

Fall 10-16-2021

Science Recommendations on Social Media Platforms

Jorge Oliveira e Sa Prof.
University of Minho, jos@dsi.uminho.pt

José Luís Pereira
Universidade do Minho, Centro Algoritmi, jlp@dsi.uminho.pt

Pedro Fontes
Universidade do Minho, a76170@alunos.uminho.pt

Follow this and additional works at: <https://aisel.aisnet.org/capsi2021>

Recommended Citation

Oliveira e Sa, Jorge Prof.; Pereira, José Luís; and Fontes, Pedro, "Science Recommendations on Social Media Platforms" (2021). *CAPSI 2021 Proceedings*. 16.
<https://aisel.aisnet.org/capsi2021/16>

This material is brought to you by the Portugal (CAPSI) at AIS Electronic Library (AISeL). It has been accepted for inclusion in CAPSI 2021 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Recomendações de Ciência em Plataformas de Redes Sociais

Science Recommendations on Social Media Platforms

Jorge Oliveira e Sá, Centro ALGORITMI, Universidade do Minho, Portugal, jos@dsi.uminho.pt

José Luís Pereira, Centro ALGORITMI, Universidade do Minho, jlmp@dsi.uminho.pt

Pedro Fontes, Universidade do Minho, Portugal, a76170@alunos.uminho.pt

Resumo

As plataformas de redes sociais podem ser vistas como fontes de recomendações de variados itens ou produtos. O *marketing* aproveita ao máximo esta potencialidade sobretudo ao tirar proveito de *influencers*, que ao experimentarem esses itens fazem avaliações dos mesmos, isto pode ser visto como recomendações. No entanto, pode-se questionar porque é que na área da ciência isto não ocorre de uma forma semelhante. Este artigo apresenta um estudo em que se recolheram partilhas/menções em plataformas de redes sociais sobre documentos científicos (artigos de revistas ou conferências, relatórios, patentes, entre outros) e dessa recolha foi obtida a polaridade desses comentários (positivos, negativos e neutros). Verificou-se que a maior parte das menções são neutras e algumas consideradas positivas. Isto permite concluir que há partilhas de menções sobre ciência e que podem servir como base de um sistema de recomendação de ciência.

Palavras-chave: sistema de recomendação; plataformas de redes sociais; análise de sentimento; ciência

Abstract

Social media platforms can be a source of recommendations for various items or products. Digital Marketing takes advantage of this, especially when using influencers to evaluate their items or products on social media platforms. Each influencer can be a recommender of items or products. However, one can question why this does not occur similarly in the science area. This work presents a study based on comments shared on social media platforms about scientific documents (for example, journal or conference articles, thesis, reports, and patents). We pretend to obtain the polarity (positive, negative, and neutral) for each scientific comment. Most mentions are neutral and some considered positive, which allows us to conclude that there are shares of comments about science and that they can serve as the basis for a science recommendation system.

Keywords: recommendation system; social media platform; sentiment analysis; science.

1. INTRODUÇÃO

A partilha de conhecimento científico é realizada, tradicionalmente, através de publicações em revistas científicas ou livros, sujeitos à revisão, à edição, à impressão e à distribuição com longos períodos de espera entre cada etapa, a restrições de acesso e a custos de adesão muito elevados. Com o desenvolvimento e a proliferação da Internet enquanto plataforma aberta, livre e global de comunicação apresenta um grande potencial para a comunicação de ciência. A Internet veio

revolucionar o acesso à ciência e permitiu colmatar as barreiras geográficas entre investigadores de diferentes países e centros de investigação, contribuindo para o desenvolvimento de uma comunicação mais efetiva e produtiva, garantindo uma ciência coletiva (Montgomery, 2009). O próprio desenvolvimento da ciência foi assim alterado com a integração da comunidade científica numa rede global (Silva, 2002) uniformizada por campos de atuação e a constituição de redes transdisciplinares de colaboração entre diversas entidades científicas e investigadores de diferentes áreas e países, sendo uma realidade atual a realização de consórcios e parcerias para a realização de projetos a nível europeu e/ou internacional. O fluxo da comunicação científica inclui deste modo a produção e publicação dos resultados da investigação realizada, divulgação para a comunidade científica e público em geral; o acesso à literatura publicada, a comunicação informal entre os próprios investigadores, comunicação com o público e a incorporação de novos conhecimentos (Morris & Ogan, 1996). Watermeyer (2010) e Bik & Goldstein (2013) referem que o contacto *online* via redes sociais é um contacto bidirecional e permite uma ligação informal entre cientistas e entre estes com o público.

A ciência pode, assim, ser enriquecida através de diferentes representações do conhecimento científico, de uma multiplicidade de opiniões e críticas que são construídas e partilhadas por toda a rede. Em suma, o fenómeno das redes sociais como consequência natural do carácter social do Homem, destaca-se no meio tecnológico de comunicação e interação social como potencial ferramenta para o acesso, partilha, difusão de conhecimento.

Desta forma, o incremento das publicações científicas (neste artigo as publicações científicas serão denominadas de resultados de investigação, aqui denominado de Research Output (RO), um RO pode ser um artigo de revista, um artigo de conferência, um livro, um capítulo de um livro, um relatório científico, uma patente, entre outros) nos últimos anos provocou a necessidade de haver recomendações sobre que ROs se devem ler. Usualmente os ROs passam por processos de avaliação pelos pares, denominados de *peer-review*, o resultado do trabalho dos *peer-review* normalmente não é tornado público, criando uma opacidade na perceção da relevância dos ROs. Isto, provoca que encontrar o RO adequado a quem procura no elevado número de ROs publicados uma tarefa complexa e árdua, sobretudo se quem procura ainda é um iniciante numa determinada área de conhecimento.

A utilização de plataformas de redes sociais leva a que indicadores bibliométricos, nomeadamente o nº de citações, tradicionalmente usados para avaliar a relevância de um RO, sejam complementados com indicadores denominados de *altmetrics* (métricas alternativas) baseadas em partilhas nas redes sociais, *likes*, visualizações, entre outros. As *altmetrics* permitem medir a atenção que um RO está a provocar tanto na sociedade em geral bem como na comunidade científica (isto

foi observado neste tempo de pandemia), dessa forma, poderão auxiliar os investigadores nos seus processos de investigação a identificar ROs mais relevantes.

Este artigo tem como objetivo recolher partilhas e menções efetuadas em plataforma de redes sociais para perceber se as plataformas de redes sociais são utilizadas como meios de divulgar a ciência e sobretudo se a comunidade científica utiliza essas plataformas de redes sociais para fazerem menções sobre ROs. Para essas menções irá ser analisada a sua polaridade de forma a identificar as positivas, neutras e negativas. Assim, será possível verificar se essas menções podem servir como base de um sistema de recomendação de ROs suportado pela experiência dos investigadores e no seu envolvimento social numa comunidade científica.

2. ESTADO DA ARTE

2.1. Recomendação

A Internet e os serviços *web* têm aumentado exponencialmente nas últimas décadas; um excedente de informações agora está agora acessível para todos. Pode ser um desafio para os utilizadores filtrarem todas estas informações para conseguirem retirar os seus aspetos essenciais. Muitas empresas de comércio eletrónico *online* recomendam produtos aos seus utilizadores, vendendo milhões de produtos através de uma plataforma *web*. Para um utilizador comum, navegar através de todas estas possibilidades pode ser um trabalho exaustivo. Os sistemas de recomendação visam resolver este problema de sobrecarga de informação enquanto personalizam a experiência do utilizador entregando recomendações precisas e personalizadas de itens/produtos de acordo com as suas preferências.

A opinião é fundamental em quase todas as atividades humanas, pois é através delas que o comportamento das pessoas é influenciado. Os clientes querem saber a opinião de outros sobre um determinado produto, antes de o comprarem e usam as plataformas de redes sociais para obterem essas opiniões (Keith Norambuena, Lettura, & Villegas, 2019).

2.2. Algoritmos de Recomendação

A filtragem colaborativa avalia produtos utilizando avaliações (implícitas ou explícitas) dadas por utilizadores. Se o utilizador A e o utilizador B avaliam itens em comum, então as suas preferências são consideradas semelhantes, conseqüentemente, se existirem produtos no histórico do utilizador A que não se encontram no histórico do utilizador B, então estes produtos podem ser recomendados ao utilizador B, por outras palavras, a filtragem colaborativa é o processo de recomendar itens utilizando a opinião dos outros (Xu, Jiang, Chen, Ren, & Liu, 2019). Existem duas categorias de filtragem colaborativa (Sen & Udgirikar, 2013):

- baseada no utilizador - o sistema de recomendação utiliza os perfis de utilizadores semelhantes para fazer a recomendação, procuram vizinhos mais próximos e, de acordo com os interesses destes, os interesses do utilizador são obtidos (Xu et al., 2019); e
- baseada no item - se um utilizador avaliar um item de forma positiva, o sistema de recomendação identifica itens semelhantes candidatos para, posteriormente serem recomendados.

Recomendações baseados em conteúdo permitem construir o perfil de um utilizador para prever produtos não vistos pelo mesmo, utilizando *tags* e palavras-chave. Não requer dados de outros utilizadores, ou seja, apenas requer a análise dos itens e perfil do utilizador para fazer as recomendações (Fayyaz, Ebrahimian, Nawara, Ibrahim, & Kashef, 2020).

A geração de uma recomendação inicia-se com a seleção dos N itens mais relevantes para o utilizador, de seguida procura a similaridade entre as características existentes no perfil dos utilizadores e dos Research Outputs (ROs) (Jomsri, Sanguansintukul, & Choochaiwattana, 2010). Outros trabalhos apresentam alterações neste algoritmo, pois, para além dos itens mais relevantes, fornecem recomendações fortuitas (serendipidade) que são itens de áreas diferentes ou distantes (Sugiyama & Kan, 2011), mas que podem ser úteis para os utilizadores. O último passo do sistema de recomendação é classificar os elementos desta lista por uma determinada ordem e os N elementos do topo dessa lista serão recomendados ao utilizador.

Existem ainda outros algoritmos de recomendação, nomeadamente:

- baseados em dados demográficos – permite recomendar itens com base em características demográficas de um utilizador: idade, género, nacionalidade, ocupação etc. Geralmente, recolhem estes dados durante o registo dos utilizadores onde são também apresentados uma série de itens para serem avaliados de forma a construir o seu perfil (Al-Shamri, 2016);
- baseados em utilidade – permite recomendar itens com maior utilidade para o utilizador, tendo por base a utilidade calculada, baseada na Multi-Attribute Utility Theory (Deng, 2015; Fayyaz et al., 2020);
- baseados em conhecimento – permite recomendar itens com base nas preferências do utilizador, estes sistemas não necessitam de muita informação sobre um utilizador uma vez que são independentes dos gostos do utilizador e são também independentes das avaliações dos mesmos, dessa forma utilizam o conhecimento explícito dos itens e dos utilizadores para gerar recomendações (Devchand, Sheehan, Gallivan, Tuncer, & Nicols, 2017). Estes sistemas são vistos como sistemas independentes e podem também são considerados complementares a outros tipos de sistemas de recomendação (Fayyaz et al., 2020);

- híbridos – consistem na combinação de duas ou mais técnicas anteriores para obter um melhor desempenho e precisão na recomendação. O seu objetivo principal é eliminar as desvantagens individuais de cada uma das técnicas apresentadas (Fayyaz et al., 2020). Por exemplo, uma recomendação baseada em colaboração com uma recomendação baseada em conteúdo (Fayyaz et al., 2020), isto permite que a recomendação diminua a sensibilidade ao número de utilizadores que avaliaram o item (Sarwar, Karypis, Konstan, & Riedl, 2002).

2.3. Sistema de recomendação baseado em menções

Destaca-se este sistema de recomendação, pois baseia-se em classificações (*ratings*) e menções sobre itens partilhados por utilizadores em plataforma sociais. Estes sistemas são utilizados para recomendar hotéis, restaurantes, e outros serviços afins (Ramzan, Bajwa, Jamil, & Mirza, 2019).

O sistema funciona de duas maneiras paralelas entre si:

1. as classificações numéricas e votos dos itens são recolhidos e harmonizados;
2. os textos das menções são processados através de Processamento de Linguagem Natural (PLN) de forma a se obter a *polarity*. Se o valor da *polarity* de uma cara característica de item for superior a 0, então essa característica é avaliada de forma positiva, caso seja igual ou inferior a 0 é avaliada de forma negativa. Assim, a *polarity* das menções de um item pode ser calculada através da do somatório da *polarity* das menções de todas as características de um item.

2.4. Sistema de recomendação baseado em Influencers

Este sistema de recomendação surge com a Internet e é uma forma de chegar aos consumidores, pois é muito mais difícil alcançá-los através dos meios de comunicação tradicionais. A era de hoje é caracterizada pelo fato de que, para atender as necessidades do mercado, os consumidores dispõem de um ambiente extremamente amplo e gama de produtos diferentes. Esse fato estimula a percepção do consumo como uma parte importante da vida social. O crescimento das redes sociais renovou por completo a forma como as pessoas interagem, comunicam e se envolvem (Arora, Bansal, Kandpal, Aswani, & Dwivedi, 2019). As redes sociais tornaram-se grandes plataformas para compartilhar informações pessoais, notícias, fotos, vídeos e são plataformas de comunicação essenciais, uma vez que facilita as interações entre os seus utilizadores, o que pode influenciar ou afetar as opiniões dos outros (Bamakan, Nurgaliev, & Qu, 2019). Pesquisas e evidências aplicadas sugerem que líderes de opinião online são importantes promotores de produtos e serviços em diferentes áreas de negócios e no próprio marketing (Byrne, Kearney, & MacEvilly, 2017; Dhanesh & Duthler, 2019). Desta forma, surge aqui um conceito mais recente: *influencer*.

Este conceito é entendido como a prática de *marketing digital* que tira proveito de utilizadores *online* com uma boa gama de seguidores, que são capazes de influenciar as atitudes dos consumidores e os processos de tomada de decisão em favor de marcas ou ideias (Wielki, 2020). De acordo com Kartajaya, Kotler, & Hooi (2019) os *influencers* têm um grande número de seguidores e as suas opiniões influenciam esses seguidores. É também aceite que os *influencers* podem criar conteúdo (Babin & Hulland, 2019) para construir a sua reputação e são considerados *experts* nas suas comunidades de seguidores. Os *influencers* podem também ser vistos como poderosas marcas humanas que impactam positivamente o desempenho das empresas associadas a eles (Ki, Cuevas, Chong, & Lim, 2020).

As empresas encontraram valor e utilidade no uso de *influencers* tornando-os porta-vozes das suas marcas uma vez que podem ser vistos como líderes de opinião. Cada vez mais as empresas alocam mais recursos financeiros no *marketing* dos *influencers*. Um estudo (Nielsen, 2016) indica que atividades envolvendo *influencers* contribuem para o aumento das vendas de produtos. Assim, há indicadores que importa analisar: 71% dos consumidores são mais propensos a fazer compras de produtos com base em menções nas redes sociais (Ewing, 2011); 78% das compras dos consumidores são impactadas pelas publicações das empresas nas redes sociais (Olenski, 2012). Outro estudo, indica que 80% dos entrevistados fizeram compras recomendadas por *influencers* através de um *link* ou imagem fornecido por este (Rakuten Marketing, 2019).

Pode-se ver um *influencer* como um recomendador, pois as pessoas ao seguirem um *influencer* estão a fazê-lo porque sentem algum tipo de confiança e de prazer com o conteúdo que ele cria e devido a esse sentimento de confiança gerado, a pessoa aceita as menções do *influencer* como recomendações.

2.5. Recomendação aplicados à Ciência

A filtragem colaborativa é utilizada para recomendar ROs baseada num sistema de *tags* (Parra-Santander & Brusilovsky, 2010). Podem ser identificadas variantes em que a similaridade entre dois utilizadores é calculada utilizando a correlação de Pearson sobre as suas avaliações de itens em comum (Schafer, Frankowski, Herlocker, & Sen, 2007) ou através de modelos probabilísticos não binário, de forma a tirar proveito do conjunto das *tags* dos utilizadores e o modelo BM25 é proposto para calcular a relevância de um RO (Klampanos, Manning, Raghavan, & Schütze, 2009).

A recomendação baseada em conteúdo é também aplicada ao contexto da ciência. Os ROs podem ser vistos como itens e os investigadores como utilizadores, através das citações de um investigador pode se iniciar a construção do seu perfil e criar a sua lista de preferências ou interesses através de palavras-chave que este usa nos seus ROs. O sistema de recomendação pode extrair as palavras-

foi efetuado através de uma conta *developer*. Depois dos acessos configurados iniciou-se o desenvolvimento da API em Python através da biblioteca *tweepy*, os recursos utilizados para a criação foram o *OAuthHandler*, *StreamListener*, e *Stream* todos eles pertencentes à biblioteca *tweepy*.

O *OAuthHandler* permite a autenticação no Twitter e vai precisar dos dados que foram criados aquando da criação da aplicação, pois é necessário a autenticação das chaves geradas *consumer key*, *consumer secret key*, *access_token*, *access token secret*. Quando a autenticação é concluída, é criado um canal de *streamListener* onde é recebida toda a informação que é criada aquando da criação de um *tweet*.

Selecionada a informação que é gerada por cada *tweet*, o próximo passo é criar um *stream* que irá receber essa informação. O *stream* para conhecer o que deve procurar tem de ter um conjunto de palavras-chave para que fique à escuta e recolha a informação, a seleção de palavras-chave é crucial para selecionar *tweets* sobre o assunto que se pretende. Numa primeira fase, restringiu-se a pesquisa a palavras como *paper*, *article*, *conference*, *journal article*, *conference paper*, entre outros, esta pesquisa resultou num conjunto de *tweets* relevantes, mas, existiam *tweets* que não estavam relacionados e precisavam de ser removidos. Assim, foi necessário estudar os *tweets*, de forma a encontrar um padrão e verificou-se que há *tweets* que identificam onde o RO foi publicado e partilham o *link* associado ao mesmo. O passo seguinte consistiu em definir palavras-chave que permitam identificar locais de publicações, ver Tabela 1. Como resultado, foram recolhidos 30.000 *tweets* no período compreendido entre a março de 2020 e abril de 2020.

Arxiv	ScienceDirect	ProjectMUSE	SSRN	OpenEditionNews
Scopus	Springerlink	UCSBLibrary	EuropePMC_news	WebOfScience
Elsevier	@academia	OCLC	dblp_org	openlibrary
IEEEExplore	OUPAcademic	ncsulibraries	ingentaconnect	ElsevierConnect
Plosone	JSTOR	mendeley_com	nberpubs	NEJM
BioMedCentral,	CambridgeCore	cssci	Directory of Open Access Journals	iopscience
nresearchnews	Pubmed	ibrarycongress	JournalTOCs	TischLibrary
ScienceMagazine	semanticsscholar	cinii_jp	PhilPapers	JAMA_current
PNASNews	medlineplus	@APA	ACMDL	AIS

Tabela 1 – Palavras Chave

Para a Validação/Transformação dos dados, começou-se por identificar o idioma original dos *tweets*, de forma a não se descartar *tweets* noutros idiomas foi efetuada a tradução dos *tweets* para inglês, tendo por base as bibliotecas de Python: *langdetect* (deteta 55 idiomas) e *googletrans*. Um dos problemas em traduzir os *tweets* em tempo real é que alguns *tweets* incluem *emojis*, provocando erros no *googletrans*, assim foi necessário excluir os *emojis* dos *tweets*.

O Armazenamento de dados é efetuada na base de dados NoSQL MongoDB. Os dados são armazenados em tempo real através do uso da biblioteca pymongo do Python. Os dados foram armazenados em JavaScript Object Notation (JSON).

A fase de Preparação consiste em preparar os dados para análise e extração de sentimentos. Nesta fase são removidos duplicados, e detetar spam, ou seja, detetar se um utilizador está a comentar um RO repetidamente de forma a viciar os resultados de uma forma negativa ou até positiva (*untruthful opinions*).

Para conseguir identificar os ROs ou o *hashtag* mais mencionados no Twitter, usou-se um algoritmo que permite para cada *tweet* identificar uma menção ou um *hashtag*, sendo essas menções e *hashtags* colocados em listas de forma a identificar as suas ocorrências. Este algoritmo para além de identificar as menções e os *hashtags* identifica ainda os *links* dos ROs para com o auxílio ao *webscraping* possa extrair, por exemplo o título e autores.

Através do uso da biblioteca do Python, Natural Language ToolKit (NLTK), é possível eliminar ambiguidade nos *tweets* através da remoção da pontuação, a pontuação introduz ruído e agrega pouco valor à capacidade de análise, mas ao remover a pontuação de um texto torna-o não estruturado. O processo de *tokenization* separa um texto em unidades, como frases ou palavras, dando estrutura a um texto anteriormente não estruturado, por exemplo, a frase "palavras num *tweet*" é dividida em *tokens* [palavras, num, *tweet*]. Esta tarefa é útil para preparar o texto a ser tratado por um analisador léxico, que é a próxima etapa. Após a *tokenization* do texto, podemos alimentar um analisador léxico para remover *stop words*, pois são palavras comumente usadas num determinado idioma e não agregam nenhum valor aos dados, a biblioteca NLTK contém uma lista de palavras irrelevantes em inglês. De seguida, o *tweet* passa por um processo de *lematization*, ou seja, reduz, por exemplo as palavras "modelagem", "modelado" e "modelador" à palavra raiz, "modelo".

Ao conteúdo dos *tweets*, devidamente tratados com as técnicas de PLN, pretende-se inferir se um *tweet* apresenta uma opinião sobre um determinado RO positiva, negativa ou neutra. Desta forma, foi definido um sistema de atribuição do sentimento, que utiliza a variável *polarity* da biblioteca Textlob. A variável *polarity* compreende valores entre [-1, 1] sendo que -1 define um sentimento negativo e 1 um sentimento positivo.

Na fase de Modelação os dados são interpretados através da sua visualização gráfica, possibilitando extrair conhecimento dos dados.

A primeira visualização consiste na origem geográfica dos diversos *tweets*, ver Figura 2. Como se pode visualizar existe uma diversidade de países, os *tweets* recolhidos estão maioritariamente localizados na América do Norte, América Central e Europa. Através da análise dos idiomas é

possível verificar que o inglês é o idioma predominante nos *tweets*, mas é possível identificar que há *tweets* em 50 idiomas distintos.

Na Figura 3 pode-se analisar os ROs com mais *tweets*.

É possível visualizar com quantos idiomas diferentes cada RO é partilhado pois, esta análise ajudará a perceber o impacto que um RO pode ter na comunidade científica devido à sua partilha em várias localizações do mundo e com esta partilha universal, o RO chegará cada vez a mais pessoas e comunidades científicas impulsionando diversas trocas de opiniões, ver Figura 4.



Figura 2 – Origem geográfica dos *tweets*

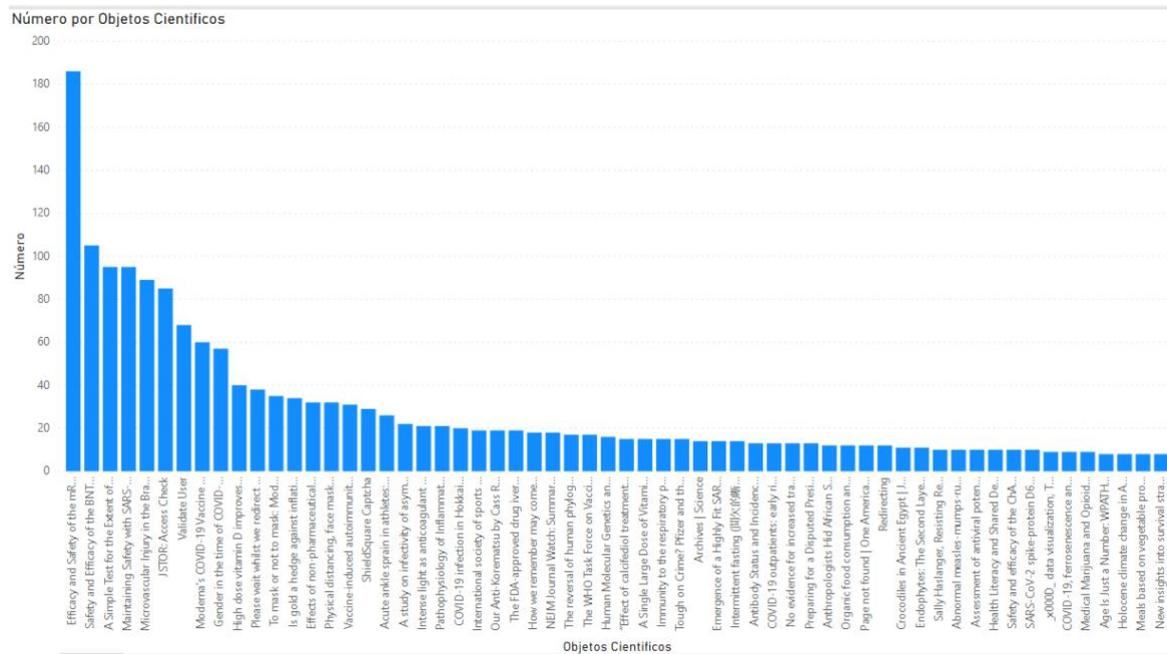


Figura 3 – ROs com mais *tweets*

Pode-se verificar na Figura 3 que o RO mais partilhado é *Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine* e na Figura 4 vê-se que esse RO foi partilhado em 15 idiomas diferentes o que

permite concluir que um RO pode ser partilhado em qualquer lugar geográfico e por isso surgem menções em diversos idiomas.

Pode-se ainda visualizar os ROs com mais menções nos *tweets* e o sentimento associado a essas menções, ver Figura 5.

Com o objetivo de tentar aumentar a precisão na extração do sentimento, alterou-se os tempos verbais dos verbos nos *tweets* para o infinito. No entanto, verificou-se que não houve qualquer tipo de alteração face à análise realizada com verbos na sua forma original.

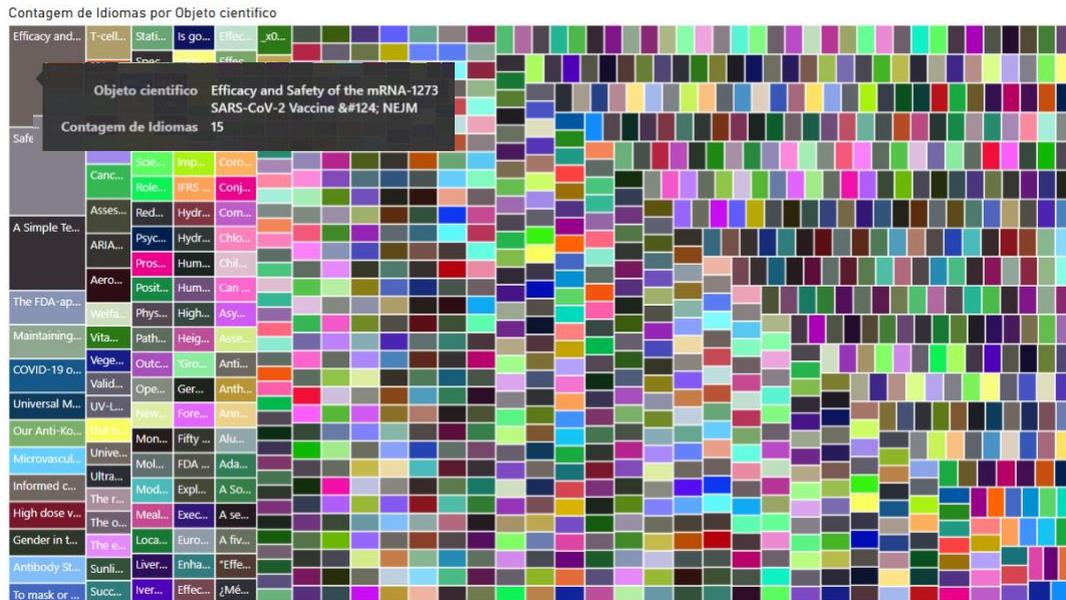


Figura 4 – ROs mais partilhados em idiomas distintos

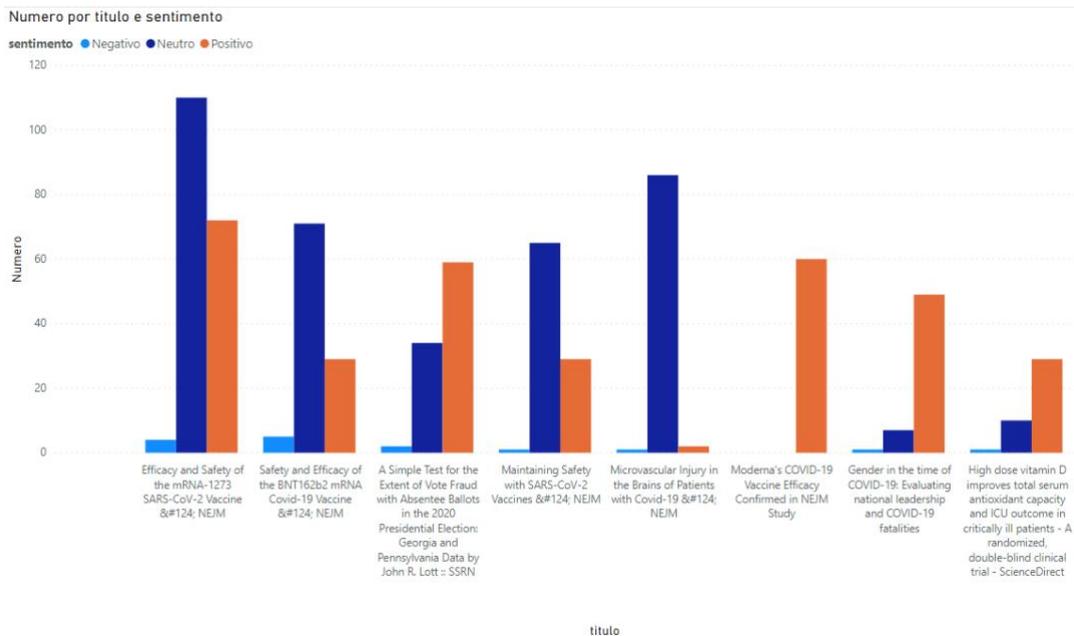


Figura 5 – ROs com mais *tweets* e sentimento associado

De acordo com os dados recolhidos e considerado o estudo realizado, apresentam-se os *hashtags* e as menções mais usadas, ver Figura 6. Percebe-se que o tema atual é referente ao estado pandémico que se está a viver e que é transversal a todo o mundo.

Como os *hashtags* que ocorrem com maior frequência estão relacionados com o tema da pandemia, os locais onde esses ROs são mencionados também confirmam esse facto, pois, por exemplo os locais @JAMA_current, @NEJM, @JAMANetwork, @sciencemagazine são locais onde habitualmente ocorre a partilha de ROs da área da saúde.

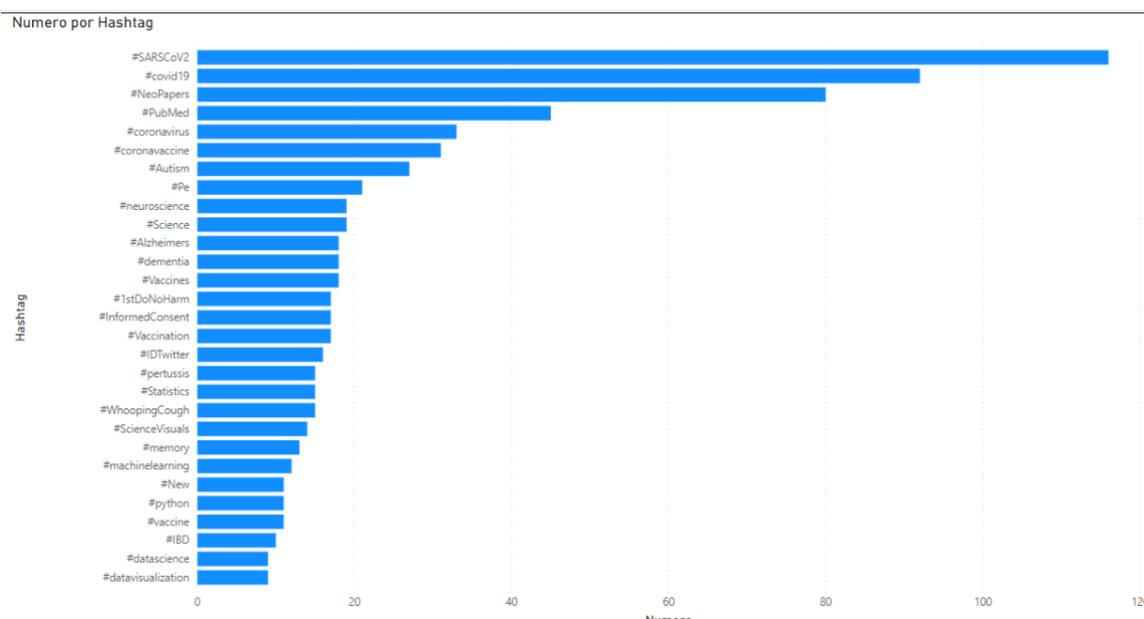


Figura 6 – Frequência de *hashtags* e menções

A fase de Avaliação tem como objetivo avaliar e interpretar os resultados obtidos na fase anterior (Modelação). Para a avaliação dos resultados foi elaborada uma apresentação a um painel de 10 investigadores associados ao projeto onde este trabalho foi desenvolvido, de forma a validar o trabalho e perceber os pontos com potencial melhoria. Os investigadores questionaram a classificação dos *tweets* em relação ao seu sentimento, pois há *tweets* que considerados positivos ou negativos, mas que ao serem lidos podem ser considerados como falsos positivos ou negativos.

Assim, foi considerada a variável *subjectivity* que permite perceber se o texto é uma opinião pessoal ou factual. A variável *subjectivity* está compreendida ente [0, 1] e quantifica a quantidade de opiniões pessoais e informações factuais contidas no texto, assim o valor 1 de *subjectivity* representa que o *tweet* contém opiniões pessoais e o valor 0 refere que o *tweet* consiste em opiniões factuais. Partindo da combinação das variáveis *polarity* e *subjectivity*, é possível obter uma relação entre as variáveis *polarity* (eixo horizontal) e *subjectivity* (eixo vertical) representada na Figura 7. Para que um *tweet* seja considerado com sentimento positivo a variável *polarity* tem de ter um valor superior a 0,8 e a variável *subjectivity* tem de ser superior a 0,5, para considerar um *tweet* com sentimento negativo é

preciso que variável *polarity* seja menor que -0,8 e a variável *subjectivity* superior a 0,5. Os valores restantes são considerados com sentimento neutro. Isto serviu para identificar os ROs que têm uma avaliação mais positiva, negativa ou neutra.

Assim, verifica-se que considerar a variável *subjectivity* na análise de sentimento provoca uma alteração nos resultados. Verifica-se que o número de comentários negativos e positivos baixa consideravelmente aumentando o número de comentários considerados neutros, ver Figura 8.

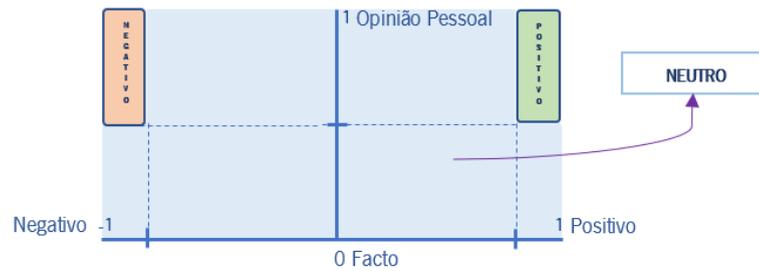


Figura 7 – Combinação das variáveis *polarity* e *subjectivity*

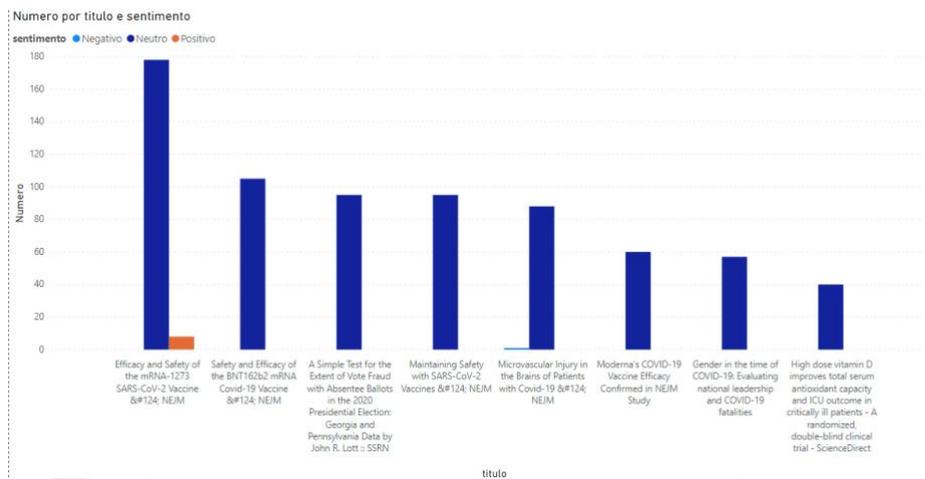


Figura 8 – ROs com mais *tweets* e sentimento associado (com *subjectivity*)

A perceção da existência de falsos positivos e negativos foi efetuada através da leitura do texto de cada *tweet* e como resultado verificou-se que os *tweets* positivos eram realmente positivos e os negativos poderiam ser considerados falsos negativos, na Tabela 2 podem ser lidos exemplos de *tweets* classificados como falsos positivos e negativos.

Positivos Falsos	<ul style="list-style-type: none"> ▪ The WHO Task Force on Vaccines for Fertility Regulation. Its formation, objectives and research activities ▪ Meaningful Engagement in the Nursing Home ▪ Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine &#124; NEJM ▪ The Ethical Defensibility of Harm Reduction and Eating Disorders
Negativos Falsos	<ul style="list-style-type: none"> ▪ Health anxiety related to problematic smartphone use and gaming disorder severity during COVID-19: Fear of missing out as a mediator – PubMed ▪ RT @AArieNugraha: Flow-Based Independent Vector Analysis for Blind Source Separation - IEEE Journals & amp; Magazine http://t.co/xWLLe4k42u

Tabela 2 – Exemplos de *tweets* classificados como falsos positivos e negativos

A consideração da variável *subjectivity* no modelo de classificação permitiu verificar que a classificação dos *tweets* em termos de polaridade foi alterada. Na Tabela 3 são apresentados exemplos de *tweets* classificados como positivos. Verifica-se que não se identificam classificações de menções negativas.

Comentários Positivos	<ul style="list-style-type: none"> ▪ RT @andresrguez: The result of Moderna 's vaccine be very very good. http://t.co/VhmUGQ2yaL Via @pmarsupia http://t.co/PaCviqq1x6 ▪ Another awesome application of MPF code: #covid diagnosis through chest CT image processing and #DeepLearning: http://t.co/xm2veVfLMf ▪ RT @robertoguerra: Looks good! Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine NEJM http://t.co/y8LrjVtBOu
--------------------------	--

Tabela 3– Exemplos de *tweets* classificados como positivos

4. CONCLUSÃO

Este trabalho apresenta uma visão de como as redes sociais podem ser usadas para partilhar opiniões sobre ROs.

Foi desenvolvido um modelo de classificação de ROs que tendo por base *tweets* partilhados no Twitter que passaram por um processo de tratamento do texto (baseado em PLN) para posteriormente ser extraída a polaridade (sentimento – positivo, neutro e negativo) nos *tweets*.

Este desenvolvimento foi elaborado tendo por base a linguagem e modelos em Python. O processo de validação dos dados foi o que exigiu mais esforço para se atingir um nível de confiança que garanta que os dados estão a ser tratados de acordo com o pretendido.

Pode-se afirmar que usar um sistema de classificação sobre menções de ROs nas redes sociais, permite determinar os ROs mais mencionados bem como a polaridade dessas menções. A polaridade das menções pode ser vista como recomendações de ROs, o que permite determinar a relevância dos ROs. Assim, menções positivas a um RO podem ser vistas como uma recomendação, mas se a polaridade for negativa, pressupõe que o RO não tem relevância para ser recomendado. No entanto, verifica-se que a maior parte das menções são neutras, isto acontece porque são menções publicadas

pelos autores ou editores a referir que um RO foi publicado, não expressando qualquer opinião sobre o mesmo.

Verifica-se ainda que não existem menções negativas a ROs, podendo concluir-se que não é normal entre a comunidade científica partilhar comentários negativos relativamente a trabalhos publicados por pares (autores), sobretudo se o autor dessas menções está identificado. Isto é relevante pois o processo de *peer-review* de ROs é um processo normalmente *blind*, isto é, o autor da revisão não é conhecido (mostrado).

É mais usual comentar positivamente o que se gosta, assim, os comentários que se consideraram positivos, são realmente positivos, pois estes transmitem uma opinião e não apenas uma apresentação factual de um RO.

Verifica-se que há muitas partilhas nas redes sociais, mas a maior parte das vezes são uma mera referência a um RO e não apresentam qualquer tipo de opinião (sentimento neutro).

Dessa forma, este estudo pode servir como base para um sistema de recomendação de ROs baseados na partilha de menções em redes sociais. Uma menção, mesmo neutra é uma referência a um RO e se for positiva, então, torna esse RO ainda mais relevante.

Este trabalho apresenta limitações, nomeadamente, o facto de só se ter utilizado uma rede social, mas como referido atrás, o acesso a dados de plataformas sociais está cada vez mais limitado.

Como trabalho futuro, sugere-se estudar quem faz menções nas redes sociais de forma a perceber quem são os *influencers* e que tipos de recomendações eles fazem.

AGRADECIMENTOS

Este trabalho foi financiado por fundos nacionais através da FCT - Fundação para a Ciência e Tecnologia no âmbito do Projeto ScopeUIDB/00319/2020.

REFERÊNCIAS

- Al-Shamri, M. Y. H. (2016). User profiling approaches for demographic recommender systems. *Knowledge-Based Systems*, 100, 175–187. <https://doi.org/10.1016/j.knosys.2016.03.006>
- Arora, A., Bansal, S., Kandpal, C., Aswani, R., & Dwivedi, Y. (2019). Measuring social media influencer index- insights from facebook, Twitter and Instagram. *Journal of Retailing and Consumer Services*, 49, 86–101. <https://doi.org/10.1016/j.jretconser.2019.03.012>
- Babin, J., & Hulland, J. (2019). Exploring online consumer curation as user-generated content: A framework and agenda for future research, with implications for brand management. *Spanish Journal of Marketing - ESIC*, 23(3), 325–338. <https://doi.org/10.1108/SJME-07-2019-0053>
- Bai, X., Wang, M., Lee, I., Yang, Z., Kong, X., & Xia, F. (2019). Scientific paper recommendation: A survey. *IEEE Access*. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2018.2890388>
- Bamakan, S. M. H., Nurgaliev, I., & Qu, Q. (2019, January 1). Opinion leader detection: A methodological review. *Expert Systems with Applications*. Elsevier Ltd. <https://doi.org/10.1016/j.eswa.2018.07.069>

- Bik, H. M., & Goldstein, M. C. (2013). An Introduction to Social Media for Scientists. *PLoS Biology*, 11(4). <https://doi.org/10.1371/journal.pbio.1001535>
- Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology*, 45. <https://doi.org/10.1002/aris.2011.1440450112>
- Byrne, E., Kearney, J., & MacEvilly, C. (2017). The Role of Influencer Marketing and Social Influencers in Public Health. *Proceedings of the Nutrition Society*, 76(OCE3), 76. <https://doi.org/10.1017/S0029665117001768>
- Deng, F. (2015). Utility-based Recommender Systems Using Implicit Utility and Genetic Algorithm (pp. 860–864). Atlantis Press. <https://doi.org/10.2991/meic-15.2015.197>
- Devchand, R., Sheehan, P., Gallivan, J. M., Tuncer, D. M., & Nicols, C. (2017). Assessment of a National Diabetes Education Program diabetes prevention toolkit: The D2d experience. *Journal of the American Association of Nurse Practitioners*, 29(9), 514–520. <https://doi.org/10.1002/2327-6924.12499>
- Dhanesh, G. S., & Duthler, G. (2019). Relationship management through social media influencers: Effects of followers' awareness of paid endorsement. *Public Relations Review*, 45(3), 101765. <https://doi.org/10.1016/j.pubrev.2019.03.002>
- Ewing, M. (2011). *71% more likely to purchase based on social media referrals*. HubSpot, United States.
- Fayyaz, Z., Ebrahimian, M., Nawara, D., Ibrahim, A., & Kashef, R. (2020). Recommendation systems: Algorithms, challenges, metrics, and business opportunities. *Applied Sciences (Switzerland)*, 10(21), 1–20. <https://doi.org/10.3390/app10217748>
- Jomsri, P., Sanguansintukul, S., & Choochaiwattana, W. (2010). A framework for tag-based research paper recommender system: An IR approach. In *24th IEEE International Conference on Advanced Information Networking and Applications Workshops, WAINA 2010* (pp. 103–108). <https://doi.org/10.1109/WAINA.2010.35>
- Kartajaya, H., Kotler, P., & Hooi, D. H. (2019). Marketing 4.0: Moving from Traditional to Digital. In *Asian Competitors* (pp. 99–123). World Scientific Publishing Co. Pte. Ltd. https://doi.org/10.1142/9789813275478_0004
- Keith Norambuena, B., Lettura, E. F., & Villegas, C. M. (2019). Sentiment analysis and opinion mining applied to scientific paper reviews. *Intelligent Data Analysis*, 23(1), 191–214. <https://doi.org/10.3233/IDA-173807>
- Ki, C. W. 'Chloe,' Cuevas, L. M., Chong, S. M., & Lim, H. (2020). Influencer marketing: Social media influencers as human brands attaching to followers and yielding positive marketing results by fulfilling needs. *Journal of Retailing and Consumer Services*, 55, 102133. <https://doi.org/10.1016/j.jretconser.2020.102133>
- Klampanos, I. A., Manning, C., Raghavan, P., & Schütze, H. (2009). Introduction to information retrieval. *Information Retrieval*, 12(5), 609–612. <https://doi.org/10.1007/s10791-009-9096-x>
- Montgomery, D. C. (2009). *Introduction to Statistical Quality Control*. 6th Edition, John Wiley & Sons, New York.
- Morris, M., & Ogan, C. (1996). The internet as mass medium. *Journal of ComputerMediated Communication.*, 1(4), 1–10. <https://doi.org/10.4324/9781315189772-25>
- Nascimento, C., Laender, A. H. F., Silva, A. S. da, & Gonçalves, M. A. (2011). A source independent framework for research paper recommendation. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries* (pp. 297–306). <https://doi.org/10.1145/1998076.1998132>
- Nielsen. (2016). *Sales Effect Study: Influencer Marketing*. TapInfluence.
- Olenki, S. (2012). *Are Brands Wielding More Influence In Social Media Than We Thought?* Forbes.
- Parra-Santander, D., & Brusilovsky, P. (2010). Improving collaborative filtering in social tagging systems for the recommendation of scientific articles. In *Proceedings - 2010 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2010* (Vol. 1, pp. 136–142). <https://doi.org/10.1109/WI-IAT.2010.261>
- Rakuten Marketing. (2019). *Influencer Marketing - Global Survey - Consumers*. Rakuten Marketing. Retrieved from <https://www.iab.com/wp-content/uploads/2019/03/Rakuten-2019-Influencer-Marketing-Report-Rakuten-Marketing.pdf>
- Ramzan, B., Bajwa, I. S., Jamil, N., & Mirza, F. (2019). An intelligent data analysis for hotel recommendation systems using machine learning. In *Scientific Programming 2019*.
- Sarwar, B. M., Karypis, G., Konstan, J., & Riedl, J. (2002). Recommender Systems for Large-scale E-Commerce : Scalable Neighborhood Formation Using Clustering. In *Proceedings of the fifth international conference on computer and information technology*.
- Schafer, J. Ben, Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering

- recommender systems. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 4321 LNCS, pp. 291–324). Springer Verlag. https://doi.org/10.1007/978-3-540-72079-9_9
- Sen, M., & Udgirakar, S. (2013). Survey Paper on Web Recommendation System. *International Journal of Science and Research*, 4, 2319–7064. Retrieved from <https://doi.org/10.21275/v4i11.nov151592>
- Silva, L. J. O. L. (2002). *Implicações cognitivas e sociais da globalização das redes e serviços telemáticos: estudo das implicações da comunicação reticular na dinâmica cognitiva e social da Comunidade Científica Portuguesa*. Centro Atlântico.
- Sugiyama, K., & Kan, M. Y. (2011). Serendipitous recommendation for scholarly papers considering relations among researchers. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries* (pp. 307–310). <https://doi.org/10.1145/1998076.1998133>
- Wang, G., He, X. R., & Ishuga, C. I. (2018). HAR-SI: A novel hybrid article recommendation approach integrating with social information in scientific social network. *Knowledge-Based Systems*, 148, 85–99. <https://doi.org/10.1016/j.knosys.2018.02.024>
- Watermeyer, R. (2010). Social network science: Pedagogy, dialogue, deliberation. *Journal of Science Communication*, 9(1), 1–9. <https://doi.org/10.22323/2.09010204>
- Wielki, J. (2020). Analysis of the role of digital influencers and their impact on the functioning of the contemporary on-line promotional system and its sustainable development. *Sustainability (Switzerland)*, 12(17), 7138. <https://doi.org/10.3390/su12177138>
- Xu, L., Jiang, C., Chen, Y., Ren, Y., & Liu, K. J. R. (2019). User participation in collaborative filtering-based recommendation systems: A game theoretic approach. *IEEE Transactions on Cybernetics*, 49(4), 1339–1352. <https://doi.org/10.1109/TCYB.2018.2800731>