

Análise Comparativa de Canais do YouTube utilizando Redes Complexas

Phelipe Rodovalho dos Santos¹, Fabíola S. F. Pereira¹

¹Faculdade de Computação – Universidade Federal de Uberlândia (UFU)
Uberlândia – MG – Brazil

{pheliperodvalho, fabiola.pereira}@ufu.br

Abstract. *How to compare two YouTube channels? Determining similar users on a social network is an increasingly important task, as it allows the personalization of the offer and consumption of content, with application in the areas of marketing and business intelligence, for example. Most existing methods in the literature to identify similarity between users are based only on published and consumed content. This work proposes a method that uses complex networks to compare YouTube channels, implemented in an automated tool. The results obtained show that it is possible to identify similarity based on the capacity to influence and the interconnection between channels.*

Resumo. *Como comparar dois canais do YouTube? Determinar usuários semelhantes numa rede social é uma tarefa cada vez mais importante, pois permite a personalização da oferta e consumo de conteúdos, com aplicação nas áreas de marketing e inteligência de negócio, por exemplo. A maioria dos métodos existentes na literatura para identificar similaridade entre usuários baseia-se apenas no conteúdo publicado e consumido. Este trabalho propõe um método que utiliza redes complexas para comparar canais do YouTube, implementado em uma ferramenta automatizada. Os resultados obtidos mostram que é possível identificar similaridade com base na capacidade de influência e na interligação entre canais.*

1. Introdução

As redes sociais, com destaque para o YouTube, vêm desempenhando um papel fundamental na comunicação global [Teixeira 2023]. Com o grande volume de dados gerados e com o crescimento de tais redes, tornou-se essencial entender as interações entre usuários. No contexto no YouTube, identificar similaridades entre diferentes canais, levando em consideração seus respectivos públicos, pode fornecer informações preciosas para melhorar a experiência dos usuários e otimizar estratégias de conteúdo e marketing [Chitra and Musco 2020].

De fato, a colaboração entre canais e influenciadores é um tema de interesse crescente [Wang et al. 2020]. Destaca-se a grande quantidade de algoritmos de recomendação de conteúdo que têm sido propostos com o objetivo de personalização da experiência do usuário [Covington et al. 2016]. Entretanto, nota-se uma lacuna na literatura quando a necessidade é voltada para a comparação direta entre dois usuários (canais). Uma comparação direta é uma abordagem interessante para

obtenção de *insights* para escolhas de campanhas de marketing e alcance de influenciadores [Balakrishnan and Griffiths 2017], por exemplo. Em especial, existe uma falta de ferramentas automatizadas para a análise comparativa de canais do YouTube [Mislove et al. 2007].

Esse trabalho busca preencher tal lacuna, oferecendo uma solução que permite, de maneira automatizada, desde a extração de dados públicos de canais do YouTube até a comparação quantitativa e qualitativa entre dois canais.

Uma das formas de modelagem das relações entre seguidores e seguidos do YouTube é por meio de redes complexas. As redes complexas são um tipo de grafo que apresentam propriedades topográficas bastante particulares, características, por exemplo, de redes sociais, não encontradas em grafos mais simples [Barabasi and Posfai 2016]. A proposta é utilizar análise de redes complexas para modelagem das relações entre canais do YouTube.

2. Trabalhos Relacionados

Em [Lopes 2022] é proposta uma ferramenta de visualização dinâmica de vídeos do YouTube para facilitar o processo de busca de informações em vídeos. Trata-se de um trabalho semelhante a este no sentido de propor uma ferramenta para manipulação do conteúdo da rede social YouTube. Porém, o foco está na recuperação de informação de forma intuitiva e não na comparação entre canais.

Diversos trabalhos propõem análises sobre o conteúdo do YouTube. A maioria deles com o foco no conteúdo textual e audiovisual para solução de tarefas de análise de sentimentos ou de polarização política. Como exemplo, em [Vasconcelos et al. 2020] é realizada uma análise da distribuição temática presente em vídeos do YouTube compartilhados via WhatsApp, com o objetivo de detectar posicionamentos em relação à vacina do COVID-19.

Em trabalhos que raciocinam sobre a teoria das redes complexas destaca-se a proposta de [Tantardini et al. 2019]. Métodos de comparação entre redes são apresentados e têm o foco em estratégias de alinhamento entre nós de duas redes. São estratégias ortogonais à utilizada neste trabalho, pois não utilizam representações vetoriais. Nenhuma aplicação dos métodos em redes reais é apresentada.

Por fim, [Li and Yang 2009] é o trabalho que mais se aproxima a este. É proposta a comparação entre redes reais baseada em métricas de redes complexas com representação vetorial. As métricas utilizadas são: número de nós, número de arestas, grau médio, coeficiente de clusterização e modularidade. O trabalho, entretanto, não aplica o método em redes sociais e não propõe uma ferramenta para comparação entre redes.

3. Metodologia

A proposta de comparação entre canais do YouTube foi desenvolvida em uma ferramenta composta por 4 módulos principais, apresentados na Figura 1 e descritos a seguir.

Coleta de dados. O processo de coleta de dados é conduzido por meio da API fornecida pelo YouTube para desenvolvedores¹. O algoritmo de coleta, desenvolvido em

¹<https://developers.google.com/youtube/v3>

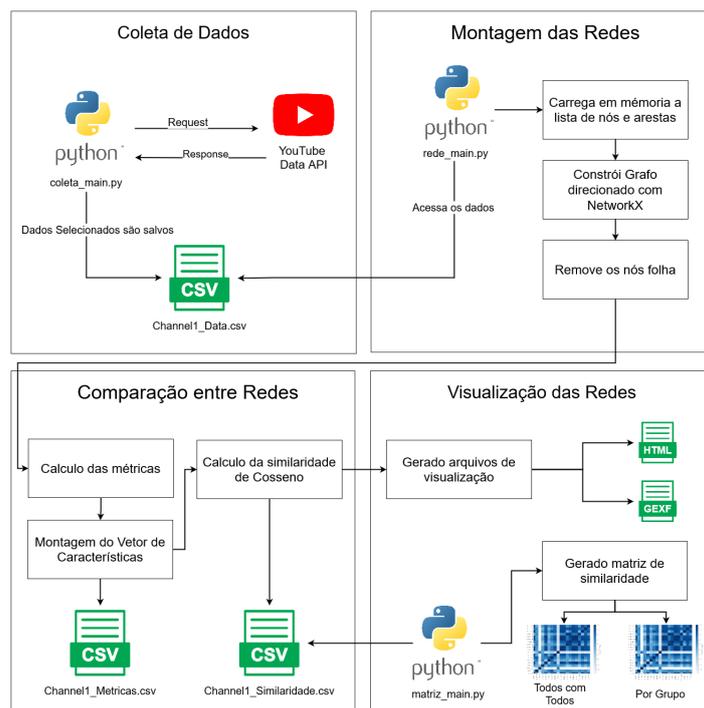


Figura 1. Módulos da ferramenta automatizada proposta para comparação entre canais do YouTube.

Python 3.11, foi projetado para coletar dados de canais de forma escalável e estruturada. O usuário fornece as informações necessárias por meio de um arquivo de configuração no formato `.ini`, que inclui uma lista de canais a serem coletados, a profundidade da busca e a chave de desenvolvedor do YouTube. O algoritmo inicia a coleta de dados a partir de um canal de referência, que é definido como “nível 0”. Para cada canal de referência, a API é usada para obter informações sobre os canais aos quais o canal de referência está inscrito. Com a conclusão da coleta de dados, os registros obtidos são exportados para arquivos no formato `.csv` para posterior análise e construção de redes complexas. As informações armazenadas no arquivo `.csv` incluem: Canais Seguidos e Seguidores, Identificadores dos Canais (ID), Descrição do Canal Seguido, Data de Criação do Canal Seguido e Profundidade do Relacionamento.

Construção da rede. Cada canal coletado nos dados é representado como um nó na rede. Se um canal segue outro canal, uma aresta direcionada é criada entre os nós correspondentes na rede, indicando uma relação de “seguir”.

Comparação entre redes. O processo de comparação é dividido em etapas distintas, incluindo cálculo de métricas, montagem de vetores de características e cálculo de similaridade. Para comparar as redes de canais do YouTube, diversas métricas foram calculadas. As métricas incluem: Grau de Centralidade, Centralidade de Proximidade, Centralidade de Intermediação, Modularidade, Densidade, Coeficiente de Aglomeração Médio, Número de Nós, Número de Arestas e PageRank. A escolha dessas métricas baseou-se na capacidade de identificar a importância dos canais na rede, sua influência sobre outros canais e o volume de conexões entre os canais. Para realizar a comparação entre as redes, os dados de cada rede foram transformados em um conjunto de vetores de características.

Cada nó de uma rede é representado por um vetor de características. Cada posição do vetor contém os valores das métricas mencionadas acima, totalizando 9 posições por nó. Dessa forma, cada rede é representada por um conjunto de características numéricas. Para analisar a similaridade entre as redes, empregou-se o cálculo da similaridade de cosseno com base nos vetores de características obtidos para cada canal. A similaridade de cosseno é uma métrica que avalia o grau de semelhança entre dois vetores, considerando o cosseno do ângulo formado por esses vetores. Quanto mais próximo o valor resultante estiver de 1, maior será a similaridade entre os canais comparados.

Visualização da rede. As redes de canais foram construídas graficamente, representando os canais como nós e as relações entre seguidores/seguidos como arestas. Isso permite uma representação visual das interações na plataforma. Antes da visualização, é feita a poda de nós-folha da rede, a fim de simplificar a visualização e focar nos canais mais centrais e relevantes. Os nós da rede foram enriquecidos com atributos como o título dos canais. Além disso, as arestas contêm informações como os nomes dos canais seguidores, profundidade da relação e data de publicação. Para viabilizar a interpretação e o compartilhamento dos resultados, as redes foram exportadas em formatos `.html` e `.gexf`. A visualização interativa em `.html` foi obtida por meio do PyVis², uma ferramenta que permite interações dinâmicas, como zoom e filtragem, proporcionando uma experiência intuitiva e imersiva na exploração das redes. Por meio dessa visualização, os usuários podem investigar de forma direta os relacionamentos entre os canais, compreendendo a estrutura e as conexões de maneira mais detalhada e prática. Além da visualização em `.html`, os arquivos no formato `.gexf` foram disponibilizados para uso no Gephi³, um software amplamente reconhecido e utilizado para análise de redes complexas [Bastian et al. 2009]. O Gephi oferece uma gama diversificada de ferramentas analíticas e de visualização, permitindo uma exploração mais aprofundada das redes. Com recursos como análise de comunidade, centralidade e distribuição de grau, os usuários podem extrair insights valiosos sobre a organização e a dinâmica das redes de canais do YouTube.

4. Resultados Experimentais

Com base nos critérios de (i) foco em personalidades públicas e transparência de interesses, (ii) abrangência e diversidade e (iii) divergência temática, foram selecionados 9 canais de cantores, descritos na Tabela 1.

4.1. Visualização das Redes

As visualizações proporcionam uma compreensão inicial da estrutura das redes, destacando aspectos como a conectividade entre os canais e a presença de comunidades ou grupos. A Figura 2, ilustra a visualização obtida da rede de Pablllo Vittar utilizando a ferramenta Gephi, onde as cores distinguem comunidades definidas pela modularidade. O tamanho das *labels* reflete o grau de entrada dos nós. *Labels* maiores denotam uma quantidade significativa de conexões entrantes, indicando canais que são frequentemente seguidos ou citados por outros na rede. Essa representação oferece uma perspectiva notável sobre a complexidade e a amplitude das conexões e inspirações da artista. É possível observar como a mesma se posiciona no centro de um ecossistema diversificado que engloba

²<https://pyvis.readthedocs.io/en/latest/>

³<https://gephi.org/>

Canal	Grupo	Qnt Inscritos ($\times 10^6$)	Qnt Vídeos
Melim	Cantores	6,21	161
Gusttavo Lima	Cantores	20,4	452
Leonardo	Cantores	3,09	376
Marquinhos Gomes	Cantores	0,664	51
Zé Ramalho	Cantores	2,91	184
Gloria Groove	Cantores	3,63	158
Pablo Vittar	Cantores	7,75	225
Adhemar de Campos	Cantores	0,278	395
Leandro Borges	Cantores	4,02	162

Tabela 1. Descrição dos 9 canais escolhidos.

canais de variados nichos e gêneros musicais. A rede evidencia conexões com canais que variam de conteúdo humorístico, como “Porta dos Fundos”, a canais de grande visibilidade como “Vevo”, refletindo o alcance e a relevância de Pablo Vittar no cenário digital. A inclusão de canais de outros influenciadores e artistas, como “whinderssonnunes” e “PewDiePie” sublinha o aspecto colaborativo e multifacetado da carreira de Pablo.

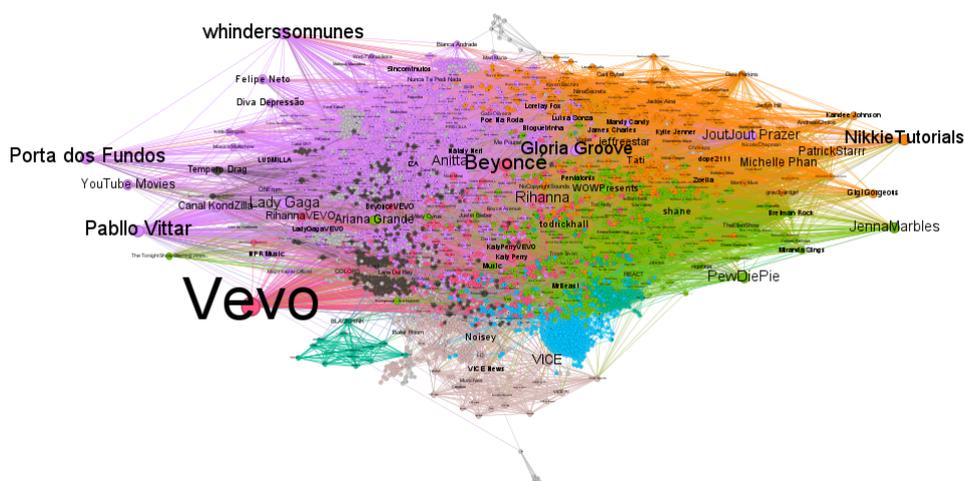


Figura 2. Uma representação gráfica da rede de influenciadores da artista Pablo Vittar, composta por diversos indivíduos de diferentes nichos e níveis de influência. As cores representam as comunidades identificadas pela métrica de modularidade. O tamanho do *label* de um nó é definido pelo grau de entrada.

4.2. Comparando duas redes

A comparação entre as redes dos diferentes canais permite identificar padrões e diferenças na estrutura das redes em relação ao tema, público e outras características.

Compreendendo a Matriz de Cantores. A análise da matriz de similaridade dos cantores, Figura 3, destaca-se a notável similaridade entre os cantores sertanejos “Gusttavo Lima” e “Leonardo”, indicando uma forte correlação nas suas redes de conexões. Da mesma forma, os artistas gospel “Marquinhos Gomes” e “Adhemar de Campos” compartilham alta similaridade, refletindo uma possível congruência de temas espirituais,

inspirações e colaborações frequentes. Por outro lado, “Pablo Vittar”, notável pela versatilidade de músicas com diferentes gêneros musicais, exibe alta similaridade com variados artistas, incluindo “Gloria Groove” no pop e “Marquinhos Gomes” no gospel, sugerindo uma abordagem artística que transita fluidamente entre gêneros diversos. Essa fusão de influências é particularmente perceptível em algumas músicas que claramente trazem inspiração gospel. Também se destaca a relação entre “Melim” e “Gloria Groove”, que apresentam similaridade relativamente baixa, sinalizando diferenças significativas nos círculos de influência e estilos musicais. Enquanto “Melim” tende para um estilo mais acústico e melódico, “Gloria Groove” se enraíza no pop moderno e em fortes referências da Arte Drag, evidenciando a diversidade de conexões que cada artista cultiva.



Figura 3. Matriz de Similaridade entre Canais do YouTube - Cantores

5. Conclusão

Este trabalho desenvolveu e aplicou uma ferramenta para a análise comparativa entre canais do YouTube, explorando a ciência das redes para analisar a influência e a interconexão entre esses canais. Através da coleta de dados via API do YouTube, organização em dados tabulares, montagem das redes, cálculo de métricas, visualização e cálculo da similaridade entre canais, foi possível avaliar e comparar a estrutura e as dinâmicas de interação entre diferentes canais da plataforma.

Trata-se de uma versão inicial da ferramenta. Estão previstas duas linhas de evolução. A primeira linha será a utilização do banco de dados (BD) Redis para o armazenamento das redes coletadas. O armazenamento será tanto dos documentos .csv, quanto dos vetores de características (RedisAI), facilitando o cálculo das similaridades entre canais. A segunda linha será o enriquecimento do vetor de características com base no conteúdo textual dos comentários dos vídeos de um canal.

Referências

Balakrishnan, J. and Griffiths, M. D. (2017). Social media addiction: What is the role of content in youtube? *Journal of behavioral addictions*, 6(3):364–377.

- Barabasi, A.-L. and Posfai, M. (2016). *Network science*. Cambridge University Press, Cambridge.
- Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks.
- Chitra, U. and Musco, C. (2020). Analyzing the impact of filter bubbles on social network polarization. In *Proceedings of the 13th International Conference on Web Search and Data Mining, WSDM '20*, page 115–123, New York, NY, USA. Association for Computing Machinery.
- Covington, P., Adams, J., and Sargin, E. (2016). Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems, RecSys '16*, page 191–198. Association for Computing Machinery.
- Li, W. and Yang, J.-Y. (2009). Comparing networks from a data analysis perspective. In Zhou, J., editor, *Complex Sciences*, pages 1907–1916, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Lopes, K. V. R. (2022). Recuperação da informação em vídeos do youtube. *Trabalho de Conclusão de Curso (Graduação em Sistemas de Informação) Universidade Federal de Uberlândia*.
- Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., and Bhattacharjee, B. (2007). Measurement and analysis of online social networks. *IMC '07*, page 29–42, New York, NY, USA. Association for Computing Machinery.
- Tantardini, M., Ieva, F., Tajoli, L., and Piccardi, C. (2019). Comparing methods for comparing networks. *Scientific Reports*, 9.
- Teixeira, Marcela C.; REIS, J. C. S. (2023). Análise do discurso de Ódio em comentários de vídeos no youtube: Um estudo de caso da cpi da covid-19 no brasil. In *SIMPÓSIO BRASILEIRO DE BANCO DE DADOS (SBBD)*, page 330–335. Sociedade Brasileira de Computação.
- Vasconcelos, M., Pereira, E., Guimarães, S., Ribeiro, M. H., Melo, P., and Benevenuto, F. (2020). Analyzing youtube videos shared on whatsapp in the early covid-19 crisis. *WebMedia '20*, page 25–28, New York, NY, USA. Association for Computing Machinery.
- Wang, X., Wu, P., Wang, F., and Wu, T. (2020). Collaborative filtering via social learning. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 50(4):1234–1247.