automatizada, as perdas já eram grandes. **Este tipo de ação mostra outro potencial disruptivo dos perfis automatizados, dessa vez para a economia, podendo gerar impactos que transbordam também para os debates políticos** (Ferrara et al., 2016).

Este tipo de atuação sugere que as redes sociais, usadas por tantas pessoas para fins de informação, podem estar na verdade contribuindo para uma sociedade menos informada, manipulando o debate público e determinando de maneira consistente os rumos do país.

COMO FUNCIONAM?

Robôs são usados nas redes sociais para propagar informações falsas, maliciosas, ou gerar um debate artificial. Para isso, precisam ter o maior número possível de seguidores. Mas como um perfil automatizado pode criar uma rede em torno de si?

Os robôs têm maior facilidade de propagação no Twitter do que no Facebook por uma série de motivos. O padrão de texto do Twitter (140 caracteres) gera uma limitação de comunicação que facilita a imitação da ação humana. Além disso, **o uso de @ para marcar usuários, mesmo que estes não estejam conectados a sua conta na rede, permite que os robôs marquem pessoas reais aleatoriamente para inserir um fator que se assemelhe a interações humanas.**

Robôs também se aproveitam do fato de que, **geralmente, as pessoas são pouco criteriosas ao seguir um perfil no Twitter, e costumam agir de maneira recíproca quando recebem um novo seguidor.** Experimentos mostram que no Facebook, plataforma na qual as pessoas costumam ser um pouco mais cuidadosas ao aceitar novos amigos, 20% dos usuários reais aceitam pedidos de amizade de maneira indiscriminada, e 60% aceitam sempre que possuem ao menos um amigo em comum. Dessa

maneira, os robôs adicionam um grande número de pessoas ao mesmo tempo e seguem páginas reais de pessoas famosas, além de seguir e serem seguidos por um grande número de robôs, de forma que acabam criando comunidades mistas - que incluem perfis reais e falsos (Ferrara et al., 2016).

Alguns robôs pretendem apenas desviar a atenção para um determinado tema e, por isso, se preocupam menos com a sua similaridade com um usuário humano do que com a intensidade e a capacidade de modificar o rumo do debate nas redes. **Outros mecanismos, contudo, possuem uma série de estratégias para imitar o comportamento humano e, assim, serem reconhecidos como tal tanto por usuários e por sistemas de detecção.**

Sabendo que o comportamento humano nas redes sociais tem algum padrão temporal na produção e no consumo de conteúdo, os perfis são programados para postar de acordo com essas mesmas regras. **Paradoxalmente, é justamente a falta de padrão tanto temporal quanto de conteúdo no longo prazo que os robôs têm mais dificuldade de imitar, e o que costuma permitir a sua identificação** (Brito, Salvador e Nogueira, 2013). Os algoritmos mais modernos vão além: conseguem identificar perfis populares e segui-los, identificar um assunto sendo tratado na rede e gerar um pequeno texto por meio de programas de processamento de linguagem natural (*natural language algorithms*) e gerar certo grau de interação.

COMO PODEM SER IDENTIFICADOS?

Não há uma característica única que indique categoricamente se determinado perfil pertence a um usuário real ou falso, automatizado. A identificação é fruto da composição de múltiplas características e indicadores inter-relacionados. As pesquisas nesse campo se distribuem entre três principais linhas de métodos: a) a partir de informações disponíveis nas próprias redes sociais; b) sistemas baseados em contribuição colaborativa (crowdsourcing) e inteligência humana para identificar os perfis de robôs; e c) através de aprendizado de máquinas (machine-learning), baseado na identificação de determinadas características que permitem a automatização da diferenciação entre robôs e pessoas (Ferrara et al, 2016).

Também há diferentes hipóteses que podem ser usadas para embasar a busca por robôs nas redes. Com o método que utiliza conexões entre perfis e dados disponíveis sobre o comportamento na rede social como fórmula de identificação de robôs, alguns sistemas assumem que estes perfis automatizados estarão primordialmente ligados a perfis semelhantes, especialmente no início da sua vida digital. **Isso porque eles precisam construir uma base de seguidores para parecerem críveis aos olhos dos usuários reais.**

Esse método, no entanto, precisa ser ponderado pelo fato de que **usuários humanos são pouco criteriosos com relação a interações e amizades com contas desconhecidas, em especial no Twitter.** Com isso, após certo tempo de existência, as contas de robôs terão redes mistas, não primordialmente compostas por robôs ou por humanos. O volume de robôs dentre os amigos do perfil, contudo, pode ser um indicativo de sua natureza.

O método de *crowdsourcing* parte do princípio que a detecção de robôs seria simples para seres humanos, cuja capacidade de compreensão e identificação do próprio comportamento ainda não foi equiparada por máquinas. Um teste realizado por Wang et al (2013) chegou à conclusão que, em um mecanismo de treinamento curto de identificação (apenas mostrando exemplos de perfis reais e falsos) e respeitando a decisão da maioria dentro de um pequeno grupo de voluntários, o número de falsos-positivos foi bem próximo a zero.

Esse sistema tem algumas dificuldades. Uma é a baixa efetividade de custo para uma rede com muitos usuários, como Twitter e Facebook. Além disso, considerando que avaliadores leigos não têm boa performance individualmente – apenas quando inseridos em sistema de voto da maioria –, **é necessária a participação de algumas pessoas treinadas para garantir o equilíbrio do sistema de voto.**

A detecção através de aprendizado de máquinas ocorre com a codificação de padrões de comportamento a partir da coleta de metadados. Desta forma, **o sistema é capaz de identificar automaticamente humanos e robôs com base no padrão comportamental do perfil.** Estes sistemas

normalmente se organizam a partir de uma base de dados na qual humanos e robôs já foram previamente diferenciados.

Os metadados dos usuários são considerados um dos aspectos mais previsíveis para diferenciar humanos e robôs e podem contribuir para uma melhor compreensão do funcionamento de robôs mais sofisticados. Identificar esses robôs ou contas hackeadas, no entanto, é difícil para estes sistemas. Além disso, a evolução constante dos robôs faz com que o sistema, construído a partir de uma base de dados estática, se torne menos preciso ao longo do tempo. No entanto, ele permite processar um grande número de correlações e padrões complexos, além de analisar um grande número de contas.

Os mecanismos mais eficientes de identificação combinam diferentes aspectos dessas abordagens, explorando múltiplas dimensões do comportamento do perfil, como atividade e padrão de horário (Boshmaf et al., 2012). Estes sistemas levam em conta, por exemplo, que usuários reais passam mais tempo na rede trocando mensagens e visitando o conteúdo de outros usuários, como fotos e vídeos, enquanto contas robôs passam o tempo pesquisando perfis e enviando solicitações de amizade.

Nesse sentido, pesquisas concluem que **as atividades das contas robôs tendem a ser menos complexas na variedade de ações que praticam**, o que adiciona mais uma possibilidade à combinação de fatores que permite que se afirme categoricamente que um determinado perfil é um robô. Esse tipo de sistema, por combinar diferentes dados, também obtém bons resultados a partir de um número menor de informações – como os 100 últimos tweets –, o que acelera a análise e a capacidade de processamento.

Os estudos sobre detecção de robôs em redes sociais se inspiram no esforço para detecção e bloqueio de spams em sistemas de mensagem eletrônica. Nesse sentido, também estão presentes a análise de links compartilhados para identificar fazendas de links (empresas que administram robôs e vendem curtidas, retuítes, etc) e dinâmicas de interações (Ghosh et al, 2012).

Realizando análises dos processos estatísticos que descrevem as interações entre usuários, diversos fatores podem ser estudados e combinados para desenvolver um modelo de detecção de robôs em redes sociais. Alguns exemplos são:

- Variedade de ações enquanto conectados na rede;
- Características do usuário, considerando o número de amigos (as pessoas reais possuem, em média, entre 100 e 1000 seguidores), a proporção e a correlação entre perfis seguidos e perfis que seguem o usuário;
- Características das amizades, analisando como os usuários daquela rede estão interagindo entre si, incluindo padrões relacionados a linguagem, popularidade e horário nos locais de interação;
- Características da rede de retuítes, menções e repetição de hashtags;
- Características temporais, como o tempo médio de produção de tweets;
- Características de conteúdo e linguagem;
- Características do sentimento expressado por meio da postagem.

COMO A DAPP ESTÁ TRABALHANDO PARA IDENTIFICÁ-LOS?

As massas de dados coletadas pela FGV/DAPP são compostas por metadados – informações a respeito dos dados em si – e, por meio deles, exploramos as possibilidades de identificação de contas que tenham atuado automaticamente durante os períodos das análises. Assim, identificamos que o metadado nomeado generator se refere à plataforma que gera o conteúdo do tuíte, o que é muito útil para a detecção de robôs.

Decidimos então verificar todos os generators utilizados em nossas bases de dados e qual a quantidade de tuítes gerada por cada um deles. A partir desses resultados, verificamos em recortes detalhados do que se tratava cada um deles e encontramos plataformas de automatização de produção de conteúdo listadas entre os generators.

Seis casos foram escolhidos para esta primeira análise:

1. O debate da Rede Globo no dia 2 de outubro de 2014 com os presidentiáveis no primeiro turno das eleições;
2. O debate da Rede Globo no dia 24 de outubro de 2014 com os presidentiáveis Dilma Rousseff e Aécio Neves, que disputavam o segundo turno das eleições;
3. As manifestações pró-impeachment realizadas no dia 13 de março de 2016;
4. O debate da Rede Globo com os candidatos a prefeito de São Paulo no dia 29 de setembro de 2016;
5. A greve geral no dia 28 de abril de 2017;
6. A votação da Reforma Trabalhista no Senado, no dia 11 de julho de 2017.

Coletadas as bases de dados referentes aos seis casos, verificamos que 1925 generators diferentes geraram os 7,8 milhões de tuítes publicados em todos eles. Deste total, 181 produziram pelo menos 100 tuítes cada, e foram estes que analisamos manualmente. Essa avaliação permitiu identificarmos 83 generators que produzem tuítes automaticamente de maneira programada ou utilizando a plataforma Twitter através de automação.

ESTUDO DE CASO

ELEIÇÕES 2014

A disputa pela Presidência da República nas eleições de 2014 foi caracterizada por um acirramento político crescente, conseqüente da efervescência das ruas ainda na esteira dos protestos de 2013. Nas redes sociais, a polarização se manifestou de forma agressiva, e parte dessa hostilidade foi provocada por robôs, que chegaram a motivar cerca de 11% das discussões.

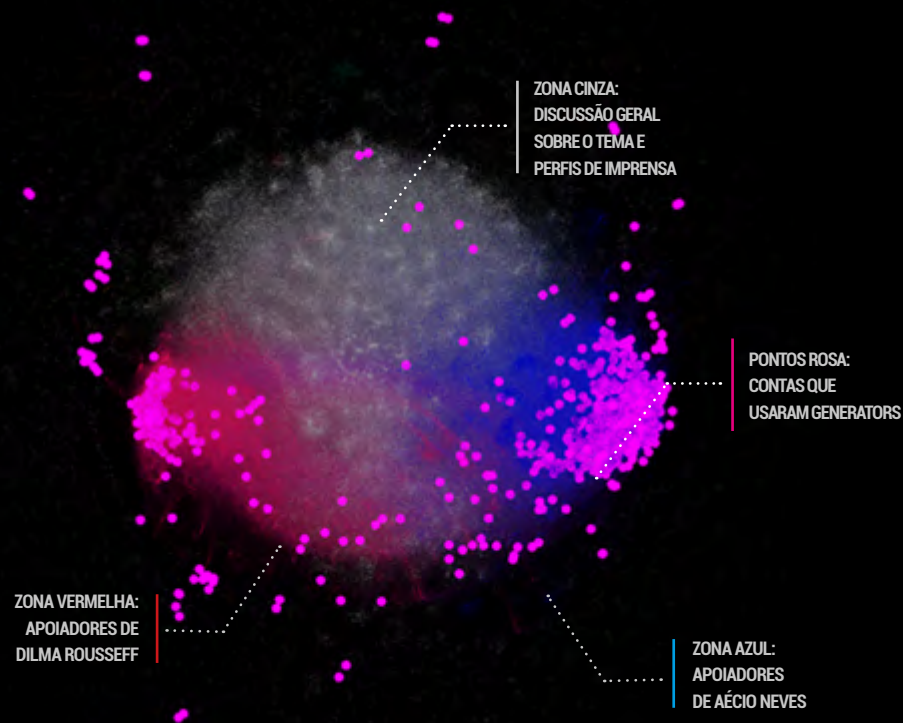
O primeiro turno foi marcado pela morte do candidato Eduardo Campos (PSB), sucedido por Marina Silva (PSB). A disputa culminou em um profundo antagonismo entre Dilma e Aécio (PSDB) no segundo turno, que resultou na vitória da então presidente com pequena margem de vantagem (cerca de três pontos percentuais).

Para analisar a disposição de possíveis robôs dentro das discussões durante as eleições de 2014, selecionamos os tuítes que trataram do debate entre Dilma e Aécio no segundo turno e elaboramos um mapa de interações a partir dos retweets. Três grandes grupos foram identificados: perfis em apoio a Dilma (vermelho), em apoio a Aécio (azul) e que faziam uma discussão geral sobre o tema – o que inclui perfis de imprensa (cinza). Selecionamos então as contas que utilizaram generators suspeitos e as destacamos em tamanho e cor (rosa).

O DEBATE DA REDE GLOBO COM OS CANDIDATOS À PRESIDÊNCIA DA REPÚBLICA

2º TURNO
24 DE OUTUBRO DE 2014

2307185 Tweets
24/10/2014 00:00 - 25/10/2014 12:00



Observando o grafo acima, notamos que as contas que produziram tweets por meio de generators suspeitos estão concentradas nos extremos dos polos de apoio aos candidatos. Praticamente nenhuma conta suspeita está no grupo que em geral não defende nenhum candidato (zona cinza).

fonte: FGV/DAPP

PONTOS VERDES:
 CONTAS QUE GERARAM
 PELO MENOS DUAS VEZES
 DOIS TUÍTES CONSECUTIVOS
 EM MENOS DE UM SEGUNDO.

De todas as interações no Twitter nas horas analisadas, 11,34% foram motivadas por tuítes ou retuítes de robôs. Entre os apoiadores de Aécio Neves (cluster azul), porém, essa parcela de interações com contas automatizadas (robôs sendo retuitados por outros robôs ou contas comuns) chegou a 19,41%. Nas discussões entre perfis em apoio a Dilma, foram 9,76%.

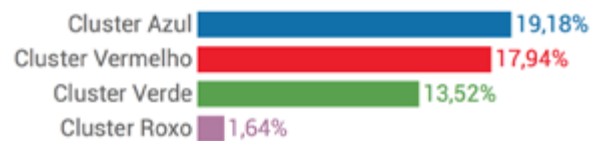


Explorando as atividades das contas suspeitas, encontramos perfis que são claramente automatizados para inflar o apoio a um candidato. Entre essas contas, identificamos as que postaram mais de uma vez por segundo, atividade suspeita de automação. E destacamos em verde, no mesmo mapa, as que produziram pelo menos duas vezes dois tuítes consecutivos com intervalo inferior a um segundo. Novamente, notamos que elas se encontram nos pólos extremos de defesa dos candidatos.

fonte: FGV/DAPP

O mesmo se repete para todos os casos analisados. Nas seguintes visualizações, também colorimos as contas que usaram generators suspeitos de rosa, as que fizeram tweets consecutivos com intervalo menor a um segundo pelo menos duas vezes de verde, e as contas que correspondem a ambos critérios de branco.

No debate do 1º turno analisado, as interações com robôs representaram apenas 6,29% da discussão no Twitter. Mais uma vez, foi entre os perfis de apoiadores de Aécio Neves que essas interações foram mais expressivas, representando 19,18% do debate no cluster azul. Entre os apoiadores de Dilma, foram 17,94%.



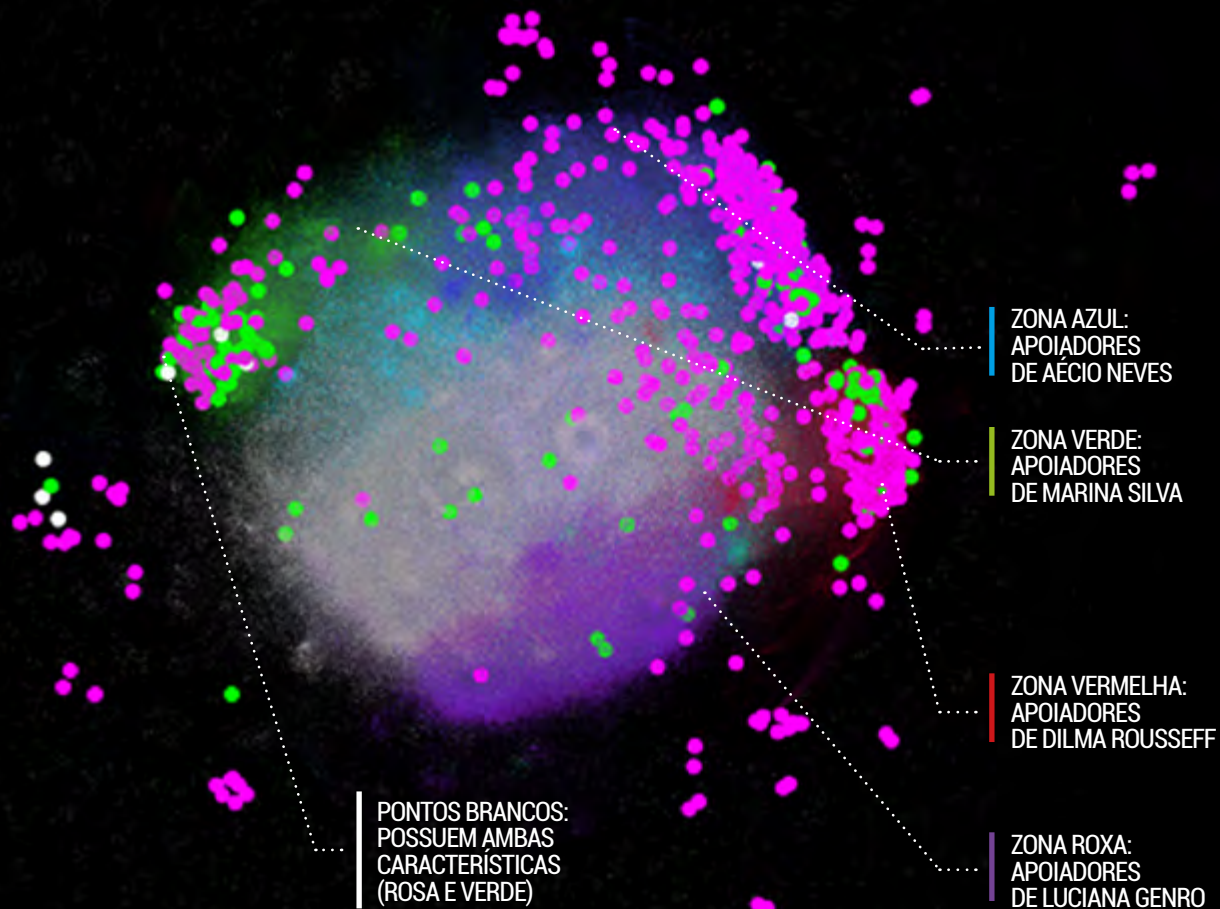
DEBATE DA REDE GLOBO PARA AS ELEIÇÕES PRESIDENCIAIS

1º TURNO

2 DE OUTUBRO DE 2014

1565773 Tweets

02/10/2014 00:00 - 03/10/2014 12:00



fonte: FGV/DAPP

ESTUDO DE CASO

IMPEACHMENT 2015

A vitória de Dilma Rousseff não freou a crescente hostilidade entre os campos políticos. A dificuldade da presidente de manter apoio político no Congresso e a crise econômica do país em recessão resultou em um processo de impeachment com apoio popular manifesto em uma série de protestos pelo país. O grafo ao lado mostra como ocorreram as discussões no Twitter no dia da maior manifestação registrada a favor do impeachment.

Pelo menos 10% das interações sobre o assunto neste dia foram impulsionadas por robôs, ou seja, retuítes de conteúdo originado por conta automatizada. No cluster de apoiadores de Dilma Rousseff, essa proporção chegou a 21,43%, o que mostra o poder de influência deste tipo de conta no debate público.

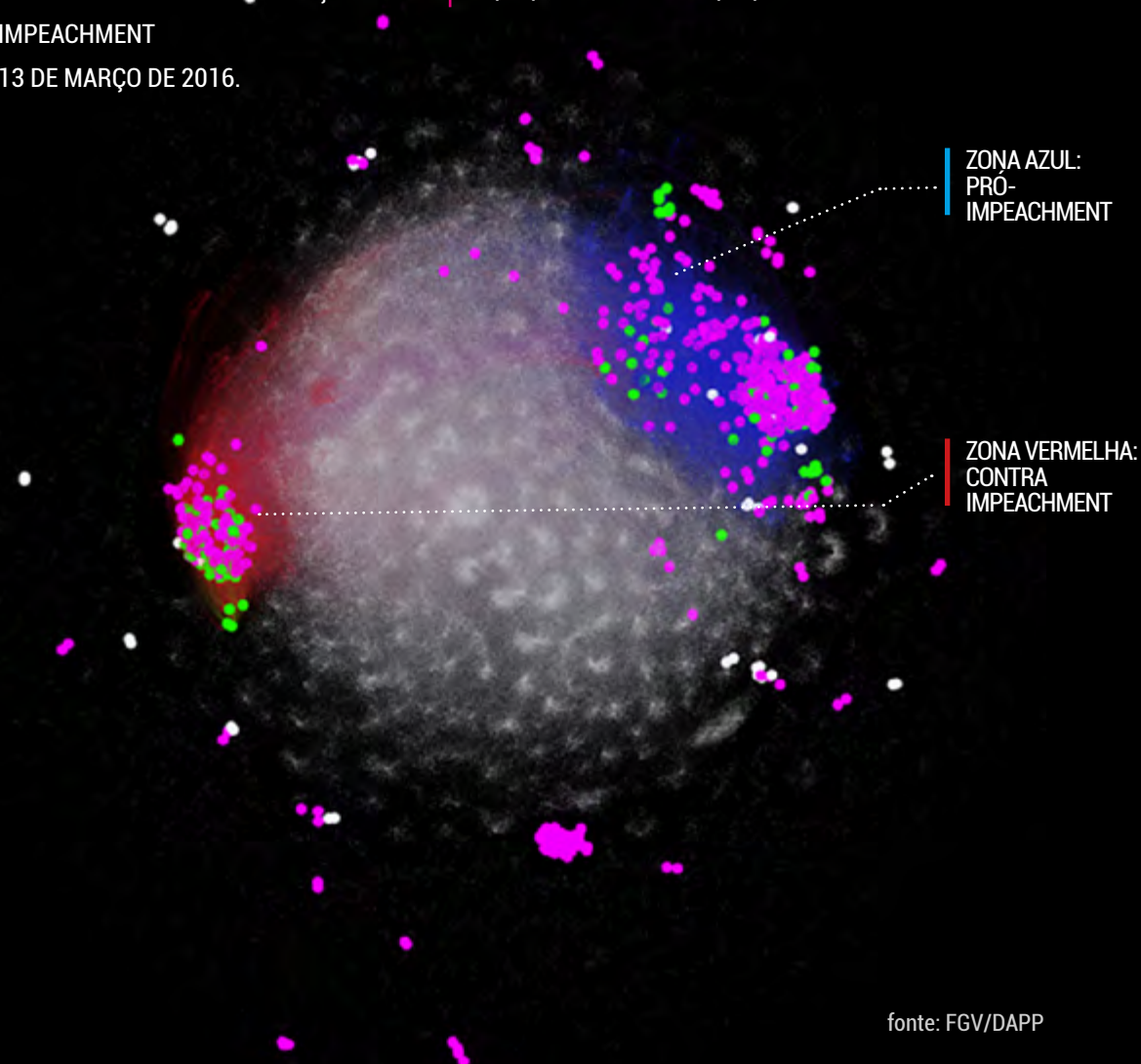


MANIFESTAÇÕES PRÓ-IMPEACHMENT

CONCENTRAÇÃO DE CONTAS SUSPEITAS
NOS POLOS DE APOIO E OPOSIÇÃO AO
IMPEACHMENT
13 DE MARÇO DE 2016.

826493 tweets

13/03/2016 00:00 – 13/03/2016 23:59



fonte: FGV/DAPP

ESTUDO DE CASO

ELEIÇÕES MUNICIPAIS DE SÃO PAULO EM 2016

A disputa pela prefeitura de São Paulo começou com uma dispersão nas intenções de voto entre o então prefeito que concorria à reeleição, Fernando Haddad (PT), João Doria (PSDB), Celso Russomanno (PRB), Marta Suplicy (PMDB) e Luiza Erundina (PSOL). Essa divisão durante todo o primeiro turno fez com que os debates eleitorais fossem marcados por uma discussão não tão antagonica entre PT e PSDB como nas eleições presidenciais. A eleição foi definida ainda no primeiro turno, com a vitória de João Doria.

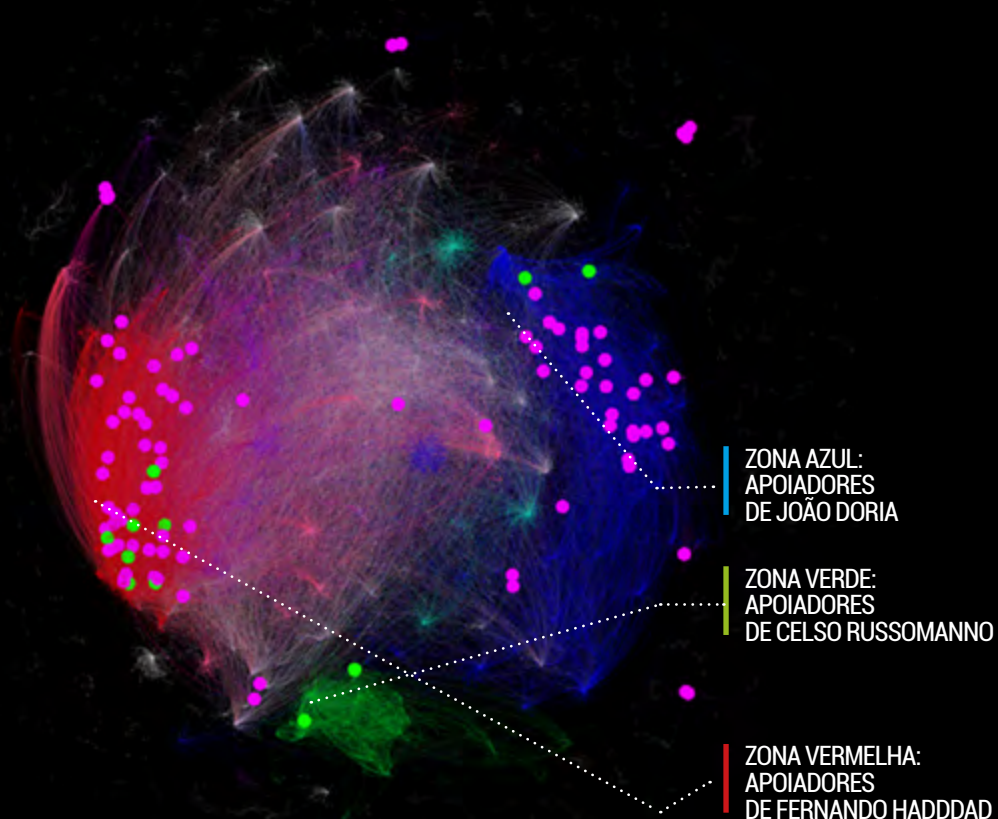
O grafo ao lado mostra como outras forças políticas tiveram influência nos debates nas redes sociais. As interações motivadas por robôs também foram mais equânimes. Entre os apoiadores de Doria, elas representaram 11,25% do debate; entre os apoiadores de Haddad, 11,54%; entre os de Russomanno, 8,40.



DEBATE DA REDE GLOBO COM OS CANDIDATOS A PREFEITO DE SÃO PAULO

1º TURNO
24/09/2016

148257 tweets
29/09/2016 00:00 – 30/09/2016 12:00



fonte: FGV/DAPP

ESTUDO DE CASO

GREVE GERAL DE 28 DE ABRIL DE 2017

Após o impeachment de Dilma Rousseff, ganhou força no Congresso Nacional o debate sobre reformas da legislação trabalhista e previdenciária. O principal argumento favorável foi que a necessidade de austeridade para superar a crise deveria ser vista como uma oportunidade para modernizar estas legislações, enquanto o principal argumento contrário identificou neste movimento perda de direitos e precarização das condições de trabalho e da rede de proteção social do estado brasileiro.

Foi a partir deste cenário que centrais sindicais e partidos contrários às reformas convocaram uma greve geral o dia 28 de abril de 2017, contando que uma alta adesão convenceria o espectro político da insatisfação popular com relação a estas pautas. Como é possível observar no grafo ao lado, os robôs tiveram mais uma vez grande presença. Entre os apoiadores da greve, 22,39% das interações foram motivadas por tuítes automatizados.



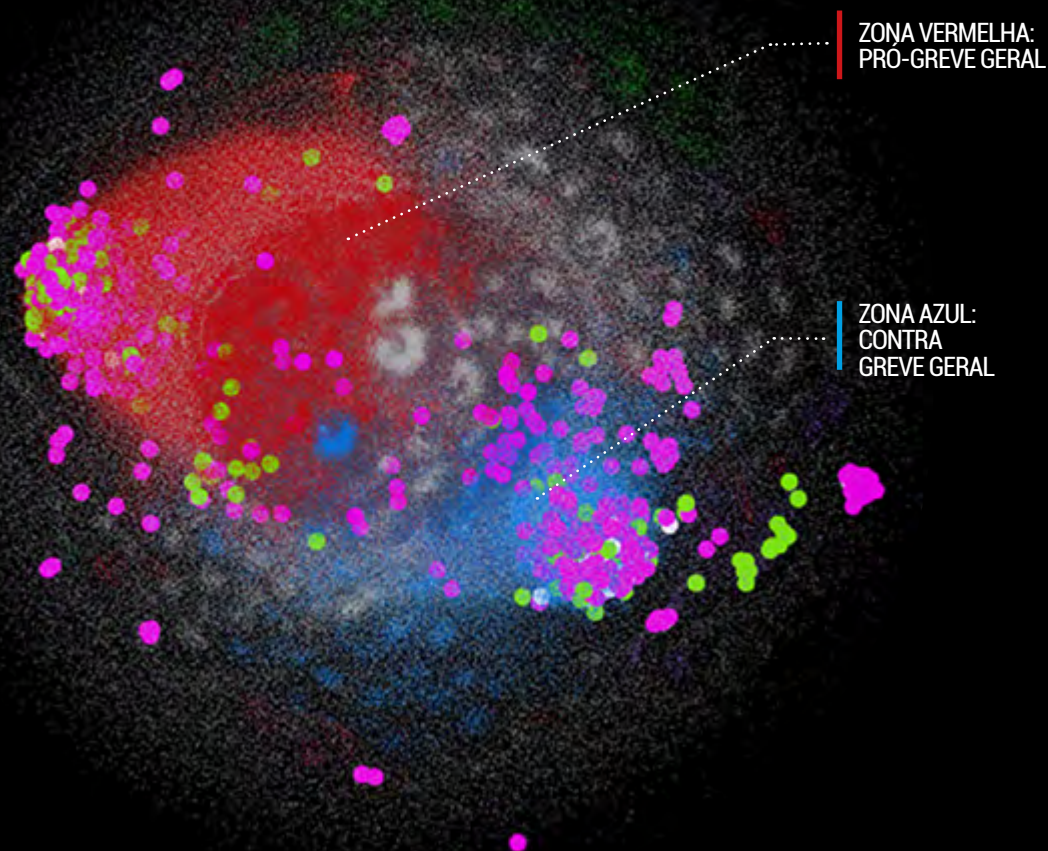
GREVE GERAL 28 DE ABRIL DE 2017

CONTAS SUSPEITAS ESPALHADAS PELO CLUSTER CONTRÁRIO À GREVE E CONCENTRADAS NO CLUSTER FAVORÁVEL

1460160 tweets

13/03/2016 00:00 – 13/03/2016 23:59

28 DE ABRIL DE 2017



fonte: FGV/DAPP

ESTUDO DE CASO

VOTAÇÃO DA REFORMA TRABALHISTA NO SENADO DE 11 DE JULHO DE 2017

Um dos principais focos da agenda de recuperação econômica do atual governo foi a aprovação de uma reforma da legislação trabalhista. **Os debates sobre esta proposta nas redes sociais seguiram a tendência de polarização já observada em outros momentos da política nacional.** Após meses de discussões, negociações e modificações, a Reforma Trabalhista foi colocada no plenário do Senado Federal para votação no dia 11 de julho.

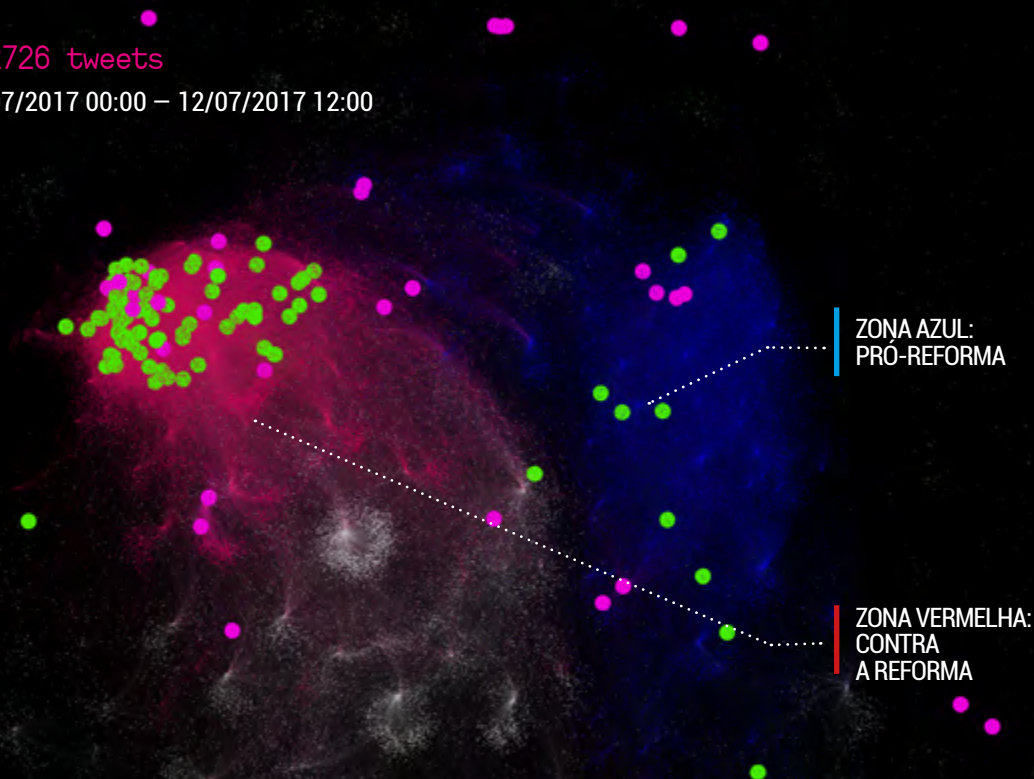
Os robôs estão mais uma vez presentes nos dois extremos do debate, estando nessa análise em maior número no polo contrário à reforma. No total, identificamos 2% das interações relacionadas a este evento como automatizadas - 3% das interações contrárias à reforma e 1% das favoráveis.



VOTAÇÃO DA REFORMA TRABALHISTA

252726 tweets

11/07/2017 00:00 – 12/07/2017 12:00



fonte: FGV/DAPP

VERIFICAÇÃO DA ANÁLISE

Para validação da análise, verificamos manualmente uma amostra de 2153 contas suspeitas nos seis casos escolhidos. Esse procedimento manual garante um nível de confiança de 95% com intervalo de confiança de dois pontos. **A verificação procurou checar se a conta produz conteúdo de forma totalmente automatizada.**

Observamos que mais de 50% destas contas possuem um aspecto de total automação. Aproximadamente 9% das contas são institucionais (de órgãos de imprensa e blogs, por exemplo), em torno de 6% já deixaram de existir, quase 2% foi suspensa pela plataforma, e pouco mais de 25% são apenas parcialmente automatizadas.

Como exemplo deste último caso, temos contas que participaram de programas de apoio a uma determinada causa e autorizaram que, em determinados momentos, tuítes automáticos fossem publicados por seus perfis na plataforma. Outro exemplo são contas que têm produção original de conteúdo mas criam gatilhos, como a publicação de notícias envolvendo certa figura pública ou instituição e as postam de maneira automática.

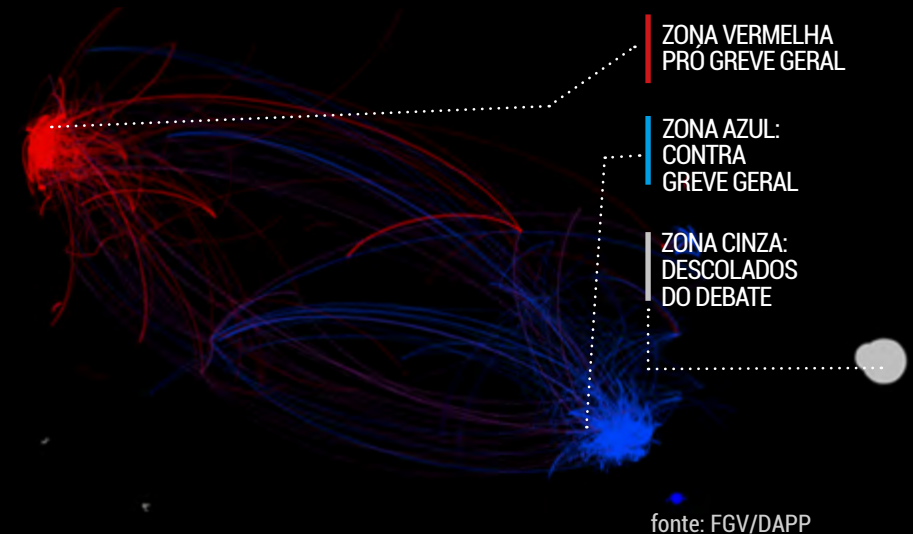
Concluimos, portanto, que a verificação de contas pode ser contraditória e, para este momento, consideramos que cada tweet deve ser avaliado quanto a sua origem (se automatizada ou não).

Voltemos agora para a exploração da topografia da rede de retuítes. No caso da greve geral, notamos que os tuítes produzidos de forma automática serviram basicamente aos extremos de discussão, como mostra o grafo a seguir, onde filtramos apenas as

GREVE GERAL 28 DE ABRIL DE 2017

1460160 tweets

28/04/2017 00:00 – 28/04/2017 23:59



Os núcleos de discussão pró e contra a greve participaram de interações com tuítes produzidos automaticamente. O grupo cinza é constituído por conteúdos automatizados que usam assuntos que estão no debate do momento para conseguir novos seguidores, sem aparente tentativa de influenciar o debatedo momento para conseguir novos seguidores, sem aparente tentativa de influenciar o debate.

CONCLUSÃO

[...]é imprescindível diferenciar robôs maliciosos de robôs com outras finalidades, como marketing digital, promoções de marcas, blogs e empresas, perfis institucionais, perfis com diversos administradores.

O surgimento de contas automatizadas permitiu que estratégias de manipulação, disseminação de boatos e difamação, comumente usadas em disputas políticas, ganhassem uma dimensão ainda maior nas redes sociais. **A participação ostensiva de robôs no ambiente virtual tornou urgente a necessidade de identificar suas atividades e, conseqüentemente, diferenciar quais debates são legítimos e quais são forjados.** Esse discernimento é essencial para que os processos sociais originados nas redes sejam efetivamente compreendidos.

Para que os robôs sejam identificados, a DAPP tem desenvolvido uma metodologia que **combina avaliações de metadados distintos para abarcar todas as possíveis estratégias de**

criação e operação de contas automatizadas. Com a dinâmica de aprimoramento da tecnologia cada vez mais veloz, os robôs têm sua atividade aprimorada a cada dia, aproximando-se do comportamento humano. Além disso, é imprescindível diferenciar robôs maliciosos de robôs com outras finalidades, como marketing digital, promoções de marcas, blogs e empresas, perfis institucionais, perfis com diversos administradores.

Outro desafio que se apresenta é a **identificação de contas ciborgue, aquelas que são parcialmente automatizadas e também manipuladas por humanos,** que postam conteúdo real para inserir o aspecto de aleatoriedade e imprevisibilidade comum

nas interações humanas. A utilização dessas contas dificulta a detecção da operação de robôs quando tentamos classificar uma conta no Twitter com a variável binária de robô ou humano, por exemplo.

A análise de interações de contas com tuítes produzidos automaticamente já denuncia e confirma o uso de robôs no debate político brasileiro. A partir da análise feita pela DAPP de metadados que denunciam a operação deles, podemos concluir que o conteúdo gerado automaticamente tem influenciado discussões no Twitter com objetivo de gerar vantagem para atores políticos.

O estudo identificou que a operação de produção de conteúdo automatizado não tem ocorrido exclusivamente apenas em um polo ou campo político. A análise dos seis casos escolhidos sugere que grupos com diferentes interesses, especialmente os localizados nos extremos do espectro político, se inflam e se atacam mutuamente com esta prática.

A análise dos seis casos escolhidos sugere que grupos com diferentes interesses, especialmente os localizados nos extremos do espectro político, se inflam e se atacam mutuamente com esta prática.

Ampliar a capacidade de identificar e reprimir a automatização maliciosa de perfis nas redes sociais deve ser uma prioridade. Análises recentes mostram que esse tipo de ação tem sido bem sucedida em direcionar o debate público, cada vez mais presente nas redes, influenciando diretamente em momentos de inflexão e determinantes para o futuro.

Dessa maneira, para que as redes sociais continuem sendo um espaço democrático de opinião e informação, é necessário identificar a organicidade dos debates. Para que as redes se tornem mais transparentes é também fundamental que os responsáveis por esse tipo de ação coordenada comecem a ser identificados, buscando compreender os interesses por trás da contratação destes serviços de automatização e propagação de desinformação.

BIBLIOGRAFIA

Many Americans Believe Fake News Is Sowing Confusion.

Disponível em:

<http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/>

Bessi, A.; Ferrara, E. **Social bots distort the 2016 US Presidential election online discussion.** Disponível em: <http://firstmonday.org/ojs/index.php/fm/article/view/7090/5653>

Pew Research Center. **News use across social media platforms 2016.** Disponível em: <http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/> Acesso em 02 de agosto de 2017.

Secretaria Especial de Comunicação Social da Presidência da República (Secom/PR). **Pesquisa Brasileira de Mídia 2016 - Hábitos de consumo de mídia pela população brasileira.** Disponível em: <http://www.secom.gov.br/atuacao/pesquisa/lista-de-pesquisas-quantitativas-e-qualitativas-de-contratos-atuais/pesquisa-brasileira-de-midia-pbm-2016-1.pdf/view>. Acessado em 16 de agosto de 2017.

Wang, A. **Detecting Spam Bots in Online Social Networking Sites: a machine learning approach.** In: Foresti, S.; Jajodia, S. Data and applications security and privacy XXIV. Springer, pp. 335-342, 2010.

Brito, F.; Salvador, I.; Rocha, E. **Detecting social-network bots based on multiscale behavior analysis.** In: Seventh International Conference on emerging security information, systems and technologies. 2013.

Lee, K.; Eoff, B.; Caverlee, J. **Seven Months with the devils: a long-term study of content polluters on twitter.** In: Fifth International Conference on Weblogs and Social Media of the Association for the Advancement of Artificial Intelligence, 2011.

Ferrara, E.; Varol, O.; Davis, C.; Menczer, F.; Flammini, A. **The rise of social bots.** In: Communications of the ACM, v.59, n.2, 2016.

Davis, C.; Varol, O.; Ferrara, E.; Flammini, A.; Menczer, F. **BotOrNot: a system to evaluate social bots.** 2016.

Tavares, G.; Faisal, A. **Scaling-laws of human broadcast communication enable distinction between human, corporate and robot twitter users.** PLoS ONE v.8, n.7, 2013.

Varol, O.; Ferrara, E.; Davis, C.; Menczer, F.; Flammini, A. **Online Human-bot interactions: detection, estimation, and characterization.** In: Eleventh International Conference on Weblogs and Social Media of the Association for the Advancement of Artificial Intelligence, 2017.

Arnaudo, D. **Computational propaganda in Brazil: social bots during elections.** University of Oxford Working Paper, n.8, 2017.

Chavoshi, N.; Hamooni, H.; and Mueen, A. **Identifying correlated bots in twitter.** In Social Informatics: 8th Intl. Conf., 14–21, 2016.

Chu, Z.; Gianvecchio, S.; Wang, H.; and Jajodia, S. **Detecting**

automation of twitter accounts: Are you a human, bot, or cyborg? IEEE Tran Dependable & Secure Comput 9(6):811–824, 2012.

Clark, E.; Williams, J.; Jones, C.; Galbraith, R.; Danforth, C.; and Dodds, P. **Sifting robotic from organic text: a natural language approach for detecting automation on twitter**. Journal of Computational Science 16:1–7, 2016.

Ratkiewicz, J.; Conover, M.; Meiss, M.; Goncalves, B.; Flammini, A.; and Menczer, F. **Detecting and tracking political abuse in social media**. In 5th Int Conf on Weblogs & Soc Med, 297–304, 2011.

Wald, R.; Khoshgoftaar, T. M.; Napolitano, A.; and Sumner, C. **Predicting susceptibility to social bots on twitter**. In Proc. 14th Intl. IEEE Conf. on Information Reuse and Integration, 6–13, 2013.

Matsubara, Y., Sakurai, Y., Ueda, N., Yoshikawa, M.: **Fast and exact monitoring of co-evolving data streams**. In: 2014 IEEE International Conference on Data Mining, pp. 390–399. IEEE, 2014

Ghosh, S., Viswanath, B., Kooti, F., Sharma, N.K., Korlam, G., Benevenuto, F., Ganguly, N., Gummadi, K.P.: **Understanding and combating link farming in the twitter social network**. In: Proceedings of the 21st International Conference on World Wide Web - WWW 2012, p. 61. ACM Press, New York, 2012

Tavares, G.; Faisal, A. **Scaling-laws of human broadcast communication enable distinction between human, corporate**

and robot twitter users. PLoS ONE v.8, n.7, 2013.

Boshmaf, Y.; Muslukhov, I.; Beznosov, K.; Ripeanu, M. Design and analysis of a social botnet. 2012.

Wired. **How Twitter bots played a role in electing Donald Trump**. Disponível em: <http://www.wired.co.uk/article/twitter-bots-democracy-usa-election> Acesso em 11 de agosto de 2017.

The New York Times. **On Twitter, a Battle Among political Bots**. Disponível em: <https://www.nytimes.com/2016/12/14/arts/on-twitter-a-battle-among-political-bots.html> Acesso em 11 de agosto de 2017

Deutsche welle. **The rise of political bots on social media**. Disponível em: <http://www.dw.com/en/the-rise-of-political-bots-on-social-media/a-19450562> Acesso em 12 de agosto de 2017.

NY Daily Newas. **The billionaire GOP patron behind Trump's social media bot army**. Disponível em: <http://www.nydailynews.com/news/politics/billionaire-gop-patron-behind-trump-social-media-bot-army-article-1.3236933> Acesso em 12 de agosto de 2017.

MIT Technology Review. **How the Bot-y Politic Influenced This Election**. Disponível em: <https://www.technologyreview.com/s/602817/how-the-bot-y-politic-influenced-this-election/>. Acesso em 12 de agosto de 2017.

Le Monde Diplomatique. **Entre trolls, robôs e ativadores: as eleições na internet**. Disponível em: <http://diplomatie.org.br/entre-trolls-robos-e-ativadores-as-eleicoes-na-internet/>. Acesso em 12 de agosto de 2017.



Instituição de caráter técnico-científico, educativo e filantrópico, criada em 20 de dezembro de 1944 como pessoa jurídica de direito privado, tem por finalidade atuar, de forma ampla, em todas as matérias de caráter científico, com ênfase no campo das ciências sociais: administração, direito e economia, contribuindo para o desenvolvimento econômico-social do país.

Escritório

Praia de Botafogo 190, Rio de Janeiro
RJ - CEP 222509000
Caixa Postal 62.591 CEP 22257-970
Tel (21) 3799-5498 | www.fgv.br

Presidente Fundador

Luiz Simões Lopes

Presidente

Carlos Ivan Simonsen Leal

Vice-Presidentes

Sergio Franklin Quintella,
Francisco Oswaldo Neves Dornelles e
Marcos Cintra Cavalcante de Albuquerque



EXPEDIENTE

EQUIPE DE EXECUÇÃO

Diretor

Marco Aurélio Ruediger

Pesquisadores

Amaro Grassi
Ana Freitas
Andressa Contarato
Carolina Taboada
Danilo Carvalho
Humberto Ferreira
Lucas Roberto da Silva
Pedro Lenhard
Rachel Bastos
Thomas Traumann

Projeto gráfico

Arielle Asensi
Luis Gomes
Rebeca Liberatori Braga