

“E o Google?” Técnicas de coleta dos resultados de pesquisa do Google¹²

“And Google?” Techniques for collecting Google search results

Max Stabile, Alexandre Arns³, Alexandre Gomes, Teógenes Moura

Resumo: Este trabalho faz parte de uma ampla agenda de pesquisa que debate os impactos das novas tecnologias em processos eleitorais e, mais especificamente, a relação entre a oferta de informações por serviços de busca e a construção de preferências eleitorais. Esta proposta tem como enfoque a apresentação de uma metodologia de pesquisa para coleta e análise dos resultados de motores de busca na Web. O caso selecionado para desenvolver a metodologia foi o motor de busca da Google. Apesar da existência de uma literatura incipiente que discute o impacto potencial dos resultados da pesquisa nas preferências políticas ainda sabemos pouco sobre esse impacto e como analisá-lo. Os resultados preliminares desta pesquisa exploratória apontam para as diversas dificuldades no procedimento de coleta, sistematicamente barrados pelo Google. Ao mesmo tempo, argumentamos que há possibilidade de monitoramento desses resultados e de análise dos seus impactos. Terminamos o artigo discutindo os limites do desenho de pesquisa utilizado e outros desenhos possíveis. Também debatemos a possibilidade da utilização de métodos mistos, a partir da triangulação de dados, como os coletados nesta pesquisa, com dados obtidos por meio de outras técnicas.

Palavras-chave: Motor de busca; Google; eleições; metodologia

Abstract: This work is part of a broad research agenda that discusses the impacts of new technologies on electoral processes and, more specifically, the relationship between the supply of information by search engines and the construction of electoral preferences. This proposal focuses on the presentation of a research methodology for collecting and analyzing the results of search engines on the Web. The case selected to develop the methodology was Google's search engine. The preliminary results of this exploratory research point to the various difficulties in the collection procedure, systematically blocked by Google. At the same time, we argue that it is possible to monitor these results and analyze their impacts. We finished the article by discussing the limits of the research design used and other possible designs. We also discuss the possibility of using mixed

¹ Trabalho apresentado ao GT 4 - Internet e Política do VIII Congresso da Associação Brasileira de Pesquisadores em Comunicação e Política (VIII COMPOLÍTICA), realizado na Universidade de Brasília (UnB), de 15 a 17 de maio de 2019.

² Agradecemos à professora do Programa de Ciência da Computação, Claudia Melo, e sua turma de alunos da disciplina Engenharia de Software pelo auxílio à viabilização deste trabalho, sobretudo o grupo de alunos responsáveis pelo desenvolvimento do código de coleta de informações das páginas de resultados de busca da Google: Douglas Alves Ferreira, Estéfane Helen, Gabriel Taumaturgo, Gabriel Almeida, Guilherme Castro, Juana, Léo Moraes da Silva, Lincoln Abreu Barbosa, Luis Braga, Luiz Filipe, Marcus Vinicius, Mikael Mello, Ricardo Rachaus, Tomas Rosário Rosemberg e Yan Trindade. O trabalho desenvolvidos por esses alunos pode ser acessado no repositório do GitHub <<https://github.com/unb-cic-esw/Observatorio-google>>.

³ Bolsista Capes pelo Instituto da Democracia e Democratização da Comunicação. A atuação na produção deste trabalho só foi possível em virtude do financiamento da Capes.

methods, based on triangulation of data, as collected in this research, with data obtained through other techniques.

Key-words: Search engines; Google; elections; methodology;

1. INTRODUÇÃO

Este trabalho é parte de uma ampla agenda de pesquisa que debate os impactos dos provedores de aplicações na internet sobre processos eleitorais e, mais especificamente, da oferta de informações por serviços de busca de conteúdos na Web e sua relação com a construção de preferências eleitorais. O propósito deste artigo é a apresentação de uma metodologia de pesquisa para coleta e análise dos resultados de motores de busca na Web. O caso selecionado para desenvolver a metodologia foi o motor de busca da Google (www.google.com.br / www.google.com), por ser o motor de busca de maior acesso no Brasil⁴. Convém destacar que o método desenvolvido para simulação das buscas de resultados no Google tomou o cuidado de não entrar em conflito com os Termos de Uso da plataforma. Considerando que a criação de contas e perfis de usuários, o treinamento destes perfis e simulação das coletas de busca foi realizada durante o processo eleitoral brasileiro, a execução da simulação tomou cuidado em não interagir com nenhum usuário não interagiram com a conta de qualquer usuário do Google.

Apesar da existência de uma literatura incipiente sobre o impacto potencial das aplicações das mídias sociais na formação das preferências dos cidadãos (EPSTEIN, 2016, s/p), a quantidade de estudos que se voltam para as aplicações de busca de conteúdo na internet é ainda menor (PAN et. al., 2007; EPSTEIN, ROBERTSON, 2015; EPSTEIN, 2015, ROBERTSON, 2016, 2018). Sabemos pouco sobre como a indexação dos resultados e das informações expostas aos usuários dos motores de busca pode influenciar sua compreensão da realidade.

Embora incipiente, a produção tem provocado reações públicas de chefes executivos das grandes corporações. Como no caso da Google, em que seu vice-presidente e chefe do

⁴ Segundo o ordenador de sítios eletrônicos de maior tráfego de dados da Alexa, a Google e o Yahoo são as únicas empresas de motor de busca dentre os 50 sítios mais acessados no Brasil. Os endereços <google.com> e <google.com.br> estão na primeira e terceira posição deste ordenamento, respectivamente. O <yahoo.com.br> está em décima primeira posição. Ver em <<https://www.alexa.com/topsites/countries/BR>> Acesso 02 abr. 2019.

grupo de pesquisa, Amitabh Kumar “Amit” Singhal⁵, negou a acusação realizada por Robert Epstein (2015) de que a empresa estaria conspirando para influenciar as eleições presidenciais estadunidenses de 2016 (SINGHAL, 2015). Parte das controvérsias envolvendo os potenciais impactos dos motores de busca sobre a formação de preferências decorre da baixa transparência da plataforma em termos dos algoritmos que ordenam as exibições de busca e das poucas técnicas disponíveis para a coleta dos dados a respeito dos resultados dessas buscas.

Discussões sobre o impacto de algoritmos no processo de tomada de decisões cotidianas, na sociedade moderna, tem acompanhado o desenvolvimento tecnológico de microprocessadores e a produção de dispositivos computacionais móveis, como os denominados ‘*smartphone*’. A acessibilidade a estes dispositivos viabiliza, também, o uso de um conjunto de serviços que consistem em aplicações algorítmicas. Tanto os dispositivos quando as aplicações utilizadas expressam, em sua forma e no seu código, princípios que organizam a perspectiva de mundo das pessoas que trabalharam na construção e podem produzir vieses em seus serviços.

Os motores de busca são uma destas aplicações algorítmicas, eles são programas que buscam catalogar e organizar as informações disponíveis na internet para que sejam de fácil acesso para os usuários. A operacionalidade deste serviço depende do trabalho de um contingente de engenheiros. São conjuntos de pessoas programando os diversos algoritmos que fazem com que o serviço de busca seja possível. Conseqüentemente, há uma multitude de vieses sendo empregados, desde os mais inofensivos até aqueles que de fato podem afetar negativamente a população. Nesse sentido, indaga-se: quais os vieses que um motor de busca contém e que forma eles incidem sobre a formação de preferências na sociedade moderna?

Buscando contribuir para a reflexão desta indagação, elaboramos uma proposta de método para analisar os resultados de busca apresentados pelo Google. A seção seguinte, denominada de *O método*, explica os instrumentos empregados para produzir uma simulação uso do motor de busca da Google. A seção explica como foram construídas as contas de usuários fictícios da Google; e explica a implementação e a execução da simulação, durante o

⁵ Singhal trabalhou na Google até 26 de fevereiro de 2016.

período eleitoral brasileiro em 2018, e a coleta de informações dos resultados das buscas realizadas.

2. O MÉTODO

O objetivo da metodologia desenvolvida é simples: coletar de forma sistemática os principais resultados de busca exibidos pelo Google para determinada palavra-chave. A tarefa para alcançar esse objetivo foi realizada em duas etapas. A primeira etapa tornou-se necessária ao nos indagarmos se ao utilizarmos uma conta “logada” do Google os resultados seriam diferentes ou não. Essa pergunta necessitou um trabalho anterior ao processo de coleta: precisaríamos então “treinar” diferentes contas para que simulassem diferentes “usuários reais”. Essa primeira etapa chamaremos de “Treinamento”. A segunda etapa foi o processo de coleta dos resultados de busca através da técnica de “raspagem das páginas”.

2.1 Treinamento e Simulação de contas específicas para a captura

Para construção das contas desta simulação, optamos pela confecção de quatro tipos distintos de históricos de navegação: “esquerda”, “direita”, “mulher”, “homem” e “neutro”. A confecção destas contas teve como base a automatização dos seus acessos a determinados sítios eletrônicos e a busca de termos específicos, com intuito de registrar um “histórico” de navegação junto ao motor da Google. Este processo de confecção do histórico das contas é definido como “treinamento” e que foi realizada desde o dia 18 de junho de 2018 até o término das eleições brasileiras, em 28 de outubro do mesmo ano. Tanto os sítios quanto os termos de buscas utilizados para o “treinamento” podem ser visualizadas no Anexo I. O treinamento específico para cada um dos tipos de contas decorre da necessidade de controle, sobre cada conta, para verificar as variações, conforme o histórico, nos resultados de busca de destas contas.

As quatro contas se inserem em duas categorias, com finalidades distintas. A conta de “esquerda” e a de “direita” integram a categoria **ideológica**, cujo papel é simular os diferentes posicionamentos ideológicos dos eleitores no Brasil, embora simplifique-os em apenas dois campos. A conta “mulher” e a conta “homem” integram a categoria **“gênero”**, cuja papel foi verificar a existência de viés com relação a caracterização de gênero. Por fim, a conta “neutro” teve a finalidade de servir de controle desta simulação. A conta “neutro”, convém

destacar, foi a única conta que não passou pelo processo de treinamento, ou seja, de formação de histórico de navegação.

Para construir a categoria **ideológica** nós nos baseamos em um mapeamento de páginas públicas no Facebook realizado pelo Instituto Brasileiro de Análise de Dados (IBPAD)⁶ (Anexo I), categorizando-as em uma rede de dois polos, entre as páginas de “esquerda” e as páginas de “direita”. A partir deste mapeamento **(i)** ordenamos, de modo decrescente, as páginas de ambos os polos pela centralidade de grau de entrada, que foi mensurado pelo número de “curtidas” que estas páginas receberam e o número de atividades que realizaram. Ainda no critério de ordenamento destas páginas, **(ii)** convém pontuar que como construímos uma simulação que simplifica o espectro ideológico em dois polos, desconsideramos as páginas referentes a pré-candidaturas e partidos políticos pela dificuldade de acomodar as agremiações e as lideranças nessa simplificação; e, sobretudo, para não interferir na identificação do viés na segunda etapa da nossa metodologia que foi baseada na busca dos nomes dos candidatos(as) a presidência durante as eleições. Por isso, consideramos as contas seriam de usuários fictícios sem preferências partidária.

Por fim, do mapeamento de cada polo, “esquerda” e “direita”, **(iii)** realizou-se o levantamento de quais páginas no Facebook possuíam, também, um sítio próprio na *Web* de modo a priorizar, no momento de “treinamento”, o acesso da conta aos domínios do sítio ao invés do domínio do Facebook. Para construir a categoria “**gênero**” utilizamos, como ponto de partida uma ferramenta do Facebook, o *Facebook Audience Insights*, que disponibiliza quais os principais páginas e interesses do público masculino e feminino dos usuários da rede em território brasileiro. Além desses termos e páginas de interesse também foram adicionadas buscas sobre doenças especificamente de cada um dos sexos.

2.1.1. Implementação das contas dos usuários

Constituídas as contas, iniciamos os respectivos treinamentos no dia 18 de junho de 2018, para os perfis “esquerda” (resocie.esquerda@gmail.com) e “direita” (resocie.direita@gmail.com); e no dia 19 de junho de 2018, para os perfis “mulher” (resocie.mulher@gmail.com), “homem” (resocie.homem@gmail.com) e “neutro”

⁶ Ver em: <<https://s3.ibpad.com.br/redes/direita-esquerda/>>

(resocie.neutro@gmail.com). Nas configurações das contas do Gmail, as duas primeiras, “esquerda” e “direita” o gênero não foi especificado; para todas as cinco contas foi definida a mesma idade, 40 anos de idade. O treinamento dos perfis se seguiram até o final da campanha eleitoral diariamente, através da execução de um algoritmo em um computador exclusivo para esta tarefa.

2.1.3. Execução do treinamento

O algoritmo de treinamento do Google passou por duas fases principais em termos de frequência de execução: a execução manual e a execução automatizada por meio de técnicas computacionais. A primeira fase foi realizada por pesquisadores do grupo manualmente e feitas com frequência variada, de modo que a consistência nessa fase não pôde ser aferida posteriormente. A segunda fase, no entanto, foi, em parceria com a turma da Ciência da Computação sob a supervisão da professora Claudia Melo, através da elaboração de algoritmos e na sua execução diária. Em função da possibilidade de execução automatizada do “treinamento” e, posteriormente, da coleta dos dados, decidimos por multiplicar o número de contas. Antes da possibilidade de automatizar a simulação, possuíamos as cinco contas já mencionadas - “esquerda”, “direita”, “mulher”, “homem” e “neutro”. Com o desenvolvimento do código de “treinamento” e coleta de dados, decidimos triplicar cada uma das contas. Em outras palavras, passamos a possuir três contas de “esquerda”, de “direita”, “mulher”, “homem” e “neutro”, totalizando quinze (15) contas de usuários fictícios. A razão para o aumento deste número de contas decorre da possibilidade de comparar os resultados das busca entre as contas de cada tipo, de modo a verificar, por exemplo, se o treinamento entre as três contas de “esquerda” produziu um resultado de busca uniforme no Google.

Selecionamos um número n , aleatoriamente⁷ e calculamos o módulo⁸ dele por 3600. O resultado dessa operação representa, então, o número de segundos que o computador espera até repetir o treinamento das contas. Deste modo, fomos capazes de executar o algoritmo de treinamento durante todo o processo de execução do projeto de pesquisa que abrangeu o

⁷ Apesar de não podermos falar estritamente em aleatoriedade em sistemas computacionais, não entraremos nesta discussão neste texto devido ao escopo do mesmo, considerando que atingimos a máxima aleatoriedade possível num computador de uso pessoal.

⁸ A operação de módulo armazena numa variável o resto da divisão de um número por outro. Então, por exemplo, o módulo de 3 por 2 (representado por $3 \% 2$) é igual a 1.

período eleitoral brasileiro. Este algoritmo foi executado automaticamente pelo processo de automatização da pesquisa, que usou entre outras a biblioteca CasperJS⁹, PhantomJS¹⁰ e a versão 3.6 da linguagem Python.

2.2. Captura dos dados

O roteiro do algoritmo, para captura dos dados das buscas realizadas pelas contas criadas, foi desenvolvido por meio da raspagem de dados da página de resultados do Google em *hyper text markup language (html)*. A execução do roteiro foi por meio do *Selenium*, um programa que emula um navegador de Web que serviu de interface para acessar o serviço de busca da Google. O procedimento lógico que o algoritmo executou está exposto no Quadro 1:

QUADRO 1

1. Para cada conta-teste:

....1.1. Acessar <https://accounts.google.com>

....1.2. Autenticar-se com Login e Senha

....1.3. Para cada termo a ser pesquisado:

.....1.3.1. Digitar texto na caixa de busca

.....1.3.2. Esperar o resultado, fazer screenshot da tela e salvar HTML

....1.4. Para cada URL na página de resultado:

.....1.4.1. Acessar a URL

.....1.4.2. Esperar o resultado e fazer screenshot da tela

....1.5. Fazer logout da conta

Fonte: desenvolvido pelos autores

Conforme exposto pelo Quadro 1, o primeiro passo que o algoritmo realizou foi acessar, via *Selenium*, o endereço <http://accounts.google.com> e inserir seu respectivo login¹¹ e senha. Uma vez autenticada a conta pela Google, a conta abria a página do motor de busca e

⁹ <http://casperjs.org/>

¹⁰ <http://phantomjs.org/>

¹¹ Como foi mencionado, foram criadas 15 contas na Google para operacionalizar esta simulação. As contas eram: resocie.direita@gmail.com; resocie.direita2@gmail.com; resocie.direita3@gmail.com; resocie.esquerda@gmail.com; resocie.esquerda2@gmail.com; resocie.esquerda3@gmail.com; resocie.homem@gmail.com; resocie.homem2@gmail.com; resocie.homem3@gmail.com; resocie.mulher@gmail.com; resocie.mulher2@gmail.com; resocie.mulher3@gmail.com; resocie.neutra@gmail.com; resocie.neutra2@gmail.com; e resocie.neutra3@gmail.com.

se inseria o termo de busca¹². O resultado da busca, então, era ‘raspada’ e armazenada em um arquivo em formato *csv*.

Para organizar o banco de dados formado a partir dos arquivos *csv*, criamos um glossário explicando o significado das informações contidas no banco. O Quadro 2 apresenta o glossário e a Figura 1 relaciona cada uma das variáveis, referidas no Quadro 2, a imagem de uma página de resultado de busca da Google. A Figura 1 foi gerada a partir do resultado de busca, na Google, do nome do ex-candidato a Presidência da República, Ciro Gomes. As legendas, escritas em vermelho na própria imagem, identificam algumas das variáveis categorizadas, na Quadro 2, com um resultado prático de busca.

QUADRO 2	
newid	Identificador único do registro de cada captura. Exemplo: Perfil 1 no termo 1 = ID 01, Perfil 1 no termo 2 = ID02
oldid	Identificador de cada registro da base, gerado pelo Selenium.
search_date	Data em que a simulação da busca foi realizada
search_timestamp	Data e hora em que a simulação da busca foi realizada
searched_term	O termo buscado no Google durante a simulação.
used_profile	A conta do Google registrada no momento da simulação da busca
result_type	O tipo de resultado encontrado pelo searched_term inserido. Os resultados foram, pelo algoritmo executado, tipificados em forma de notícias, propaganda, resultado e video.
result_position	A posição do resultado de um tipo específico, na primeira página de resultados da simulação. Ver Figura 1.

¹² Os termos de busca utilizados para esta coleta até o dia 08 de outubro foram: “Bolsonaro”, “Lula”, “Marina Silva”, “Cabo Daciolo”; “Haddad”; “Alckmin”; “Boulos”; “Ciro Gomes”; “João Amoêdo”; “Álvaro Dias”; “Candidatos a Presidência”; “Eleições 2018”, “Local de votação”, após o dia 08 foram utilizados apenas, “Haddad”, “Bolsonaro”, “fraude nas urnas”, “local de votação”, “Candidatos a Presidência” e “Eleições 2018”.

result_url	A URL dos resultado de um tipo específico.
result_title	O título do resultado.
result_description	A descrição do resultado que aparece logo abaixo do título do resultado.
result_image	A imagem relacionada ao resultado, quando esta possui.
subresults	Subresultados são as “subseções” de um sítio eletrônico que o Google reconhece no processo de indexação.
is_ad	Valor binário que indica se o referido resultado específico é ou não um anúncio pago no Google.

Fonte: tabela organizada pelos autores

Figura 1 - Imagem da Primeira Página de Resultados de Busca da Google

The image shows a Google search interface for 'Ciro Gomes' on May 15-17, 2019. The search results are categorized into 'NOTÍCIA', 'VIDEO', and 'OTHER'. Annotations include:

- result_type**: Points to the 'Principais notícias' section.
- user_profile**: Points to the user's profile picture in the top right.
- result_position**: Points to the first three search results.
- result_title**: Points to the title of the first result: 'Ciro Gomes | Notícias | EL PAÍS Brasil'.
- result_description**: Points to the description of the first result: 'Notícias, entrevistas, artigos de opinião e informações sobre Flávio Bolsonaro, político do Ceará e pré-candidato do PDT à Presidência da República nas eleições ...'.
- result_image**: Points to the image of Flávio Bolsonaro in the second result.
- subresults**: Points to the 'Pesquisas relacionadas a ciro gomes' section at the bottom.

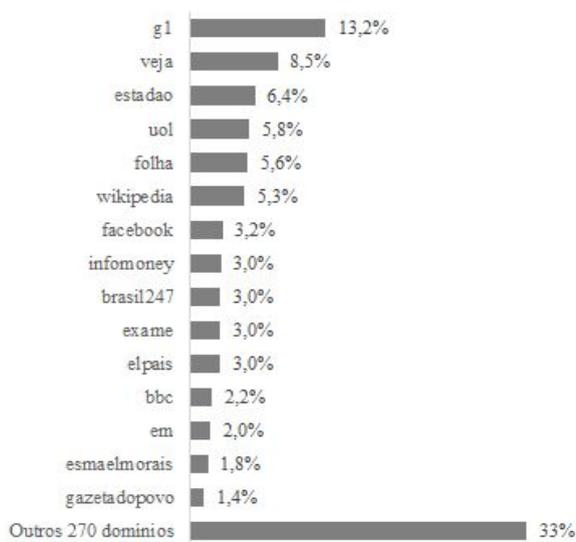
Fonte: elaborada pelos autores

3. Avaliação Preliminar

No escopo do presente artigo, apresentaremos uma análise preliminar do que identificamos com a simulação. Outra análise preliminar do banco de dados formado pela simulação pode ser encontrada no trabalho, também apresentado na *Compolítica8*, de Alana Fontenelle, Beatriz Franco e Gabriel Elias (2019), cujo trabalho, intitulado “‘Dá um ‘Google’ aí’: buscando a relevância do Google nas eleições brasileiras de 2018” (2019).

Ao todo foram realizadas 13.564 buscas durante o período de 17 de agosto a 30 de outubro de 2018 que contabilizaram 235.570 mil linhas na base de dados. Cada linha na base de dados é uma observação que representa um endereço eletrônico (URL), em uma posição específica na página de resultados, para um termo de busca de um dos logins feitos. Deste universo de URLs, identificamos que a quantidade de URLs únicas totalizam um universo 8.883 endereços eletrônicos distintos.

Gráfico 1 - Proporção de domínios para buscas aos candidatos



Dentre estes 8.883 URLs, quando analisados apenas os resultados referentes a notícias e resultados, excluindo, portanto, resultados de anúncio e de vídeos, apenas para as buscas dos candidatos, identificamos que existem 285 nomes domínios distintos. Ao analisarmos a frequência de quantas vezes determinado domínio apareceu, temos que os 15 principais domínios representaram 65% das buscas simuladas. O Gráfico 01 demonstra a proporção de cada domínio no conjunto total das buscas:

Tabela 1 - Quantidade de termos buscados por dia e mês.

Dia/Mês	Agosto	Setembro	Outubro
1		111	
2		53	95
3		373	179
4		332	103
5		273	366
6		272	375
7		428	193
8		0	155
9		137	138
10		192	150
11		131	145
12		186	99
13		336	109
14		196	135
15		243	110
16		275	144
17	325	333	180
18	214	368	163
19	216	319	132
20	124	303	151
21	395	250	143
22	399	349	72
23	603	203	
24	252	301	
25	20	266	
26		109	
27		314	
28	194	127	
29	113	144	75
30	144	57	113
31	59		

A Tabela 1, ao lado, mostra a quantidade de buscas feitas em cada dia dos meses de Agosto, Setembro e Outubro. As buscas se referem as realizadas por cada uma das 15 contas e 13 palavras-chaves até o dia 08 de outubro e os 15 perfis e 6 termos ao longo do segundo turno.

Como toda técnica de *webscraping*, houveram algumas dificuldades ao longo da captura: 1. Problemas de retorno do google e pausa no processo de coleta por razões ainda não conhecidas, 2. problemas de conexão com a internet; 3. problemas com os computadores responsáveis pela captura.

Por essas razões a tabela ao lado mostra diferentes quantidades de buscas por dia e em alguns dias as buscas não foram realizadas, como o período de 23/10 a 28/10.

Ao longo do processo o horário das buscas foram definidas de forma aleatória assim como a ordem de login de cada uma das contas e termos buscados. Isso foi um dos aprendizados do processo de coleta, pois se eventualmente o Google barrasse depois de determinada requisição, garantiríamos a aleatoriedade das buscas e perfis coletados.

4. Considerações finais

Como considerações finais do trabalho apresentado neste artigo, avaliamos que existem diferentes possibilidades de análise do banco de dados gerados a partir do método proposto. Uma dessas possibilidades é fazer a análise textual do conteúdo encontrado, no momento de coleta. Uma das categorias de dados que salvamos durante o processo de coleta foi o de "título", que se refere ao título de cada um dos resultados que aparecem para determinada conta ao realizar uma busca. Com essa informação podemos analisar como foi retratado o contexto geral das eleições, na medida em que os dados coletados foram obtidos durante a disputa eleitoral. Uma possível abordagem é entender como os dois principais atores do segundo turno das eleições estão relacionados um ao outro por meio de algoritmos de proximidade textual, dentro do campo do processamento de linguagem natural, como o Word2Vec¹³.

A segunda possibilidade é trabalhar com as análises e a diferença entre as médias de posicionamento das URLs para cada perfil/termo buscado. Nas análises preliminares existe indícios de que existe, sim, diferença entre os perfis. Algo que será abordado em um próximo paper.

A terceira possibilidade e a mais robusta de todas é a utilização da estatística de Kendall Tau (2014) para avaliar se existe diferença no ordenamento das páginas entre os perfis, para os mesmos termos em intervalos próximos de data.

5. Referências

Alvo, M., & Yu, Philip L.H. (2014), Statistical Methods for Ranking Data

PAN, Bing; HEMBROOKE, Helene; JOACHIMS, Thorsten; LORIGO, Lori; GAY, Geri; GRANKA, Laura. In Google We Trust: Users' Decisions on Rank, Position, and Relevance. **Journal of**

¹³ <https://skymind.ai/wiki/word2vec>.

Computer-Mediated Communication. Vol. 12, Nº 3. 1º de Abril de 2007. pp.801-823. Disponível em: <<https://academic.oup.com/jcmc/article/12/3/801/4582975>> Acesso 13 mar. 2019

EPSTEIN, Robert; ROBERTSON, Ronald E. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. **PNAS**. 04 de Agosto de 2015. Disponível em: <<https://www.pnas.org/content/early/2015/08/03/1419828112?sid=2096de02-a382-4aeb-89e0-df794a4ca03a>> Acesso 12 dez 2018

EPSTEIN, Robert. How Google Could Rig the 2016 Election. **Politico Magazine**. 19 de agosto de 2015. Disponível em: <<https://www.politico.com/magazine/story/2015/08/how-google-could-rig-the-2016-election-121548>> Acesso 13 mar. 2019

EPSTEIN, Robert. **The new mind control**. Aeon. Essay. 18 de Fevereiro de 2016. Disponível em: <<https://aeon.co/essays/how-the-internet-flips-elections-and-alters-our-thoughts>> Acesso 13 mar. 2019

ROBERTSON, Ronald E.. When will Google Defend Democracy?. **The Conversation**. 11 de Junho de 2018. Disponível em: <<https://theconversation.com/when-will-google-defend-democracy-96838>> Acesso 13 mar. 2019

ANEXO 1

Termos de busca para os perfis de “direita”

imposto zero, direito armas, pena de morte, ideologia de genero, escola sem partido, intervencao militar, marxismo cultural, direito a vida, reducao maioridade, prisao perpetua, esquerdopatas, petralhas, estado minimo, reducao impostos, ditadura venezuela, ditadura cuba, feminazi, lula ladrao, privatiza tudo, liberdade economica

Termos de busca para os perfis de “esquerda”

lula livre, reforma agraria, direito a moradia, passe livre, diretas ja, fora, temer, volta dilma, coracao valente, quem matou marielle franco?, legalizacao do aborto, socialismo, feminismo, pre-sal e nosso, descriminalizacao das drogas, laicidade do Estado, anula STF, luta contra o racismo, contra o golpismo, defesa da universidade publica, lute como uma menina

Termos de busca para os perfis “masculinos”

como fazer churrasco, tabela brasileirão, como dar um no de gravata?, roupas masculinas, câncer de próstata, saúde masculina, impotência sexual, viagra, paternidade acessórios masculinos, ejaculação precoce

Termos de busca para perfis “femininos”

como amamentar?, roupas femininas, sintomas de menopausa, tensão pré-menstrual, cólica menstrual, maternidade, depressão pós-parto, câncer de mama, câncer de útero, saúde feminina, acessórios femininos.