XX ENANCIB
21 a 25 Outubro/2019 – Florianópolis
A Ciência da Informação e a era da Ciência de Dados
ISSN 2177-3688

GT-5 – Política e Economia da Informação

ESTUDO SOBRE ALGORITMOS EM CAMPANHAS ELEITORAIS: ANÁLISE COM IRAMUTEQ

STUDY ON ALGORITHMS IN ELECTION CAMPAIGNS: ANALYSIS WITH IRAMUTEQ

Marco Schneider – Instituto Brasileiro de Informação em Ciência e Tecnologia
Marcos Ramos – Instituto Brasileiro de Informação em Ciência e Tecnologia
Priscila Carvalho – Instituto Brasileiro de Informação em Ciência e Tecnologia

Modalidade: Resumo Expandido

Resumo: O estudo relata a pesquisa empírica durante 2008-2018, com base no banco de dados Scopus, por meio de técnicas da Scientometrics para identificar, quantificar e analisar a produção científica sobre o uso de algoritmos em campanhas eleitorais. Além disso, a análise de conteúdo dos resumos dos artigos foi realizada com o software Iramuteq, de código aberto e gratuito. O resultado do estudo mostrou a eficiência das palavras-chave escolhidas na recuperação de informações, devido à identificação da convergência dos temas dos artigos. Além disso, reconheceu-se a necessidade de mais estudos sobre o assunto devido aos possíveis efeitos dos algoritmos na democracia participativa.

Palavras-Chave: Algoritmo; Campanha eleitoral; Eleição; Iramuteq.

Abstract: The study reports the empirical research during 2008-2018 based on the Scopus database, through Scientometrics techniques to identify, quantify and analyze the scientific production on the use of algorithms in election campaigns. In addition, content analysis of the abstracts of the articles was performed using the free and open-source Iramuteq software. The result of the study showed the efficiency of the chosen keywords in information retrieval, due to the identification of the convergence of the themes of the articles. In addition, it was recognized the need for further study on the subject due to the possible effects of algorithms on participatory democracy.

Keywords: Algorithm; Election campaign; Election; Iramuteq.

1 INTRODUÇÃO

Os questionamentos sobre o papel e os efeitos dos algoritmos na sociedade tornaram-se mais presentes em pesquisas científicas devido ao seu uso massivo como, por exemplo, o emprego de robôs para disseminação de informações em redes sociais; uso na coleta de dados dos cidadãos; o surgimento de aplicações inteligentes capazes de manipular áudio e imagem para criação de vídeos; uso na previsão de comportamento humano e de resultados em campanhas eleitorais.

Segundo Finn (2017),

> [...] a palavra algoritmo frequentemente abrange uma variedade de processos computacionais, incluindo vigilância rigorosa do comportamento do usuário, agregação de "*big data*" das informações resultantes, mecanismos de análise que combinam várias formas de cálculo estatístico para analisar dados e, finalmente, um conjunto de ações, recomendações e interfaces voltadas para o ser humano que geralmente refletem apenas uma pequena parte do processamento cultural acontecendo nos bastidores. (FINN, 2017, p.16)[1].

O'Neil (2016) afirma que os algoritmos já tomam decisões importantes que podem ter impactos profundos na vida das pessoas, porém os modelos matemáticos são opacos e o seu funcionamento invisível. "O modelo em si é uma caixa preta, seu conteúdo um segredo corporativo ferozmente guardado." (O'NEIL, 2016, p.14)[2]. Por isso, é importante que sejam criados meios de se promover mais transparência.

Diante dos paradigmas e desafios gerados pelo uso de algoritmos, o presente artigo buscou ilustrar a pesquisa empírica com base na Cientometria realizada na Scopus, durante 2008 a 2018, com objetivo de identificar, quantificar e analisar a produção científica do tema algoritmos na política, em particular, em campanhas eleitorais.

Utilizou-se o método de Análise de Conteúdo nos resumos dos artigos recuperados na base de dados através do programa de análise textual, gratuito e de código aberto chamado Iramuteq (*Interface de R pour analyses Multidimensionnelles de Textes et de Questionnaires*), criado por Pierre Ratinaud no Laboratório de Estudos e Pesquisa Aplicada em Ciências Sociais em Toulouse, nas linguagens de programação R e Python. (CAMARGO; JUSTO, 2013).

---

[1] Tradução nossa: […]the word algorithm frequently encompasses a range of computational processes including close surveillance of user behaviors, "big data" aggregation of the resulting information, analytics engines that combine multiple forms of statistical calculation to parse that data, and finally a set of human-facing actions, recommendations, and interfaces that generally reflect only a small part of the cultural processing going on behind the scenes.

[2] Tradução nossa: The model itself is a black box, its contents a fiercely guarded corporate secret.

2 PROCEDIMENTOS METODOLÓGICOS

Para alcançar os objetivos da pesquisa utilizou-se a Cientometria como metodologia por ser voltada para os aspectos quantitativos da ciência enquanto uma disciplina (MACIAS-CHAPULA, 1998, p.134). Ademais, usou-se o conceito de Análise do Conteúdo de Bardin (2016) que, no prefácio, define como "um conjunto de instrumentos metodológicos cada vez mais sutis em constante aperfeiçoamento, que se aplicam a discursos (conteúdos e continentes) extremamente diversos" (BARDIN, 2016, p.15).

As variáveis de pesquisa foram escolhidas pelo processo de indexação por assunto, partindo do binómio conceito-termo, que por análise reconhecem os conceitos do conteúdo temático de um documento (MENDES; SIMÕES, 2002, p.23), bem como considerou-se a questão da linguagem natural e o meio digital na recuperação da informação relevante (MOENS, 2002, p.16). Assim, os termos selecionados para recuperação dos artigos foram: *algorithm, political campaign, election campaign e presidential election*.

A base de dados selecionada foi a Scopus da Elsevier, devido ao padrão de curadoria de dados e pela sua abrangência mundial. Já para escolha do período foi utilizado como referência o artigo de Nilckerson e Rogers (2013), em que destacaram o avanço significativo do uso de algoritmos e *big data* a partir da campanha de Barack Obama em 2008, por isso, foram coletados os dados de 2008 até 2018, no dia 09 de junho de 2019 (NICKESON; ROGERS, 2013).

3 ANÁLISE DE RESULTADOS

O resultado da recuperação foi de 190 artigos, mas após tratamento dos dados verificou-se 40 títulos repetidos, portanto, a análise final foi composta por 150 artigos.

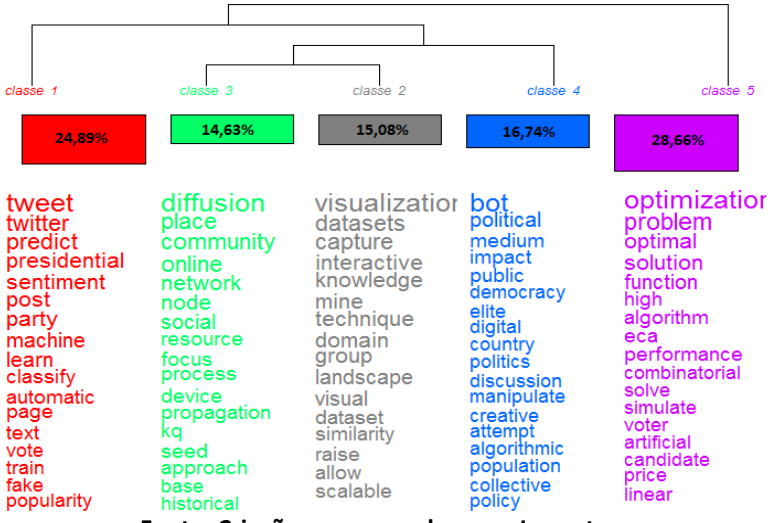
A primeira etapa foi a análise estatística do *corpus* textual que apresentou: 150 textos; 738 segmentos de textos; 25.424 ocorrências de palavras divididas em 2.801 ativas (substantivo, adjetivos e advérbios), 324 suplementares (artigos, pronomes e preposições) e 1.334 palavras com ocorrência única.

Os segmentos de textos são as principais unidades de análise textual do Iramuteq, normalmente formados por três linhas, originados a partir da frequência das palavras, apresentando vocabulário semelhante entre si e vocabulário diferente dos segmentos de textos das demais classes (CAMARGO; JUSTO, 2013).

A segunda etapa foi a Classificação Hierárquica Descendente (CHD) representada no dendrograma, com resultado de 5 classes compostas por 663 segmentos de textos: classe 1 (vermelho) com 165 segmentos; classe 2 (cinza) com 100 segmentos; classe 3 (verde) com 97 segmentos; classe 4 (azul) com 111 segmentos; e classe 5 (roxa) com 190 segmentos.

A análise de CHD utiliza a correlação das palavras em segmentos de textos no *corpus* textual, comparando com a lista de formas reduzidas, dividindo os segmentos com relação à frequência de palavras e apresentando como resultado o esquema hierárquico de classes.

Figura 1: Dendrograma de Classificação Hierárquica Descendente



Fonte: Criação nossa com base no Iramuteq.

As palavras em destaque de cada classe foram as seguintes: classe 1 (vermelho) *tweet*; classe 2 (cinza) *visualization*; classe 3 (verde) *diffusion;* classe 4 (azul) *bot;* e classe 5 (roxa) *optimization*. Além disso, notou-se que o termo *algorithm* apareceu na classe 5 em uma posição mais destacada do que na classe 4.

A análise da classe 1 mostrou os temas: análise de *tweets*; classificação de sentimento de *tweets*; análise de redes sociais para prever orientação política; uso de aprendizado de máquina para identificar *fake news*; análise da opinião pública no Twitter; redes neurais para análise de dados eleitorais; análise linguística da frequência de palavras; classificação de memes; e *software* de coleta de dados de blogs.

A análise da classe 2 apresentou os temas: linguagem natural e redes neurais para análise de dados; mapeamento de conversas em redes sociais; visualização de dados em *cluster*s; visualização para identificar padrões de associação de dados espaciais; mineração de dados em redes sociais; uso de *colonel blotto game* para competições e eleições; aplicação da teoria dos jogos; uso de algoritmos na desinformação em campanha eleitoral.

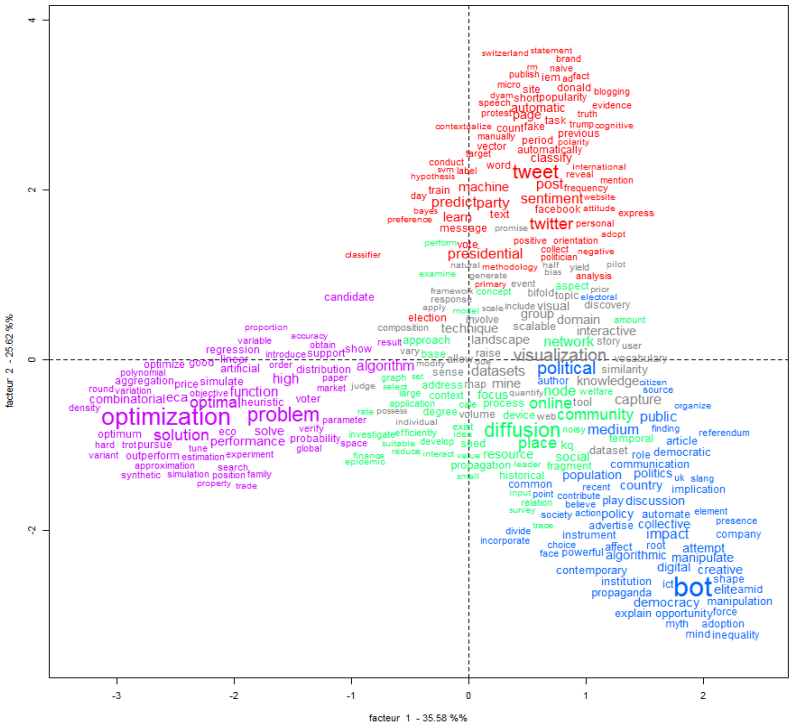
A análise da classe 3 revelou os temas: teoria de ação; desenvolvimento de modelos confiáveis e replicáveis para aplicação em *big data*; algoritmos para analisar fatores sociais e políticos; análise de influenciadores em comunidades online; estudos do comportamento em redes sociais; estudo da dinâmica das ações coletivas no cyberespaço.

A análise da classe 4 apontou os temas: estudos de redes sociais; análise de blogesfera da comunidade feminina mulçumana; análise de movimentos coletivos no cyberespaço; análise e identificação de robôs no Twitter; análise de impacto de campanhas publicitárias em redes sociais; ferramentas de comunicação para manipulação das discussões no cyberespaço; análise de influência negativa do robôs nas discussões; identificação de tipos de estratégias de mobilização política e de militância; comparativo entre o volume de investimento *versus* o resultado eleitoral; estudo de como identificar e retirar conteúdos de desinformação online.

A análise da classe 5 indicou os temas: aplicação de teoria estatística de *Bayes*; uso de algoritmo genético para soluções de problemas de otimização; uso de algoritmo *kruskal tenson decomposition* para análise linguística de *tweets;* otimização de algoritmo na eleição presidencial; otimizar a performance de algoritmos evolutivos e genéticos; avaliação da robustez do modelo de simulação de conjuntos de dados; teoria matemática para aprendizagem de algoritmo de predição de mercados; algoritmo heurístico que simula comportamento necessário para que o candidato obtenha o maior nível de apoio na eleição.

A terceira etapa foi a Análise Fatorial de Correspondência que representa as diferentes palavras e variáveis associadas a cada uma das classes.

Figura 2: Análise de Fator de Correspondência

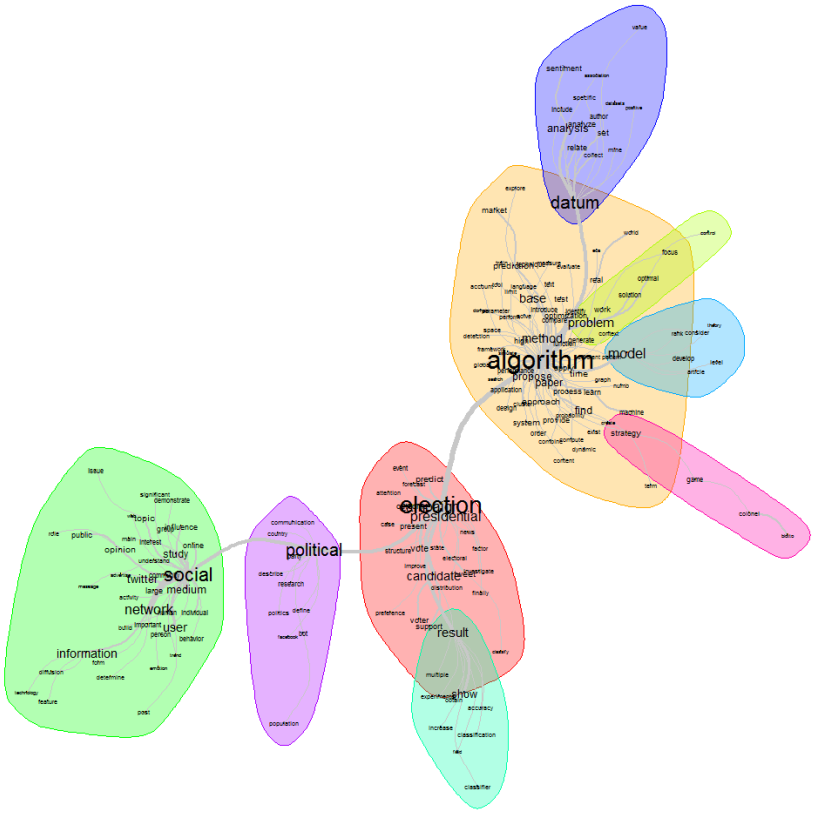


Fonte: Criação nossa com base no Iramuteq.

Observou-se a centralidade das classes 2 (cinza) e 3 (verde), bem como que os termos possuem maior conexão entre si do que os demais localizados nas extremidades. Identificou-se que na classe 1 (vermelho), *tweet* apresenta um campo semântico cuja centralidade é composta pelos termos: *post, sentiment, classify, predict party, presidential, vote e facebook.* Na classe 4 (azul), *bot* demonstrou uma forte ligação semântica com os termos *propaganda, digital, democracy, creative, elite, manipulation e algorithm*. A classe 5 (roxa) apontou um alinhamento dos termos *optimization* e *problem* refletindo a forte ligação entre eles, pois sem problema não há otimização. Além disso, notou-se a ligação entre os termos *solution, performance, optimal, probability, heuristic, combinatorial, function* e *algorithm.*

A quarta etapa foi a Análise de Similitude que se refere a co-ocorrência de palavras no corpus textual reproduzida abaixo.

Figura 3: Análise de Similitude



Fonte: Criação nossa com base no Iramuteq.

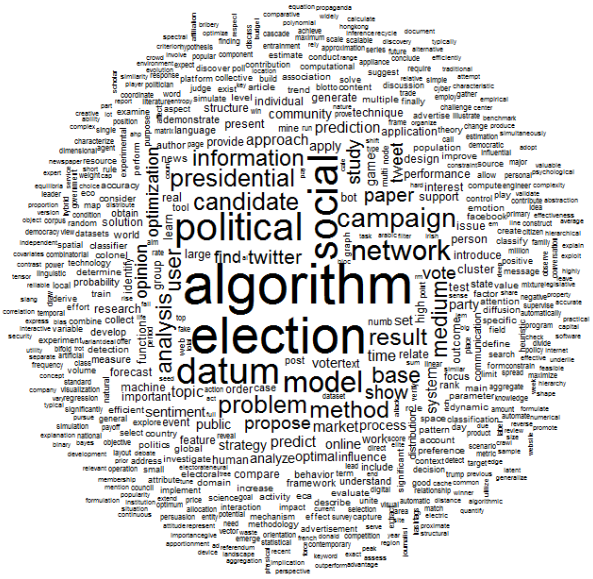
Percebeu-se que os termos *algorithm, election, political* e *social* representam *clusters*, sendo que o *cluster algotithm* possui quatro sub *clusters datum, problem, model e strategy*. Ademais, inferimos a existência de uma relação de dependência semântica por meio da convergência dos termos dependentes em relação ao termo central *algorithm*.

Notou-se que o *cluster election* possui um sub *cluster result* evidenciando uma forte presença e ligação de palavras que se referem ao tema eleição tais como: *presidential, candidate, vote, result* e *show.* Assim, deduziu-se que os artigos recuperados na Scopus têm relação com estudos científicos sobre o período eleitoral.

Cabe mencionar que a análise de Similitude destacou termos em um alinhamento léxico que sugere a seguinte tradução semântica: algoritmos em eleições presidenciais, na política e em redes sociais.

A quinta análise foi a frequência de palavras ilustradas em uma nuvem, tratando-se de uma análise lexical simples, porém graficamente bastante interessante na medida em que possibilita rápida identificação das palavras-chave de um corpus textual (CAMARGO; JUSTO, 2013).

Figura 4: Nuvem de Palavras



Fonte: Criação nossa com base no Iramuteq.

A palavra central da nuvem é *algorithm,* mas a palavra *election* está diretamente ligada e equilibrada no que tange à ligação semântica e representatividade no corpus textual. Além disso, notou-se a formação de um círculo semântico composto pelos termos: *datum, model, problem, method, base, show, result, network, campaign, social, political, candidate, presidential, information*.

4 CONSIDERAÇÕES FINAIS

A pesquisa identificou que palavras escolhidas para recuperação de artigos na Scopus foram eficientes, pois resgataram 190 artigos, sendo apenas 21% repetidos. Ademais, a análise do corpus textual pelo programa Iramuteq demonstrou correspondência positiva na identificação dos termos que se referem ao objeto de estudo.

Notou-se que os temas discutidos nos artigos eram em sua maioria convergentes, pois estavam voltados para soluções de problemas tais como: criação e otimização de modelos algorítmicos confiáveis e replicáveis de predição de resultados; e coleta de dados em redes sociais, em particular o Twitter, usado para análise do comportamento de indivíduos com o propósito de prever a tendência ou atitudes dos eleitores.

Vale pontuar que a recorrência dos temas reflete a importância dos estudos sobre algoritmos em campanhas eleitorais, visto que algumas práticas como *big data* e redes sociais

já são amplamente utilizadas com objetivo de aumentar os lucros das empresas, porém no âmbito político o uso indiscriminado de algoritmos pode vir a comprometer a democracia.

Acreditamos que os dados podem fornecer subsídios úteis para outras pesquisas em torno da mesma problemática. Nesse sentido, o estudo prevê a disponibilização dos dados de pesquisa no repositório Zenodo do Laboratório em Rede de Humanidades Digitais[3].

REFERÊNCIAS

BARDIN, Laurence. **Análise de conteúdo**. São Paulo: Almedina, 2016.

CAMARGO, Brígido V.; JUSTO, Ana Maria. **IRAMUTEQ: um software gratuito para análise de dados textuais**. Temas em Psicologia, v. 21, n. 2, Ribeirão Preto, 2013, p. 513-518. Disponível em: http://dx.doi.org/10.9788/TP2013.2-16. Acesso em: 5 ago 2019.

FINN, Ed. **What algorithms want:** imagination in the age of computing. Cambridge, MA: MIT Press: 2017.

MACIAS-CHAPULA, Cesar A. **O papel da informetria e da cienciometria e sua perspectiva nacional e internacional.** Ci. Inf. [online]. vol.27, n.2, 1998. ISSN 0100-1965. Disponível em: http://dx.doi.org/10.1590/S0100-19651998000200005. Acesso em: 9 ago 2019.

MENDES, Maria Teresa P.; SIMÕES, Maria da Graça. **Indexação por assuntos: princípios gerais e normas**. Lisboa: Gabinete de Estudos a&b, 2002. (Estudos a&b; 1). ISBN 972-98827-0-3. Disponível em: https://estudogeral.sib.uc.pt/handle/10316/20805. Acesso em: 9 ago 2019.

MOENS, Marie-Francine. **Automatic indexing and abstracting of documents texts**. Kluwer Academic Publishers, Boston, MA, 2002.

NICKERSON, David; ROGERS, Todd. **Political Campaigns and Big Data**. HKS Working Paper No. RWP13-045, February 25, 2014. Disponível em:http://dx.doi.org/10.2139/ssrn.2354474. Acesso em: 5 ago 2019.

O'NEIL, Cathy. **Weapons of Math Destruction**: How Big Data Increases Inequality and Threatens Democracy. New York: Crown-Archetype, 2016.

[3] Repositório do Laboratório em Rede de Humanidades Digitais. Disponível em:<https://zenodo.org/communities/larhud/about/>. Acesso em: 10 ago 2019.