



# Social4Science: Desvendando Conexões e Tendências na Difusão de Conhecimento Científico nas Redes Sociais



## Rafael Gonçalo Pereira Ribeiro

Técnico em Informática pelo Centro Federal de Educação Tecnológica de Minas Gerais (CEFET/MG) – Divinópolis, MG – Brasil.

Discente do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET) – Divinópolis, MG – Brasil.

<http://lattes.cnpq.br/6995985875605301>

## Thiago Magela Rodrigues Dias

Doutor em Modelagem Matemática e Computacional pelo Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG) – Belo Horizonte, MG – Brasil.

Professor do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET) – Divinópolis, MG – Brasil.

<http://lattes.cnpq.br/4687858846001290>

## Patrícia Mascarenhas Dias

Doutora em Modelagem Matemática e Computacional pelo Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG) – Belo Horizonte, MG – Brasil.

Professora da Universidade do Estado de Minas Gerais (UEMG) – Divinópolis, MG – Brasil.

<http://lattes.cnpq.br/6871965805554986>

## Ronaldo Ferreira Araújo

Doutor em Ciência da Informação pela Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brasil.

Professor da Universidade Federal de Alagoas (UFAL), Maceió, AL, Brasil.

<http://lattes.cnpq.br/3328212638040851>

## Emerson de Sousa Costa

Doutor em Engenharia Elétrica pela Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brasil.

Professor do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET) – Divinópolis, MG – Brasil.

<http://lattes.cnpq.br/9306302347633373>

Submetido em: 14/11/2023. Aprovado em: 08/05/2024. Publicado em: 18/07/2024.

## RESUMO

Com o aumento significativo do uso das mídias sociais, torna-se cada vez mais crucial compreender como as publicações científicas são disseminadas e debatidas nessas plataformas online. A análise dessas interações e da circulação da pesquisa científica tem sido objeto de estudo na área de altmetria, oferecendo resultados importantes sobre a percepção e compartilhamento da ciência pelo público. Este trabalho visa apresentar uma plataforma inovadora para a coleta e análise de dados sociais relacionados a publicações científicas, com ênfase na plataforma de compartilhamento de vídeos, o *YouTube*. A plataforma, denominada Social4Science, foi desenvolvida para coletar e analisar dados sociais do *YouTube*, proporcionando uma compreensão mais profunda de como as publicações científicas são divulgadas e discutidas nas mídias sociais. Ao correlacionar dados sociais com informações científicas, a Social4Science oferece uma análise abrangente de diversas

métricas, possibilitando a identificação de tendências e padrões nas discussões sobre publicações científicas. Os resultados obtidos indicam que a plataforma proposta é altamente promissora, fornecendo uma compreensão mais abrangente da interação entre a ciência e o público. Além disso, abre novas perspectivas para futuras pesquisas nessa área, destacando seu potencial para avanços significativos.

**Palavras-chave:** produção científica; mídias sociais; altmetria.

## INTRODUÇÃO

O repasse das descobertas e o processo de iniciação científica são extremamente importantes para o desenvolvimento tanto social quanto cultural. A comunicação entre o meio acadêmico e a sociedade é crucial, uma vez que todo conhecimento e pesquisa desenvolvida visa garantir um retorno à sociedade. Logo, a forma de repassar os resultados precisa estar alinhada à sociedade e às necessidades sociais para que o público perceba a sólida relação entre sociedade e ciência (Silva Neto, 2018).

Neste sentido, o *YouTube* tem oferecido uma oportunidade interessante ao movimento de divulgação científica na Internet. A plataforma é o maior site de compartilhamento de vídeos do mundo, comportando conteúdos de diversos assuntos e temáticas. O Brasil é um dos principais consumidores dessa plataforma. Graças a isso, a divulgação científica encontrou um espaço de grande reverberação, com relativa demanda de público (Fonseca; Bueno, 2021).

A disseminação da ciência por meio do *YouTube* possibilita a criação de vídeos educativos, debates, entrevistas e demonstrações práticas, promovendo a interação e o diálogo entre os pesquisadores e o público interessado.

Para os autores Reale e Martyniuk (2016), a divulgação científica por meio do *YouTube* é uma excelente ferramenta para democratizar o conhecimento científico.

Ao extrair informações de artigos científicos citados em vídeos do *YouTube*, é possível coletar uma série de dados, incluindo título do artigo científico, autores do artigo científico, nome do periódico em que o artigo foi publicado, ano de publicação do artigo, número de citações recebidas pelo artigo, dentre outros.

Com esses dados, é possível realizar diversas análises, incluindo a identificação de tendências mencionadas nos vídeos, a identificação de autores ou revistas de maior destaque mencionadas nos vídeos do *YouTube*.

Outro aspecto importante é a análise da relação entre a popularidade dos vídeos no *YouTube* e o número de citações recebidas pelos artigos científicos mencionados nesses vídeos. Essa correlação pode revelar a influência dos vídeos online na disseminação e no reconhecimento da pesquisa acadêmica. A compreensão dessa relação contribui para uma visão mais completa da interação entre a divulgação científica e o impacto da pesquisa.

A análise dos dados coletados pode revelar lacunas na comunicação científica, indicando áreas em que há uma desconexão entre a produção científica e sua divulgação online. Essas lacunas podem direcionar esforços para melhorar a comunicação e o engajamento do público, promovendo uma maior compreensão e apreciação da ciência.

A plataforma proposta denominada Social4Science visa preencher uma lacuna significativa na pesquisa científica, fornecendo uma ferramenta eficiente para explorar o vasto universo das mídias sociais e compreender como as informações científicas são disseminadas, discutidas e percebidas pelo público. Com a coleta e análise dos dados provenientes dessas plataformas, é possível obter resultados valiosos sobre as tendências, padrões e interações relacionadas à ciência.

Além disso, a plataforma abrange uma ampla gama de funcionalidades que permitem a identificação de influenciadores, a análise do impacto e da relevância das publicações científicas, a detecção de temas emergentes, entre outras análises importantes. Com base nesses dados, os pesquisadores poderão tomar decisões embasadas, desenvolver estratégias de divulgação mais eficazes e aprimorar a comunicação entre a comunidade acadêmica e a sociedade.

Portanto, Social4Science representa uma abordagem significativa para explorar o potencial das mídias sociais no contexto da pesquisa científica. Ela oferece uma visão abrangente e aprofundada das interações entre ciência e sociedade, impulsionando a divulgação científica, promovendo um diálogo mais inclusivo e estabelecendo uma sólida ponte entre a academia e o público.

O intuito da ferramenta é coletar e analisar dados das mídias sociais, como o *YouTube*, visando compreender como as publicações científicas são disseminadas e discutidas nessas plataformas. Especificamente, buscamos investigar as características dos vídeos publicados no *YouTube* que fazem referência a um DOI (*Digital Object Identifier*), identificando tendências e padrões relevantes.

Por meio da coleta de dados do *YouTube* e da aplicação de técnicas de análise, este trabalho visa obter resultados sobre como a ciência é comunicada e discutida nesse ambiente online. Ao analisar as características dos vídeos que mencionam DOIs, podemos entender melhor como a informação científica é transmitida, quais tópicos são abordados e como o público interage com esse conteúdo.

Indicadores de atenção online têm sido debatidos no contexto de estudos altmétricos, que focam na compreensão do impacto social de resultados de pesquisas na web social (Araújo, 2020). Essas análises podem ser úteis para pesquisadores, editores de periódicos e outros profissionais envolvidos na divulgação científica, pois podem ajudar a compreender melhor como a ciência é percebida e compartilhada pelo público e a identificar oportunidades para aumentar a visibilidade das publicações.

Estudos desenvolvidos com essas abordagens mais contextuais estão crescendo na literatura e sinalizam a preocupação no campo altmétrico em contribuir para o aprofundamento da análise e investigação de onde e como os artigos são utilizados por diversas comunidades que interagem com os artigos online (Araújo, 2020).

## **METODOLOGIA**

Este trabalho utilizou o portal da Altmetric por meio da plataforma Altmetric Explorer como ferramentas para realizar a busca de publicações científicas que foram citadas em vídeos publicados no *YouTube*. Essa relação entre os vídeos e os artigos científicos é estabelecida quando um vídeo menciona um artigo utilizando o DOI (*Digital Object Identifier*), que geralmente é incluído na descrição do vídeo.

A utilização do DOI como identificador único permite vincular de forma precisa um vídeo específico a um artigo científico correspondente. Ao realizar a busca no portal da Altmetric por vídeos no *YouTube* que mencionam DOIs, foi possível identificar e coletar os dados para as análises e o estudo das interações entre a mídia social e a pesquisa científica.

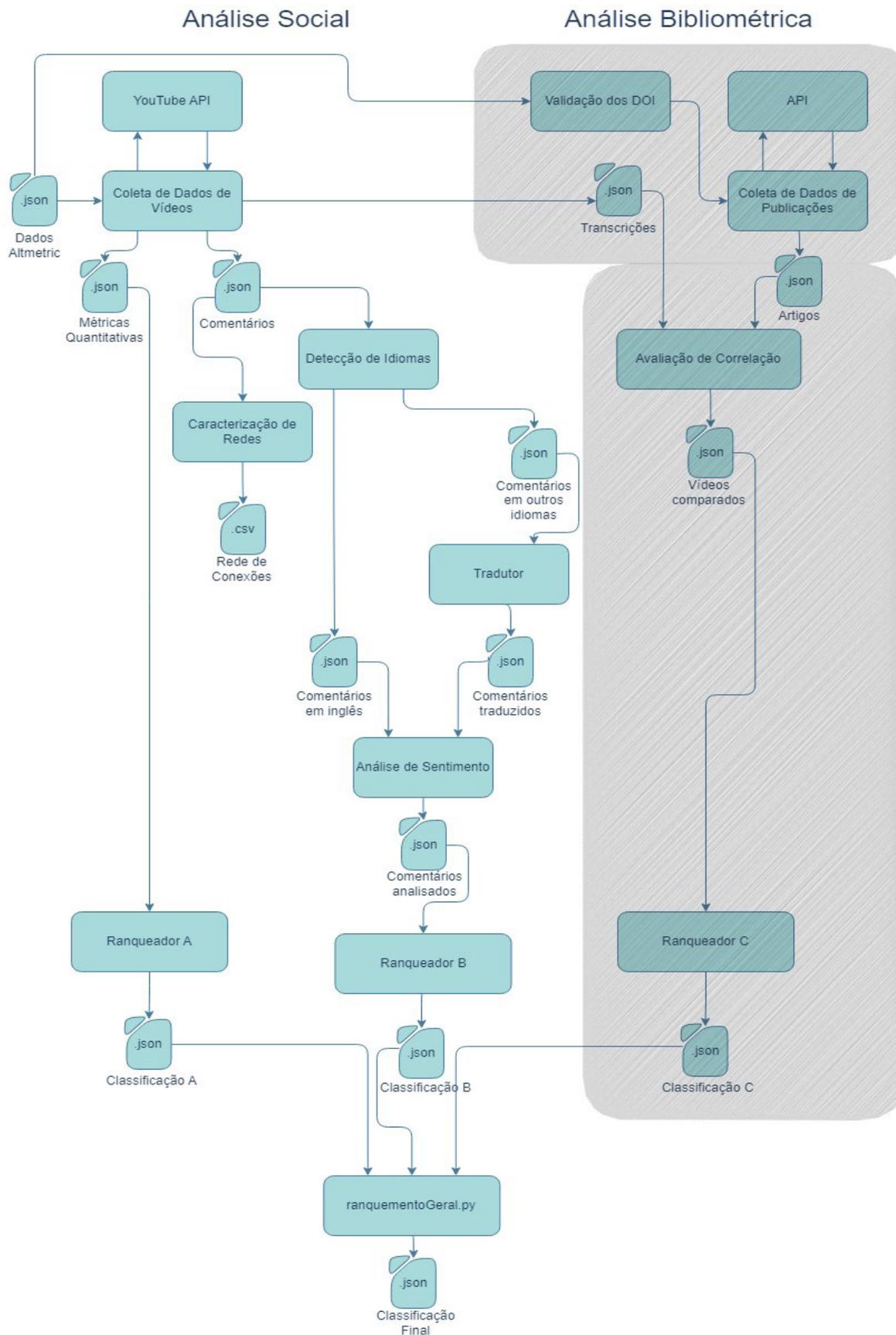
Essa abordagem de buscar referências de artigos científicos em vídeos do *YouTube* por meio do DOI é uma forma eficaz de identificar a presença e o alcance da ciência nessa plataforma. A plataforma Altmetric Explorer oferece recursos que facilitam a coleta e o tratamento dos dados, possibilitando a realização de análises detalhadas sobre as características dos vídeos e as citações dos artigos científicos.

A partir desta relação extraída da Altmetric, contendo um arquivo com Identificador do vídeo e o DOI de uma publicação, todo o processo de extração dos dados é iniciado. A plataforma Social4Science recebe esta relação como entrada e dá início a todo o processo de coleta e análise dos dados, dividido em dois segmentos:

1. Análise Social: coleta de dados dos vídeos do *YouTube*.
2. Análise Bibliométrica: coleta de dados dos artigos científicos.

A arquitetura da Plataforma Proposta pode ser observada na **FIGURA 1**.

**FIGURA 1 – Arquitetura geral da Plataforma Social4Science**



Fonte: Autores (2023).

O processo de coleta e tratamento dos dados realizado pela plataforma se inicia com a entrada de um arquivo fornecido pela Altmetric, contendo os identificadores dos vídeos e os DOIs dos artigos científicos. Esses DOIs são utilizados na etapa de “Análise Bibliométrica”, enquanto os identificadores dos vídeos são utilizados na etapa de “Análise Social”.

A etapa de “Análise Bibliométrica” consiste na utilização dos DOIs dos artigos científicos para obter informações relevantes sobre as publicações, como título, autores, periódico de publicação, ano de publicação e número de citações recebidas. Esses dados bibliométricos são essenciais para compreender a relevância e o impacto dos artigos científicos mencionados nos vídeos do *YouTube*.

Por sua vez, a etapa de “Análise Social” utiliza os identificadores dos vídeos para explorar os aspectos sociais e de interação relacionados aos vídeos que citam os artigos científicos. Essa análise social pode incluir a identificação de tendências, a análise da popularidade dos vídeos, a avaliação das interações dos usuários, como curtidas, compartilhamentos e comentários, e a identificação de influenciadores ou canais relevantes na divulgação científica.

Ao separar as etapas de análise bibliométrica e análise social, a plataforma permite uma abordagem abrangente e aprofundada na compreensão do impacto e da disseminação da ciência nas mídias sociais, especialmente no *YouTube*. Combinando informações bibliométricas e dados sociais, é possível obter informações importantes sobre como as publicações científicas são recebidas, discutidas e compartilhadas nessa plataforma, contribuindo para o avanço da divulgação científica e para a compreensão das interações entre a ciência e a sociedade.

Na Análise Social (1) os dados dos vídeos são coletados via uma *Application Programming Interface* (API) pública do *YouTube*, quando são gerados alguns extratos de dados. Eles são utilizados para o cálculo de diversas métricas e podem ser exportados para outras ferramentas de análises e visualizações, permitindo outras formas de aprofundamento.

Como exemplo destes extratos, destacam-se os conjuntos contendo dados quantitativos dos vídeos, como, por exemplo, quantidade de visualizações, quantidade de comentários e curtidas de cada vídeo, bem como, extratos contendo dados dos canais em que os vídeos foram publicados, das redes de interações identificadas a partir dos comentários de cada vídeo, extratos das descrições dos vídeos, das transcrições de cada áudio e, por fim, um conjunto de dados padronizados no idioma inglês de todos os comentários extraídos.

Já na Análise Bibliométrica (2), o conjunto dos DOI é verificado via API, a fim de validá-los. Caso seja um DOI válido, seus dados são enviados para a API da OpenAlex, recuperando, desta forma, informações sobre o artigo em questão, como, por exemplo, seu título, autores, ano de publicação, resumo, palavras-chave, periódico de publicação, dentre outros. Visando complementar os dados, uma nova requisição do mesmo DOI é enviada para a API da OpenCitations, recuperando as citações do artigo.

Todo este conjunto de dados são armazenados em extratos de dados que também são objeto de análises por diversas métricas implícitas na própria plataforma e são disponibilizados em formatos que possam ser importados por outras ferramentas de análise e visualização.

Os dados quantitativos desempenham um papel fundamental na plataforma, permitindo a realização de diferentes tipos de ranqueamento e a análise de correlações entre as Análises Sociais e as Análises Bibliométricas. Essas medidas quantitativas fornecem insumos valiosos sobre a popularidade, engajamento e alcance dos vídeos e das publicações científicas mencionadas neles.

Por outro lado, os conjuntos de dados que contêm informações textuais dos vídeos, como títulos, comentários, descrição e transcrição, são correlacionados com os dados textuais das publicações científicas, como títulos, resumos e palavras-chave. Nesse contexto, são adotadas medidas de correlação, como a distância de Levenshtein ou o cálculo da similaridade do cosseno, para explorar as relações entre os textos.

A distância de Levenshtein é uma métrica que calcula a diferença entre duas sequências de caracteres, como títulos de vídeos e títulos de publicações científicas. Essa medida permite avaliar a similaridade ou dissimilaridade entre os textos, fornecendo informações sobre a proximidade temática entre os vídeos e as publicações.

A similaridade do cosseno, por sua vez, é uma medida que quantifica a similaridade entre dois vetores de palavras, como os termos presentes nos comentários dos vídeos e as palavras-chave das publicações científicas, permitindo identificar associações semânticas e relações entre os textos.

Ao utilizar essas medidas de correlação, a plataforma Social4Science pode revelar conexões entre os conteúdos dos vídeos e das publicações científicas, identificar padrões temáticos e explorar como as informações são transmitidas e discutidas nas mídias sociais.

Dessa forma, a combinação de dados quantitativos e textuais proporciona uma análise abrangente e enriquecedora, permitindo compreender tanto os aspectos quantitativos quanto os textuais envolvidos na disseminação e na discussão das publicações científicas no *YouTube*.

Como um estudo de caso inicial, foram coletados da Plataforma Altmetric em março de 2022 um conjunto contendo 65.534 DOI's que possuíam na época citações de vídeos do *YouTube*. Deste conjunto de DOI's foram verificadas diversas características das publicações científicas. Considerando a análise do tipo de publicação, verificou-se que a maioria era de artigos (94,9%), seguido em menor quantidade por livros (3%) e capítulos de livros (1%). Ressalta-se ainda um total de 45 Conjuntos de Dados que também eram referenciados.

## **RESULTADOS**

Na Análise Social, todos os dados dos vídeos são analisados, como quantidade de curtidas, quantidade de visualizações e quantidade de comentários. Por outro lado, na análise bibliométrica, todos os dados quantitativos dos artigos são analisados, como a validação do DOI, recuperação da quantidade de citações recebidas por outros artigos e a quantidade de vídeos que citam o referido artigo. A partir desses dados, é possível identificar tendências e

padrões nas discussões sobre as publicações científicas nas mídias sociais. Por exemplo, pode-se identificar quais publicações são mais populares nas mídias sociais, quais assuntos geram mais discussões e quais são os principais influenciadores.

Considerando ainda as análises bibliométricas, foi possível verificar a representatividade dos principais periódicos em que os artigos haviam sido publicados (**FIGURA 2**).

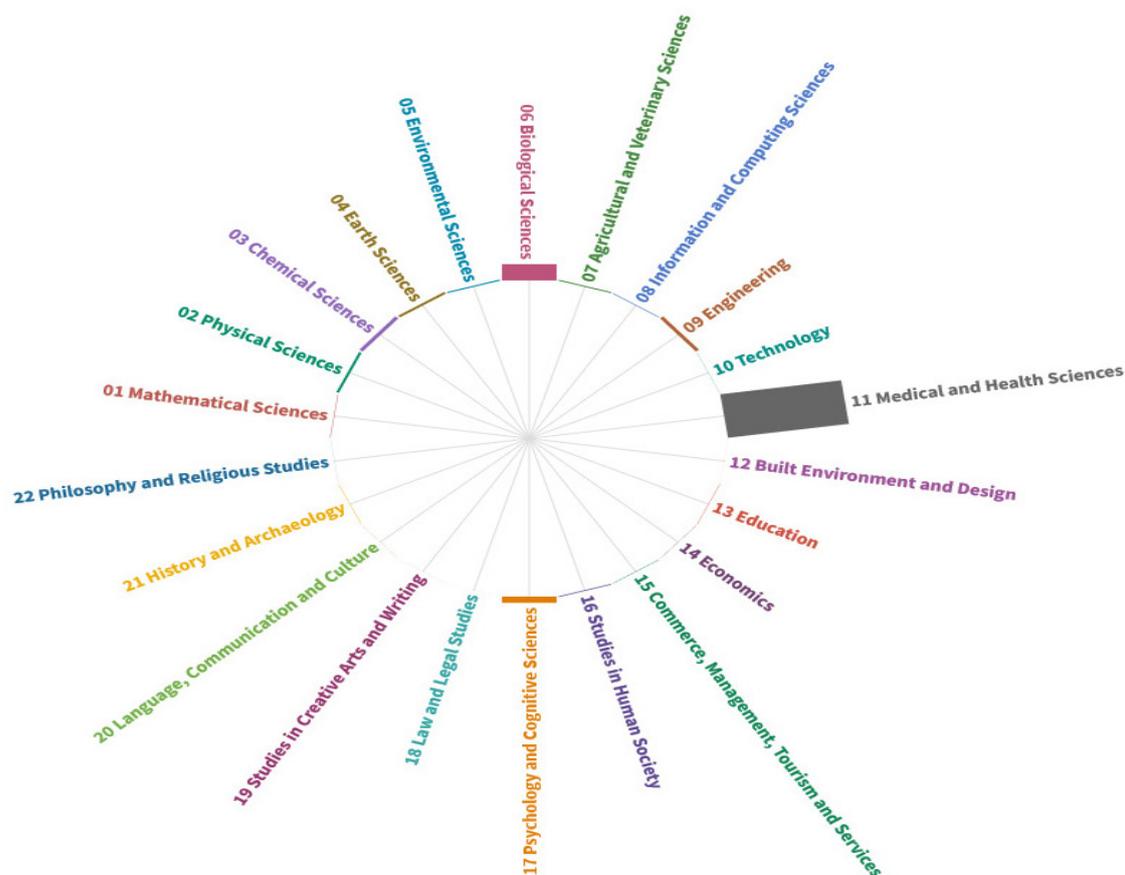
**FIGURA 2** – Representatividade dos periódicos dos artigos referenciados



Fonte: Autores (2023).

É possível observar a representatividade de alguns periódicos de prestígio como Nature, American Journal of Clinical Nutrition, Plos One, Nutrients, Journal of Strength & Conditioning Research e Science. Outra observação está relacionada às áreas do conhecimento dos artigos analisados que tendem a ter uma maior utilização do *YouTube* como ferramenta de divulgação de artigos (**FIGURA 3**).

**FIGURA 3** – Principais áreas dos artigos referenciados.



Fonte: Autores (2023).

Utilizando a classificação de áreas do conhecimento dos artigos, conforme dados coletados, percebe-se uma grande concentração em duas áreas principais, Ciências Médicas e da Saúde (69%) e Ciências Biológicas (11,5%), sendo apenas estas duas áreas detentoras de aproximadamente 80% de todo o conjunto.

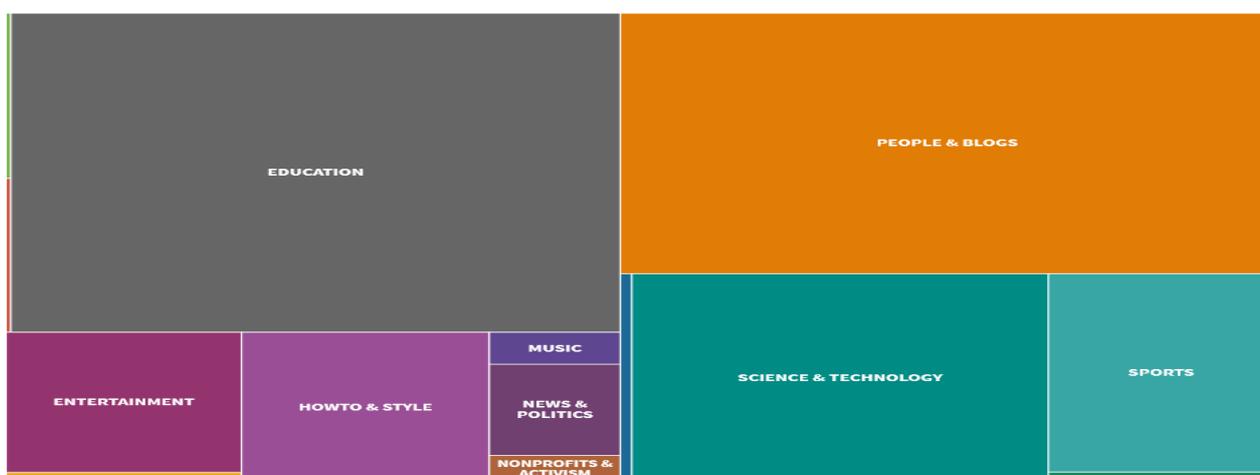
Diversas outras análises bibliométricas também foram realizadas, como, por exemplo, a validação dos DOI que são citados, objetivando verificar a real existência da publicação e, após esta etapa, a coleta e análise dos dados contidos nos títulos, resumos, palavras-chave, bem como, informações do número de citações desses artigos por outras publicações. E ainda, dados dos periódicos em que foram publicados, como fator de impacto e Qualis. Estes dados são automaticamente coletados por API públicas de diversas fontes e de forma automática.

Já as análises sociais levam em conta informações dos vídeos publicados no *YouTube* que fazem referência a algum DOI. Logo, as análises sociais se concentram em informações obtidas diretamente do *YouTube*, utilizando-se para isso a sua API pública de requisição dos dados. Este processo de extração dos dados do *YouTube* também ocorre de forma automática a partir da listagem inicial informada. A partir dos dados de entrada, os identificadores dos vídeos são extraídos e as requisições são realizadas na API do *YouTube*.

Foi possível observar nos dados coletados como os vídeos estão classificados em categorias. Tais categorias são referentes aos canais em que estes vídeos são publicados (**FIGURA 4**).

Percebe-se que os canais que possuem vídeos que citam publicações com DOI estão em sua maioria categorizados como Educação, Pessoas e Blogs, e ainda Ciência e Tecnologia. Para melhor compreensão desta categorização e seu impacto, é importante analisar qual a representatividade destes canais, uma vez que a quantidade de inscritos no canal, a quantidade de vídeos publicados ou a data de registro do canal no *YouTube* podem agregar diversas novas informações.

**FIGURA 4** – Categorias dos canais dos vídeos.



Fonte: Autores (2023).

Considerando ainda os dados da Análise Social, são caracterizados pela plataforma desenvolvida, redes de comentários. Os comentários de cada um dos vídeos são analisados e as ligações entre os canais são identificadas. Logo, a partir de um conjunto de comentários, é possível identificar com análise de redes como os canais interagem, tendo em vista os comentários que realizam ou recebem.

Além das redes caracterizadas, diversas outras análises quantitativas são realizadas, como, por exemplo, a quantidade de visualizações que um determinado vídeo recebe, a quantidade de comentários, a quantidade de curtidas, a duração e o idioma dos vídeos, dentre outros dados.

Considerando ainda a Análise Social, o conteúdo dos vídeos também é objeto de análises, como, por exemplo, o título dos vídeos, a descrição dos vídeos, bem como a transcrição do áudio. Tais elementos são importantes, pois viabilizam diversas outras análises que visam correlacionar os conteúdos dos vídeos com dados do conteúdo dos artigos, que também são coletados, como o título, resumo e palavras-chave.

## CONCLUSÕES

A plataforma Social4Science proposta neste trabalho permite a coleta e análise de dados científicos em mídias sociais, fornecendo resultados valiosos sobre a divulgação e disseminação do conteúdo científico. Através da análise desses dados, é possível identificar tendências, padrões e lacunas de conhecimento nas discussões sobre publicações científicas nas mídias sociais. Essa ferramenta oferece informações importantes para pesquisadores e profissionais da área científica, permitindo ajustes nas estratégias de comunicação e promoção do conhecimento científico, estabelecendo uma conexão mais efetiva com o público.

Através dos dados coletados pela plataforma proposta, é possível estabelecer correlações significativas entre diferentes variáveis. Essas correlações fornecem uma compreensão mais aprofundada sobre como a popularidade de um vídeo no *YouTube* está relacionada às características do artigo científico mencionado, considerando, fatores como área de pesquisa, tipo de publicação e país de origem. Essas análises permitem uma análise abrangente do impacto e da repercussão das publicações científicas nas mídias sociais, contribuindo para a compreensão do processo de disseminação e alcance do conhecimento científico.

Todo o ferramental desenvolvido com o código-fonte de todos os módulos do *framework* será disponibilizado em um repositório do *GitHub* para toda comunidade de interesse.

## **REFERÊNCIAS**

ARAUJO, R. F. Communities of attention networks: introducing qualitative and conversational perspectives for altmetrics. **Scientometrics**, [s. l.], v. 124, p. 1793-1809, 2020. DOI: 10.1007/s11192-020-03566-7.

FONSECA, A. A.; BUENO, L. M. Breve panorama da divulgação científica brasileira no YouTube e nos podcasts. **Cadernos de Comunicação**, Santa Maria, v. 25, n. 2, p. 2-19, maio/ago. 2021. Disponível em: [https://scholar.archive.org/work/kci5gaqbyvbxhonfsiyjvf3fai/access/wayback/https://periodicos.ufsm.br/ccomunicacao/article/download/63121/pdf\\_1](https://scholar.archive.org/work/kci5gaqbyvbxhonfsiyjvf3fai/access/wayback/https://periodicos.ufsm.br/ccomunicacao/article/download/63121/pdf_1). Acesso em: 12 mar. 2023.

SILVA NETO, J. R. Alcance da divulgação científica por meio do YouTube: estudo de caso no canal Meteoro Brasil. **Múltiplos Olhares em Ciência da Informação**, Belo Horizonte, v. 8, n. 2, 2018. Disponível em: <https://periodicos.ufmg.br/index.php/moci/article/download/16885/13644>. Acesso em: 12 mar. 2023.

REALE, M. V.; MARTYNIUK, V. L. Divulgação Científica no Youtube: a construção de sentido de pesquisadores nerds comunicando ciência. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 39., 2016, São Paulo. **Anais [...]**. São Paulo: Intercom – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação. Disponível em: <https://www.portalintercom.org.br/anais/nacional2016/resumos/R11-0897-1.pdf>. Acesso em: 12 mar. 2023.