
 <p>ESCOLA DE COMUNICAÇÃO, ARTES E DESIGN FAMECOS</p>	<h1>REVISTA FAMECOS</h1> <p>mídia, cultura e tecnologia</p> <p>Revista FAMECOS, Porto Alegre, v. 32, p. 1-21, jan.-dez. 2025 e-ISSN: 1980-3729 ISSN-L: 1415-0549</p>
<p> https://dx.doi.org/10.15448/1980-3729.2025.1.45633</p>	

CIBERCULTURA

Algoritmos e técnicas de visualização de dados da audiência de ficção televisiva em *streaming*

Algorithms and data visualization techniques for audiences of television fiction on streaming

Algoritmos y técnicas para visualizar datos de audiencia en la transmisión de ficción televisiva en streaming

Maria Immacolata

Vassallo de Lopes¹

orcid.org/0000-0003-3477-1068

immaco@usp.br

Claudia Freire²

orcid.org/0000-0002-1222-7734

cfreire@ucp.pt

Recebido em: 01 nov. 2024.

Aprovado em: 11 jul. 2025.

Publicado em: 23 set. 2025.

Resumo: Este artigo considera o futuro da pesquisa de audiência na era do *streaming*, questionando a dinâmica e as potencialidades do paradigma *big data* frente ao universo de conteúdo gerado por usuários em plataformas digitais. Partindo dessa premissa, analisa-se o potencial de algoritmos e novas técnicas de visualização para a transmissão de dados de audiência de ficção televisiva à luz de uma lacuna crítica entre os dados de audiência e a própria audiência.

Palavras-chave: recepção, ficção televisiva, *streaming*, técnicas, algoritmos, visualização de dados.

Abstract: This article considers the future of audience research in the streaming era, questioning the dynamics and potentialities of the big data paradigm against the universe of content generated by users on digital platforms. Starting from this premise, the potential of algorithms and new visualization techniques for audience studies is analyzed in light of a critical gap between the data and the audience.

Keywords: reception, television fiction, streaming, techniques, algorithms, data visualization

Resumen: Este artículo considera el futuro de la investigación de audiencias en la era del *streaming*, cuestionando las dinámicas y potencialidades del paradigma *big data* frente al universo de contenidos generados por usuarios en plataformas digitales. Partiendo de esta premisa, se analiza el potencial de algoritmos y de las nuevas técnicas de visualización para la transmisión de datos de audiencia de ficción televisiva a la luz de una brecha crítica entre los datos de audiencia y la audiencia misma.

Palabras clave: recepción de ficción televisiva, streaming, técnicas, algoritmos, visualización de datos.

Questões atuais relacionadas à produção e audiência de ficção televisiva em *streaming*

O mundo dos mercados audiovisuais *on-line* está mudando rapidamente com fornecedores de conteúdo audiovisual que adentram à oferta de *streaming* em um cenário dominado pelos grandes players Netflix, YouTube, Disney+, Pluto/Viacom, Amazon Prime, Apple TV+, HBO Max.



Artigo está licenciado sob forma de uma licença
[Creative Commons Atribuição 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/).

¹ Professora Titular Sênior da Escola de Comunicações e Artes da Universidade de São Paulo – USP. Pesquisadora 1A do CNPq, Docente do Programa de Pós-Graduação em Ciências da Comunicação. Diretora da revista MATRIZES. Coordenadora do Observatório Ibero-americano da Ficção Televisiva (OBITEL), da Rede Obitel Brasil e do Centro de Estudos de Telenovela (CETVN-ECA-USP).

² Pós-Doutoranda no Centro de Estudos de Comunicação e Cultura – CECC – da Faculdade de Ciências Humanas da Universidade Católica Portuguesa em Lisboa.

Nessa indústria em rápida mudança, as inter-relações de competição entre formatos e ofertas de serviços são tão diferentes que a pesquisa acadêmica torna-se relevante ao propor análises e reflexões.

Questões atuais relacionadas à produção de ficção televisiva para *streaming* entram em cena, como apontou a greve iniciada em 14 de julho de 2023, pelo sindicato dos atores americanos, SAG-AFTRA, que expôs disputas trabalhistas entre a indústria criativa versus atores e roteiristas. O principal ponto de discórdia são os pagamentos residuais, uma demanda que infringe o modelo de negócios atual do *streaming*.

No modelo tradicional, atores e roteiristas são pagos pelo trabalho que fazem em um programa de TV ou filme e recebem resíduos quando o conteúdo ficcional é revendido como reprise. Algumas vezes, a renda residual pode ser uma quantia enorme, dependendo da popularidade ou do número de reprises de um programa.

Na era do *streaming*, o modelo mudou de negócio quando as plataformas passaram a produzir conteúdo. Os atores e roteiristas continuam a receber um valor residual, mas trata-se essencialmente de uma taxa fixa. Um dos objetivos da

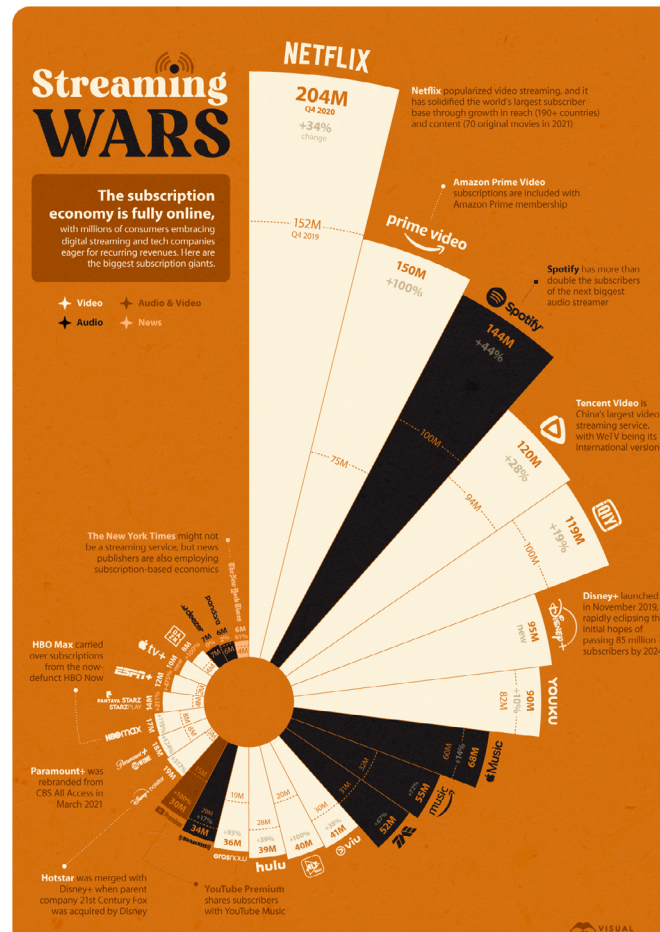
greve é que esses pagamentos sejam baseados na popularidade do programa, ou seja, no indicador de visualizações. Outro grande ponto crítico para os atores é o uso da inteligência artificial. Os atores querem proteções para que suas imagens não sejam reutilizadas digitalmente sem aprovação nem pagamento.

A greve de atores e roteiristas terminou em 9 de novembro de 2023 resultando na maior interrupção das indústrias criativas de televisão e cinema desde o impacto da pandemia de Covid-19. As consequências serão observadas daqui para a frente, seja na América Latina ou na Europa. Diversos filmes que seriam lançados no Natal de 2023 foram cancelados, e o que se observa em 2024 nos catálogos das provedoras de *streaming* é um crescimento da oferta de filmes antigos, remakes e *reality shows*.

Em meio a esse debate, a audiência revela o sucesso dos conteúdos ficcionais nas plataformas. Em 2021, a VisualCapitalist² comparou os dados de assinantes de *streaming* em todo o mundo e propôs uma visualização interessante da audiência para o cenário da chamada *Streaming Wars* ou *Batalha do Streaming*.

² A VisualCapitalist[©] é uma empresa de mídia e de notícias que desenvolve conteúdo visual baseado em dados com foco em tendências globais, investimentos, tecnologia e economia. Disponível em: <https://www.visualcapitalist.com/about/>. Acesso em: 28 jul. 2025.

Figura 1 - Streaming Wars



Fonte: VisualCapitalist© (Wallach, 2021).

De acordo com relatórios posteriores da VisualCapitalist©, em 2022, a Netflix perdeu assinantes pela primeira vez em uma década (menos 200.000) enquanto os rivais continuavam a crescer. Segundo a plataforma de *streaming*, as causas foram a suspensão dos serviços na Rússia em março do mesmo ano, o aumento da competição e o compartilhamento de contas. Em 2023, a maior rival da Netflix é Disney+ (Century Fox, Hulu, Star+ e ESPN+).

De acordo com o último relatório de *streaming* da Conviva (2022), o tempo de *streaming* global aumentou 10% no primeiro trimestre de 2022 em comparação com o ano anterior. O crescimento mais rápido ano a ano no tempo de exibição ocorreu na Ásia (+172%), África (+55%) e América do Sul (+34%).

Na Europa, um estudo comparativo realizado entre a TV tradicional e dois dos serviços *on-line* de maior destaque, YouTube e Netflix, demonstrou que a faixa etária é um indicador fundamental no consumo de *streaming*. Os números de usuários e visualizações de vários países europeus mostram que os jovens usam amplamente portais como o YouTube e têm acesso a serviços pagos de *streaming on-line*, enquanto os grupos etários com mais de 50 anos se conectam significativamente menos a esses serviços, de acordo com Budzinski, Gaenssle e Lindstädt Dreusicke (2021).

O Observatório Europeu do Audiovisual monitora, de forma contínua, a composição de catálogos de VoD, para filmes ou conteúdos televisivos – um tema fundamental, uma vez que os players de VoD são obrigados a oferecer 30% de

conteúdo europeu em seus catálogos, conforme exigido pela revisão da diretiva AVMS³. Os filmes e especialmente lançamentos, são “o pão com manteiga” dos serviços de vídeo transacional sob demanda (TVoD), com uma média de 5.216 títulos de filmes para serviços de TVoD em comparação com uma média de 1.031 para serviços de vídeo sob demanda por assinatura (SVoD) (European Audiovisual Observatory, 2022).

De acordo com o relatório, os serviços TVoD também tiveram uma participação ligeiramente maior de filmes europeus em seus catálogos em comparação com os serviços SVoD, com 31% e 29%, respectivamente. Os maiores exportadores de filmes nos catálogos TVoD, na Europa, foram a França, com uma quota de 30%, a Alemanha, com 22%, a Itália, com 9%, e Espanha, com 7%. Para SVoD, esses mesmos quatro países representaram 66% de todos os filmes exportados da União Europeia.

O Brasil é o segundo país com o maior número de assinantes individuais de plataformas de *streaming*, conforme Lopes, Piñon e Burnay (2022). O *streaming* faz parte da rotina de 71% dos brasileiros, com predominância entre dois players: 91% para Netflix e 53% para Amazon Prime Video. As novidades são o principal motivo de acesso às plataformas para 70% dos entrevistados, principalmente quando se trata de filmes e séries, de acordo com Lopes e Abrão (2023).

Em nosso país somos 149 milhões de usuários de internet, 84% da população brasileira. Destes, 142 milhões de usuários se conectam quase todos os dias. A TV é o segundo dispositivo mais utilizado para acessar a rede no país, atrás apenas do celular. Entre as crianças e adolescentes de 9 a 17 anos, 82% acessaram a internet para assistir a vídeos, programas, filmes ou séries *on-line* (NIC.br, 2024). Embora a demanda por serviços de *streaming* tenha aumentado no país, os principais impedimentos ao seu acesso continuam sendo o custo do serviço e a qualidade do acesso à internet.

Em seu Relatório de Engajamento, a Netflix (2023) provou a importância da diversidade do catálogo oferecido aos assinantes, já que as pessoas assistem, em média, a seis gêneros de conteúdos diferentes por mês e que os títulos mais assistidos em nível global pertencem a doze gêneros diferentes.

A imagem da *Streaming Wars* apresentada na Figura 1 é um exemplo de visualização de dados da audiência de *streaming* que nos interessa neste artigo, envolvendo tecnologia e design gráfico aliados a características persuasivas da informação. Em pesquisas acadêmicas, o trabalho de visualização de resultados, muitas vezes, não é prioridade no Campo da Comunicação. Diante dessa realidade, a proposta desta investigação orienta-se por três objetivos:

- Identificar principais indicadores da audiência de *streaming*;
- Verificar as possibilidades oferecidas por algoritmos e técnicas de visualização de dados para análise da audiência *streaming*;
- Encaminhar formas mais expressivas de apresentação de dados da audiência de ficção televisiva de *streaming*.

A abordagem metodológica proposta para este estudo foi de caráter exploratório com o objetivo de investigar a literatura sobre o tema, bem como *sites* especializados no assunto, e filtrar um número significativo de ferramentas ou técnicas classificadas por pesquisadores e especialistas como sendo adequadas diante das propostas apresentadas de visualização de dados da audiência.

Visualização de dados da audiência de *streaming*

Hoje estamos testemunhando um aumento no uso da visualização de dados em uma variedade de domínios e gêneros. A visualização de dados

³ Diretiva (UE) 2018/1808: do Parlamento Europeu e do Conselho Europeu, de 14 de novembro de 2018, que regulamenta, entre todos os Estados-membros, a oferta de serviços de comunicação social audiovisual, que inclui canais de televisão por subscrição, plataformas de partilha de vídeos e serviços audiovisuais VOD (“video on demand”). Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A32018L1808>. Acesso em: jul. 2022.

(também chamadas de Dataviz ou DV) é utilizada para facilitar o entendimento e a compreensão rápida da informação e também tem como objetivo persuadir o leitor. Consequentemente, pode-se compreender a visualização de dados como artefato cultural com características semióticas, estéticas e sociais (Kennedy; Engebretsen, 2020).

A visualização de dados é uma tecnologia – ou um conjunto de tecnologias – que, tal qual artefatos como o relógio, a bússola ou o mapa, é capaz de transformar a maneira como vemos e nos relacionamos com a informação. É por isso que criar ou desenvolver tecnologias não consiste apenas em criar objetos; em vez disso, quando as tecnologias surgem, “novos mundos estão sendo criados” (Cairo, 2020, p. 18). Mas que “novos mundos” a visualização de dados é capaz de gerar?

Uma filosofia da visualização de dados pode derivar temas, metodologias e linguagem a partir de uma ampla gama de disciplinas: epistemologia, sociologia, semiótica, história, ética, campos da teoria crítica, como a cartografia crítica, ou da filosofia da ciência, estatística, arte e, talvez mais que qualquer outro campo, a filosofia da tecnologia (Cairo, 2020). Os filósofos da visualização devem raciocinar sobre a história, suposições, convenções, práticas e impactos da visualização em indivíduos, culturas e sociedades.

Novas ferramentas e técnicas para coletar, filtrar, analisar e visualizar dados tornam os processos criativos mais fáceis e baratos. Também estamos testemunhando novas arenas para disseminação e engajamento com dados, cada vez mais usados para fins informativos, persuasivos e retóricos em campanhas políticas, comunicação em saúde, educação e em redações, onde novas equipes dedicadas à visualização de dados estão sendo implementadas. (Kennedy; Engebretsen, 2020).

Big data tornou-se um conceito famoso caracterizado por três Vs: volume, variedade e velocidade. Mais recentemente, Vs adicionais foram propostos, como variabilidade e valor (Kitchin, 2014). O *big data* é exaustivo em escopo, abrange populações ou sistemas inteiros (n=todos);

refinado em resolução e possibilita indexação única para identificação. Relacional por natureza, os dados contêm campos comuns que permitem diferentes combinações, são flexíveis, mantendo os traços de extensibilidade (aceita facilmente a adição de novos campos) e possui escalabilidade (pode expandir em tamanho rapidamente).

A análise de *big data* permite uma abordagem epistemológica inteiramente nova para dar sentido ao mundo; em vez de testar uma teoria analisando dados relevantes, as novas análises de dados buscam obter *insights* teóricos a partir dos dados e inauguraram uma nova era de empirismo, em que as técnicas analíticas são capazes de permitir que os dados falem por si mesmos, sem a necessidade de um olhar teórico.

Neste sentido, o *big data* traria consigo o renascimento do empirismo. Anderson (2008) argumenta que o dilúvio de dados torna o método científico um método obsoleto; que os padrões e relacionamentos contidos no *big data* produzem inerentemente conhecimento significativo e perspicaz sobre fenômenos complexos atuais. Tal visão empirista tem ganhado credibilidade, especialmente nos círculos empresariais, mas suas ideias também se enraizaram no novo campo da ciência de dados e em outras ciências.

A visualização de dados também é capaz de criar significados por meio de códigos visuais e de gerar sentimentos, ou seja, respostas emocionais. Significados e sentimentos são inseparáveis em nossas interações com os textos. Um estudo de Kennedy e Hill (2017) revelou que as visualizações de dados despertam uma ampla gama de sentimentos nas pessoas, ativados tanto pelo conteúdo textual das visualizações como pelos fatores contextuais dos usuários ou pela situação física e psicossocial de uso.

Na análise dos resultados de suas pesquisas, os autores Kennedy e Hill (2017) argumentam que a emoção pode ser entendida como um recurso epistêmico, uma forma de conhecimento que é valiosa para a construção crítica do mundo, no qual, sem emoções, a capacidade de tomar decisões racionais fica prejudicada. Sendo assim, as emoções são componentes vitais para a com-

preensão do mundo social e suas representações, o que inclui a visualizações de dados.

Manovich (2020) ressalta a importância de pesquisadores das ciências humanas começarem a usar sistematicamente a visualização de dados na pesquisa, no ensino e na apresentação pública de resultados. Ou seja, toda pesquisa teria um produto midiático como resultado, em vez de representá-lo apenas por gráficos. A combinação de técnicas de análise visual com conjuntos de dados culturais e questões de pesquisa em Humanidades possibilita uma nova abordagem, que o autor chama de *cultural analytics*, que assume os desafios de como acessar e visualizar grandes coleções de mídia cultural de forma mais eficiente (Yamaoka *et al.*, 2011).

De acordo com Van Es, Wieringa e Schafer (2018), a ideia de que as tecnologias são tentenciosas não é nova e tem sido estudada por várias décadas dentro do campo dos New Media Studies. Entretanto, os autores sublinham a importância de refletir sobre as ferramentas que utilizam e estendem um apelo à ação de colegas acadêmicos para que se envolvam em tais críticas propondo a reflexividade.

Nosso engajamento emocional com visualizações de dados também está intimamente ligado aos seus aspectos estéticos. As formas, cores e arranjos das visualizações de dados acionam nossos sentidos de maneira particular. Por sua vez, a interação entre os aspectos semióticos de criação de significado das visualizações de dados e as emoções que evocam estão intimamente relacionadas à capacidade de eliciar o engajamento social.

O conceito de engajamento tem várias camadas, de acordo com Kennedy e Engebretsen (2020), pode referir-se à interação real com uma visualização de dados, engajando-se com ela, ou a respostas emocionais e práticas, envolvendo-se por ela. O engajamento por meio das representações visuais de dados se inicia com o processo de olhar, ler, interpretar e pensar, quando as pessoas tentam dar sentido à proposta de representação visual.

A audiência mundial de *streaming* pode ser

caracterizada como *big data*, e os relatórios sobre o comportamento dessa audiência ganharam relevância cada vez maior no setor do audiovisual. Os relatórios da audiência contêm informações multimodais aliando texto, gráfico e imagens que facilitam sua compreensão (WU *et al.*, 2021).

Em nosso estudo identificamos as seguintes métricas na maioria dos relatórios analíticos de *streaming* de acordo com Pham *et al.* (2021):

- Número de visualizações (Stream ou Playback Stream)
- Dispositivos do Usuário (device viewership or end user devices)
- Minutos/dispositivo único (Minutes/unique device)
- Plataformas ativas (Delivering Platform or Active Platforms) e navegadores (Browsers) e URLs
- Rastreamento de visitas (Leads)
- Dados demográficos da audiência (Demographic Data)
- Métricas de Qualidade de Serviço (Quality of Service – QoS)
- Métricas de Qualidade da Experiência (Quality of Experience – QoE or Session Details)
- Métricas de engajamento (Engagement Metrics)

É fato que o crescimento da visualização de dados trouxe grandes avanços para muitos campos do conhecimento, especialmente a Comunicação. E isso só aconteceu devido à contribuição de várias comunidades de pesquisadores, entre elas a computação gráfica e a estatística, ambas aliadas à ciência da computação, que avançou no desenvolvimento de sistemas que permitem aos usuários especificar e interagir com programas de visualização, tais como: Vega-Lite, VizQL, Tableau, Microsoft Power BI, Hyper DB, Projeto Falcon (Qin *et al.*, 2020).

Apesar de o produto final ser conciso, as técnicas empregadas para a visualização de dados de *streaming* são complexas. Novos sistemas estão sendo desenvolvidos para aperfeiçoar a leitura

de dados. Khalid e Zeebaree (2021) revisaram trabalhos de pesquisa relacionados à visualização de *Big Data* e concluíram que métodos de abordagem da visualização podem incluir até mesmo recursos de realidade virtual para observação e análise de estruturas de dados diversas e complexas.

Algoritmos e a visualização da informação

Ao observar o avanço da metodologia analítica, é possível dizer que a pesquisa em Comunicação há algum tempo convive com técnicas de pesquisa (algoritmos e soluções) de agências de monitoramento ou iniciativas em código aberto que fornecem relatórios já prontos, ou seja, realizam a coleta, o trabalho com os dados e apresentam os resultados.

Sobretudo, ao falar em metodologia de pesquisa para a internet no Campo da Comunicação, estamos também a falar de algoritmos, querendo ou não, sabendo ou não, pois enquanto pesquisadores entramos em contato com esses blocos de código para seleção, organização e coleta de dados.

De início, a pressuposta neutralidade, cientificidade e, mais ainda, a rapidez provida pelos algoritmos parece ter solucionado grande parte dos problemas de pesquisa sobre os dados de audiência.

Mas é bom lembrar que as empresas e as agências de monitoramento são as detentoras das soluções. Observando esse processo podemos dizer, de certa forma, que perdemos o acesso ao *como*, que é propriamente a natureza e o aspecto da metodologia de pesquisa, à medida que acessamos apenas o relatório final, com as categorias de análise já propostas e os indicadores e dados já trabalhados. Resta-nos uma análise final, mas ausente das possibilidades de conhecer, a fundo, o método.

De forma que é preciso atentar para o fato

de que os resultados da audiência de *streaming* obtidos por meio de registros das ações dessa mesma audiência e pela análise de algoritmos estão na ordem do dia, o que demanda um estudo de algoritmos no Campo da Comunicação e nas agendas de pesquisa. Estudos guiados pelo ponto de vista do processo comunicativo e de produção e consumo de cultura, como pesquisadores que acompanham e observam o desenvolvimento de software como um produto cultural de nosso tempo, a exemplo de Manovich (2020), Schäfer e Van Es (2017) e Corrêa e Bertocchi (2012).

Mas o que são algoritmos? Informalmente, um algoritmo é um bloco de código computacional que define um procedimento para resolver um problema específico a partir da entrada de um valor ou conjunto de valores (*input*) e da saída de algum valor ou conjunto de valores (*output*). Um algoritmo é, portanto, uma sequência de etapas computacionais que transformam o *input* no *output* (Cormen *et al.*, 2001).

Também podemos definir um algoritmo como uma ferramenta para resolver um problema computacional bem especificado. O enunciado do problema deve especificar, em termos gerais, a relação de entrada/saída desejada. Qual algoritmo é melhor para um determinado aplicativo depende – entre outros fatores – do número de itens a serem classificados, até que ponto os itens já estão classificados, possíveis restrições nos valores dos itens e o tipo de dispositivo de armazenamento a ser usado, memória, *hardware* (Cormen *et al.*, 2001).

Mas o conceito de algoritmo é, em si, uma questão para a área da Filosofia da Tecnologia, pois o desenvolvimento ou modelagem de algoritmos exige altos níveis de abstração⁴ (Francik, 1999; Hill, 2016). Dessa forma, a noção de algoritmo não pode ser rigorosamente definida em toda a sua generalidade, e a razão desse fato é que o conceito de algoritmo está em constante expansão.

⁴ Abstração: de acordo com Francik (1999), a abstração é uma operação da mente que isola um elemento de uma noção enquanto negligencia outros. Na visualização de software, a abstração consiste em ignorar alguns aspectos de um programa ou algoritmo que está sendo visualizado, caso não constitua um meio eficaz para a compreensão do algoritmo. Por outro lado, consiste também em introduzir na visualização alguns elementos que não estão presentes no programa executado ou em seus dados, mas que ajudam a compreender o *software*.

Para melhor compreensão, os autores fazem analogia entre algoritmos e números. Muitos tipos de números foram introduzidos ao longo da história: números inteiros positivos, números naturais, racionais, reais, números complexos, quatérnios, cardinais infinitos, ordinais infinitos, etc. Da mesma forma, muitos tipos de algoritmos foram introduzidos. Além dos algoritmos sequenciais clássicos, temos agora algoritmos paralelos, interativos, distribuídos, em tempo real, analógicos, híbridos, quânticos, etc.

Novos tipos de números e de algoritmos estão sendo desenvolvidos e introduzidos na Ciência da computação de maneira que as noções de números e algoritmos não se cristalizaram para suportar definições rigorosas, o que talvez nunca aconteça (Moschovakis, 2001).

A questão sobre a compreensão do conceito de algoritmo é tão séria que professores e programadores desenvolveram soluções que permitem aos alunos da Ciência da Comunicação e da Matemática visualizarem algoritmos em ação.

Os primórdios da visualização de algoritmos remontam à década de 1940, com aumento significativo do desenvolvimento de algoritmos a partir dos anos 1980. Mas o maior avanço na área pode ser observado nos últimos 20 anos (Šimoňák, 2014). Algumas das soluções contemporâneas para visualização de algoritmos incluem sistemas como VizAlgo⁵, TRAKLA23, ANIMAL Algorithm Visualization⁶, JAWAA5 ou Algoritmos em Ação e também os aplicativos Algorithms: Explained and Animated⁷ e Algo – Algorithm Interpreter⁸.

É fato que, por mais experientes que os cientistas de dados possam ser, não há como fazer páreo ao processamento bruto e ao poder de análise de algoritmos, que percorrem milhões de pontos de dados em segundos e muitas vezes podem encontrar *insights* mais profundos de conjuntos de dados maiores. Ao adicionar um componente de aprendizado de máquina, os algoritmos de um software de visualização tornam-se mais inteligentes e podem aprender

a procurar conexões mais profundas em um conjunto de *big data*.

Como foram projetados para melhorar automaticamente suas análises à medida que verificam as informações, as ferramentas de aprendizado de máquina dão conta de fluxos enormes de dados constantes e fornecem análises em tempo real. Os algoritmos responsáveis por esse tipo de análise foram, ironicamente, chamados por alguns autores de *streaming algorithms* (Kejariwal; Kulkarni; Ramasamy, 2015). Entre os exemplos desses sistemas estão: o Mill Wheel, desenvolvido pelo Google; Apache Samza, desenvolvido pelo LinkedIn; o AKKA; o MOA; o Apache Spark, que saiu de Berkeley para o Hadoop; o Apache Storm, sistema amplamente popular e de código aberto desenvolvido pelo Twitter; o Pulsar, mecanismo de análise em tempo real desenvolvido pelo eBay; Flink; Heron.

Mas muitas pessoas relacionam o aprendizado de máquina e algoritmos como sendo a tecnologia por trás do Instagram, do YouTube e do Facebook responsável por selecionar posts de preferência dos usuários ou realizar o preenchimento automático de termos favoritos nos mecanismos de pesquisa. Pode-se dizer que esse é um dos frutos do relacionamento entre os algoritmos e a visualização da informação: a persuasão.

No início, o *feed* do Instagram era visto por todos os usuários a partir do critério cronológico de inserção dos posts, ou seja, de acordo com a ordem de postagem. Em 2016, os algoritmos do Instagram mudaram, e a rede social passou a mostrar o conteúdo por relevância, ou seja, os algoritmos passaram a identificar os temas mais relevantes para o usuário e permitiam apenas a visualização destes. De acordo com a empresa, a intenção dessa metodologia não era limitar o alcance dos stories, vídeos ou posts no *feed*, mas trazer a melhor experiência possível ao usuário, entregando diretamente conteúdo de preferência e eliminando o que não interessava.

⁵ Disponível em: <https://visualgo.net/en>. Acesso em: 31 jul. 2025.

⁶ Disponível em: <http://www.algoanim.net/>. Acesso em: 31 jul. 2025.

⁷ Disponível em: https://play.google.com/store/apps/details?id=wiki.algorithm.algorithms&hl=en_US&gl=US. Acesso em: 31 jul. 2025.

⁸ Disponível em: https://play.google.com/store/apps/details?id=com.abdo.algo&hl=en_US&gl=US. Acesso em: 31 jul. 2025.

Assim, o usuário desenvolveria afinidade com o conteúdo e teria mais chances de voltar ao aplicativo. Alguma semelhança com os algoritmos de recomendação da Netflix?

Ao inserir o valor relevância em algoritmos nas plataformas de OTT, fazendo com que se torne a regra primeira de saída da informação, o que ocorre é que a audiência está sempre exposta a mais informações sobre o mesmo (seja esse "mesmo" um filme, a série, um tema, gênero, artista, roteiro, formato, personagem, etc.). E a audiência está sujeita a sempre mais do mesmo, apesar de o número de itens cadastrados na base de dados ou de o número de ficções ser de milhares, justamente porque os algoritmos identificam o gosto e a preferência do usuário. Ora isso não é novidade e parece funcionar muito bem, pois quantas vezes assistimos a um filme ou a uma série mais de uma vez ou até inúmeras vezes porque gostamos e coincidentemente essa opção de conteúdo aparece logo nas primeiras linhas de visualização da plataforma de *streaming*.

Quando se trata de aprendizagem de máquinas, há uma série de algoritmos que podem ser utilizados para o trabalho de organização e visualização de dados. O aprendizado consiste na tarefa de inferir uma função a partir de dados de treinamento ou um conjunto de exemplos. Os algoritmos de aprendizado de máquina são classificados em 4 tipos: (a) aprendizagem supervisionada, (b) aprendizagem não supervisionada, (c) aprendizagem semissupervisionada, (d) aprendizado por reforço.

A Classificação de Algoritmos é uma técnica de Aprendizagem Supervisionada utilizada para identificar a categoria de novas observações de dados com base nos exemplos de treinamento. Na Classificação, um programa aprende, a partir de um determinado conjunto de exemplos de dados, a classificar os novos dados em várias categorias, rótulos ou grupos. Desse modo, a classificação orienta uma entrada e uma saída correspondente. Ela pode ser binária, *multi-class*

ou *multi-label*.

A seguir apresentamos a classificação de algoritmos de aprendizagem de máquina mais utilizados atualmente. Essa lista não pretende ser exaustiva, mas apenas mostrar um retrato de algoritmos mais aplicados para organização e visualização da informação em nosso tempo atual de pesquisa⁹.

Linear regression (LR): um algoritmo de regressão linear estabelece uma relação entre variáveis independentes e dependentes, ajustando-as a uma linha. Essa linha é conhecida como linha de regressão e representada por uma equação linear. A regressão linear tem muitos usos práticos. Se o objetivo for predição, previsão ou redução de erro, a regressão linear pode ser usada para ajustar um modelo preditivo a um conjunto de dados observados de valores da resposta e variáveis explicativas. Se o objetivo é explicar a variação na variável de resposta, a análise de regressão linear pode ser aplicada para quantificar a força da relação entre a resposta e as variáveis explicativas e, em particular, para determinar se algumas variáveis explicativas podem não ter nenhuma relação linear com a resposta, ou para identificar quais subconjuntos de variáveis explicativas podem conter informações redundantes sobre a resposta. A regressão linear é uma técnica estatística muito poderosa e pode ser usada para gerar *insights* sobre o comportamento do consumidor, entender o negócio e os fatores que influenciam a lucratividade. Podem ser usadas nos negócios para avaliar tendências e fazer estimativas ou previsões, avaliar o risco em serviços financeiros ou no ramo de seguros e serviços de crédito.

Regressão logística (RLog): a regressão logística é usada para estimar valores discretos (geralmente valores binários como 0/1) de um conjunto de variáveis independentes. Ele ajuda a prever a probabilidade de um evento ajustando os

⁹ Fontes: IBM Cloud Learn Hub for Artificial Intelligence, JavaPoint Classification Algorithm in Machine Learning, Brownlee (2021), Shrestha e Mohmood (2019), Valero Medina e Alzate Atehortúa (2019), Pouyanfar *et al.* (2018), LeCun, Bengio, Hinton (2015), Kranen (2011), Madhulatha (2011), Otukey e Blaschke (2010), Di Battista *et al.* (1994) além de pesquisa extensiva em *blogs* dedicados à classificação de algoritmos e aprendizagem de máquinas para organização da informação e visualização de dados.

dados a uma função lógica. Em essência, quando se tem um grande conjunto de dados que se deseja categorizar, recomenda-se a aplicação da regressão logística. Os usos práticos desses algoritmos podem ser observados no sistema financeiro em pontuações de crédito, na medicina em testes rápidos de sangue e testes biomoleculares, nos editores de texto e processamento natural da linguagem, em *sites* comparativos de reservas de hotéis e na indústria de games.

Decision tree: esse algoritmo é usado para classificar problemas, temas, decisões. Dividir e conquistar é a meta, utilizando tarefas de classificação e regressão. Dessa forma, ele organiza os dados dividindo-os em dois ou mais conjuntos homogêneos com base nos atributos/variáveis independentes mais significativos. Possui uma estrutura hierárquica de árvore, que consiste em um nó raiz, ramos, nós internos e nós folha. Uma árvore de decisão começa com um nó raiz que não possui ramificações de entrada. As ramificações de saída do nó raiz alimentam os nós internos, também conhecidos como nós de decisão. Com base nos recursos disponíveis, ambos os tipos de nós conduzem a avaliações para formar subconjuntos homogêneos, que são denotados por nós folha ou nós terminais. O resultado é um fluxograma ramificado que torna fácil avaliar as tomadas de decisões diante do problema, permitindo que diferentes grupos em uma organização entendam melhor por que uma decisão foi tomada. Exemplos desse algoritmo são o C4.5 Top Down Decision Tree e o J48.

Support Vector Machine algorithm (SVM): essa classe de algoritmo foi criada por Vladimir N. Vapnik e Alexey Ya. Chervonenkis em 1963. Desde então, os sistemas têm sido utilizados na classificação de texto, hipertexto e imagem. Os SVMs podem identificar e trabalhar com caracteres manuscritos e têm sido usados em laboratórios de biologia para realizar tarefas como classificar proteínas. São também usados em chatbots, carros autônomos, em programas de reconhecimento facial e robôs. No algoritmo SVM os dados brutos são plotados como pontos em um espaço

n-dimensional (n =número de recursos). O valor de cada recurso é então vinculado a uma coordenada específica, facilitando a classificação dos dados. São usadas para classificar dois grupos de dados por classificação semelhante. O algoritmo desenha linhas (hiperplanos) para separar os grupos de acordo com padrões. Um SVM pode construir um modelo de aprendizagem que atribui novos exemplos a um grupo ou outro. Por essas funções, o SVM é chamado de classificador linear binário não probabilístico. Como outras máquinas de aprendizado supervisionado, um SVM requer que dados rotulados sejam treinados. Os grupos de materiais são rotulados para classificação. Depois de processar vários exemplos de treinamento, os SVMs podem realizar aprendizado não supervisionado. Os algoritmos tentarão obter a melhor separação de dados com o limite ao redor do hiperplano sendo maximizado e até mesmo entre os dois lados.

Naive Bayes algorithm: de acordo com os especialistas, esse é o algoritmo mais simples que se pode aplicar aos dados. Como o nome sugere, o algoritmo faz uma suposição, pois todas as variáveis no conjunto de dados são "ingênuas", ou seja, não correlacionadas entre si. Esse algoritmo funciona muito rápido e pode prever facilmente a classe de um conjunto de dados. Devido à rapidez e eficiência, é muito comum usá-lo em soluções para fazer previsões em tempo real. Também é muito popular para previsões multiclasse. Os serviços de e-mail (como o Gmail) usam esse algoritmo para descobrir se um e-mail é spam ou não. Esse algoritmo é excelente na Análise de Sentimentos de um grupo-alvo (clientes, público, etc.). A filtragem colaborativa e o algoritmo Naive Bayes trabalham juntos para construir sistemas de recomendação. Esses sistemas usam mineração de dados e aprendizado de máquina para prever se o usuário gostaria de um determinado conteúdo ou não; isso faz com que o Bayes ingênuo seja amplamente aplicado na prática.

KNN algorithm: esse algoritmo pode ser aplicado tanto para problemas de classificação

quanto para problemas de regressão. Ele armazena todos os casos disponíveis e classifica quaisquer novos casos realizando uma votação majoritária de empates entre os k vizinhos. O caso é então atribuído à classe com o qual tem mais em comum. É mais eficiente em bases de dados menores, pois é lento na aprendizagem. Aplicações práticas: mineração de texto, dados do setor de agricultura, finanças, medicina, reconhecimento facial, sistemas de recomendação (Amazon, Hulu, Netflix, etc.).

K-means: esse algoritmo classifica os conjuntos de dados em um determinado número de clusters de forma que todos os pontos de dados dentro de um cluster sejam homogêneos e heterogêneos em relação aos dados em outros clusters. K-means é um método que consiste em particionar um conjunto de n objetos em $k \geq 2$ clusters, de modo que os objetos de um cluster sejam semelhantes entre si e diferentes daqueles em outros clusters. O termo "k-means" foi usado pela primeira vez por James MacQueen em 1967, embora a ideia remonte a Hugo Steinhaus em 1956. O algoritmo padrão foi proposto pela primeira vez por Stuart Lloyd, do Bell Labs, em 1957 e, em 1965, Edward W. Forgy publicou essencialmente o mesmo método, razão pela qual às vezes é chamado de algoritmo Lloyd-Forgy. O K-means forma uma família que inclui algoritmos que formam clusters a partir de particionamento, clustering baseado em densidade, clustering baseado em modelo de distribuição, agrupamento hierárquico, propagação por afinidade e fuzzy clustering – Fuzzy C-means algorithm ou Fuzzy K-means algorithm. Exemplos: o algoritmo de Louvain (aplicado no estudo de Regattieri, Malini e Medeiros, 2014), o Force Atlas, o Yifan Hu e o Fruchterman Reingold, que estão disponíveis em programas como o Gephi e o NodeXL, o CLARANS, CURE, STREAM, e os algoritmos de Floyd (1962) e algoritmo de Dijkstra (1959). Algoritmos de clustering são utilizados também pela Amazon e Netflix em seu sistema de recomendação aos usuários a partir do histórico anterior de pesquisa de produtos.

Random forest algorithm: o primeiro algoritmo random forest foi criado em 1995 por Tin Kam Ho. Em 2019, uma extensão do algoritmo foi desenvolvida por Leo Breiman e Adele Cutler, que registraram o termo *Random Forests* em 2006, que passou a ser propriedade da Minitab Inc. a partir de 2019. É um algoritmo que constrói "florestas" a partir de um conjunto de dados oriundos de *decision trees*. Simplificando: o random forest lida com conjuntos de árvores de decisão e as mescla para obter uma previsão mais precisa e estável. Uma grande vantagem é que pode ser usado para problemas de classificação e regressão, que formam a maioria dos sistemas de aprendizado de máquina atuais. O algoritmo é aplicado em sistemas financeiros (bancos), mercado de ações, medicina e comércio eletrônico. Em finanças, por exemplo, é usado para detectar clientes com maior probabilidade de pagar suas dívidas em dia, usar certos serviços com mais frequência e para detectar fraudes. Na bolsa de valores ou Bitcoin, o algoritmo pode ser usado para determinar o comportamento futuro de uma ação. Na área da saúde, é usado para identificar a combinação correta de componentes na medicina e analisar o histórico médico de um paciente com a finalidade de identificar doenças. O random forest também é usado no comércio eletrônico para determinar se um cliente realmente vai gostar ou não de um novo produto. Ultimamente também tem sido utilizado em aplicações para leitura e visualização de dados de satélite.

Artificial Neural Networks (ANN): algoritmos de redes neurais, também conhecidos como redes neurais artificiais (ANNs) ou redes neurais simuladas (SNNs), são um subconjunto do aprendizado de máquina que também engloba os principais algoritmos de deep learning: Convolutional Neural Networks (CNNs), Long Short Term Memory Networks (LSTMs), Recurrent Neural Networks (RNNs), Generative Adversarial Networks (GANs), Radial Basis Function Networks (RBFNs), Multilayer Perceptrons (MLPs), Self Organizing Maps (SOMs), Deep Belief Networks

(DBNs), Restricted Boltzmann Machines (RBMs), Autoencoders. O nome *redes neurais e estrutura* são inspirados no cérebro humano, imitando a maneira como os neurônios biológicos realizam a sinapse uns com os outros. As redes neurais artificiais (RNAs) são compostas por camadas de nós, contendo uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada nó, ou neurônio artificial, conecta-se a outro e tem um peso e um limiar associados. Se a saída de qualquer nó individual estiver acima do valor limite especificado, esse nó é ativado, enviando dados para a próxima camada da rede. Caso contrário, nenhum dado é passado para a próxima camada. A complexidade da rede pode ser alterada dependendo do número de camadas e do número de unidades em cada camada. As redes neurais precisam de treinamento para aprender e melhorar a precisão ao longo do tempo. No entanto, uma vez que esses algoritmos de aprendizado são ajustados, tornam-se ferramentas poderosas para classificar e agrupar dados em alta velocidade. Uma das redes neurais mais conhecidas é o algoritmo de busca do Google. Pense em cada nó individual com seu próprio modelo de regressão linear, composto de dados de entrada, pesos, um viés (ou limite) e uma saída. O Multilayer Perceptron é o classificador de rede neural mais utilizado nos dias atuais. Dentre as aplicações práticas podemos citar o reconhecimento de caligrafia e caracteres manuscritos, a compressão de imagem, previsões da Bolsa de Valores, reconhecimento facial, amplamente aplicado em *sites* de mídia social, em projetos aeroespaciais, operações militares e de defesa, na área da saúde (Bayesian Penalized-Likelihood – BPL – algoritmo) e na previsão do tempo. R-CNN é um algoritmo da família neural, desenvolvido por Girshick *et al.*, da UC Berkeley em 2014, utilizado para a detecção de objetos em também a posição deles em uma determinada área ou espaço. Ele classifica imagens, detecta objetos, localiza a presença de objetos em um ambiente, enfim, é o principal algoritmo utilizado para auxiliar os robôs a verem.

Maximum likelihood classification – MLC algorithm: vários algoritmos de classificação baseados em semelhança de pixels foram desenvolvidos, muitos dos quais produzem certas otimizações e amadurecem gradualmente. A classificação de máxima verossimilhança (MLC) é o método mais utilizado para processamento de dados de satélite, mapeamento e avaliação da cobertura do solo.

Dimensionality reduction algorithms: esses algoritmos auxiliam no processo de aprendizagem de máquina quando há muitos fatores com base nos quais a classificação final é feita. Esses fatores são basicamente variáveis chamadas de características. Quanto maior o número de recursos, mais difícil fica visualizar o conjunto de treinamento e depois aperfeiçoá-lo. Às vezes, a maioria desses recursos é redundante. É aqui que os algoritmos de redução de dimensionalidade entram em ação. A redução de dimensionalidade é o processo de redução do número de variáveis aleatórias em consideração, por meio da obtenção de um conjunto de variáveis principais. Ele pode ser dividido em seleção de recursos e extração de recursos, para organizar o treinamento.

Boosting algorithms: é uma técnica de aprendizado de conjunto sequencial na qual diferentes modelos são adicionados iterativamente para melhorar o desempenho geral de classes de algoritmos. Exemplos: AdaBoosting algorithm, que aproveita os casos classificados com erro do modelo anterior e cria uma nova amostra de dados ponderada na qual os casos classificados com erro têm mais peso. O AdaBoost é um dos primeiros algoritmos de reforço desenvolvidos. Raramente é usado nos dias de hoje, mas é a base da maioria dos algoritmos de reforço. Gradiente boosting, comparado ao AdaBoost, o aumento de gradiente não penaliza os casos classificados como perdidos, mas usa a função de perda. Os mais atuais são os XGBoost e Light GBM.

A visualização de um resumo sobre a Classificação de Algoritmos de Aprendizagem de Máquina foi realizada por Brownlee (2021) em

um mapa mental:

Figura 2 - Classificação de algoritmos de aprendizagem de máquina



Fonte: Brownlee (2021).

Essa breve aproximação sobre algoritmos teve como objetivo prover certo entendimento sobre esses blocos de códigos e expressões matemáticas, como podem ser utilizados para organizar a informação a partir de suas classificações inerentes. A grande maioria deles é desenvolvida em Python¹⁰ e em R¹¹. Mas a genialidade dessas soluções está na combinação de classes de algoritmos, ou seja, na adaptação que vai compor uma forma mais apurada e mais rápida de saída de dados, além de permitir uma predição mais acurada. É o exemplo de combinações feitas utilizando Supervised Tensor Learning (STL) a

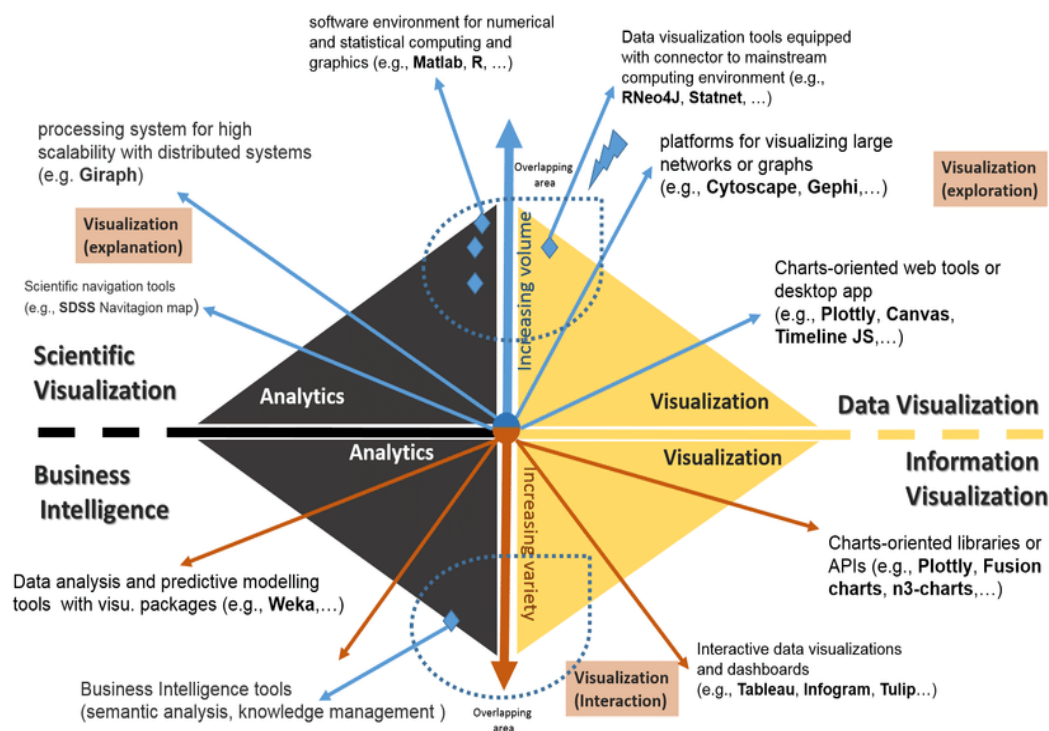
algoritmos de R-CNN, métodos Bayesianos com logística regressa e Stochastic Gradient Descent/ Back propagation.

Resultados: soluções e técnicas para a visualização de dados da audiência de streaming

Em 2017, Caldarola e Rinaldi publicaram um panorama com mais de 36 soluções para visualização de *Big Data* e que podem ser aplicados à audiência de *streaming*. Os resultados foram apresentados sob a forma de resumo por meio da figura a seguir:

¹⁰ Python: linguagem de programação, disponível em: <https://www.python.org/>. Acesso em: 31 jul. 2025.

¹¹ R: linguagem de programação estatística e gráfica, disponível em: <https://www.r-project.org/>. Acesso em: 31 jul. 2025.

Figura 3 - Soluções para visualização de *Big Data*

Fonte: Caldarola e Rinaldi (2017, p. 299).

No estudo, Caldarola e Rinaldi (2017) separaram os resultados entre as seguintes áreas de visualização: interativa, explicativa e visualização científica ou exploratória.

Nos quadrantes sul-sudeste, foram agrupadas ferramentas de visualização interativa e painéis juntamente com bibliotecas ou APIs orientadas a gráficos, categorizadas como ferramentas ou módulos de visualização de informações. Os autores descrevem cada uma dessas ferramentas, mas aqui apenas citaremos os resultados encontrados para essa categoria:

1. Tableau – <https://www.tableau.com/products>
2. Infogram – <https://infogr.am/>
3. ChartBlocks – <http://www.chartblocks.com/en/>
4. Plottly – <https://plot.ly/>
5. [D3.js](https://d3js.org/) – <https://d3js.org/>
6. Ember Charts – <http://addepar.github.io/ember-charts/>
7. Google charts – <https://developers.google.com/chart/>
8. FusionCharts – <http://www.fusion-charts.com/>
9. [chart.js](http://www.chartjs.org/) – <http://www.chartjs.org/>
10. Leaflet – <http://leafletjs.com/>
11. [Chartist.js](https://gionkunuz.github.io/chartist-js/) – <https://gionkunuz.github.io/chartist-js/>
12. Sigma JS – <http://sigmajs.org/>
13. n3-charts – <https://github.com/n3-charts>
14. Polymaps – <http://polymaps.org/>
15. GeoJSON – <https://geojson.org/>
16. [Processing.js](http://processingjs.org/) – <http://processingjs.org/>

17. dygraphs – <http://dygraphs.com/>

Na região nordeste da Figura 3, os autores agruparam as ferramentas web orientadas a gráficos ou aplicativos de desktop juntamente com plataformas para visualização de grandes redes ou gráficos e ferramentas de visualização de dados equipadas com conectores para fazer a interface com ambientes de computação convencionais, como Matlab¹² ou R. A essa categoria também pertence o Plottly já citado anteriormente. Nesse agrupamento os autores mencionam:

1. Timeline – <http://timeline.knightlab.com/>
2. Canvas – <http://canvasjs.com/>
3. Commetrix – <http://www.commetrix.de/>
4. Cuttlefish – <http://cuttlefish.sourceforge.net/>
5. Cytoscape – <http://www.cytoscape.org>
6. Gephi – <https://gephi.org/>
7. Graph-tool – <https://graph-tool.skeuwed.de/>
8. Graphviz – <http://graphviz.org/>
9. JUNG (Java Universal Network/Graph Framework) – <http://jung.sourceforge.net/>
10. Keynetiq – <https://www.keynetiq.com/>
11. Netlytic – <https://netlytic.org/home/>
12. NetMiner – <http://www.netminer.com/main/main-read.do>
13. Network Workbench – <http://nwb.cns.iu.edu/>
14. NodeXL – <https://www.smrfoundation.org/nodexl/>

15. UCINET – <https://ucinet-for-windows.software.informer.com/>

16. Pajek – <https://pajek.software.informer.com/>

17. Statnet – <http://statnetproject.org/>

18. Tulip – <http://tulip.labri.fr/TulipDrupal/>

19. Visone – <http://visone.info/html/demo.html>

Dentro da região noroeste da Figura 3, foram coletadas as ferramentas que os autores classificaram como sendo voltadas à visualização científica e exploratória, como sistemas de processamento para alta escalabilidade com sistemas distribuídos, por exemplo: Giraph¹³, SocNetV¹⁴ e Sentinel Visualizer¹⁵.

Em outros estudos, Khalid e Zeebaree (2021) também mencionaram o SAAS Visual Analytics e Microsoft Power BI. Sallaberry (2020) cita SentiCompass¹⁶, HydroQual¹⁷, EpiNews, e Kennedy e Hill (2017) citam uma lista de mais de 100 ferramentas¹⁸ para visualização de dados, e Kucher e Kerren (2015) oferecem outra lista com mais de 400 técnicas¹⁹ para análise e visualização de textos, além das bases de dados DataViz²⁰ e R Graph Gallery²¹. Essas listas são verdadeiros catálogos e ainda precisam ser atualizadas e verificadas quanto à adaptação para processamento de dados de audiência de *streaming*.

A pesquisa em *sites* e *blogs* dedicados ao *streaming* data ainda cita: MOA (Massive Online Analysis), Weka, Quadrigam, CartoDB, Tableau Public, Seaborn, R + ggplot2, Cyfe, Databox, Dattawrapper, Easy Insight, Exago, GeoVista, Golden Software's Grapher, Geckoboard, Grow, GoodData, iDashboards, Klipfolio, Pi Platform, Toucan Toco, Visme, Zoho Analytics, RiverML, GAENARI, RapidMiner, ArcGIS, Aspectum, DisplayR, Vega, Vega-Lite, ZingChart (JS), Flourish, Infogram,

¹² Matlab: plataforma de programação e computação numérica usada por engenheiros e cientistas para analisar dados, desenvolver algoritmos e criar modelos. Disponível em: <https://www.mathworks.com/products/matlab-online.html>. Acesso em: 31 jul. 2025.

¹³ Disponível em: <https://giraph.apache.org/>. Acesso em: 01 ago. 2025.

¹⁴ Disponível em: <https://socnetv.org/>. Acesso em: 01 ago. 2025.

¹⁵ Disponível em: <http://www.fmsasg.com/>. Acesso em: 01 ago. 2025.

¹⁶ Disponível em: <https://youtu.be/ZaMF6VNO7tA>. Acesso em: 01 ago. 2025.

¹⁷ Disponível em: <https://www.lirmm.fr/~poncelet/publications/papers/HydroQual.pdf>. Acesso em: 01 ago. 2025.

¹⁸ Disponível em: <https://www.visualisingdata.com/resources/>. Acesso em: 01 ago. 2025.

¹⁹ Disponível em: <https://textvis.lnu.se/>.

²⁰ Disponível em: <https://datavizcatalogue.com/>.

²¹ Disponível em: <https://r-graph-gallery.com/>.

plotDB, Slemma, Pandas, Qulick Sense, Domo, Heatmap.py, gheat, HeatMapAPI.

Ao realizar esta investigação, foi possível observar tendências para a visualização do *streaming* data a partir de quatro variáveis principais, também indicadas pelos dados de recepção. São elas: tempo × comportamento da audiência × conteúdo × localização geográfica. Dessa forma a sugestão é que as pesquisas de *streaming* possam explorar as visualizações utilizando: (a) streamgraphs, (b) heatmaps, (c) wedge stack graphs.

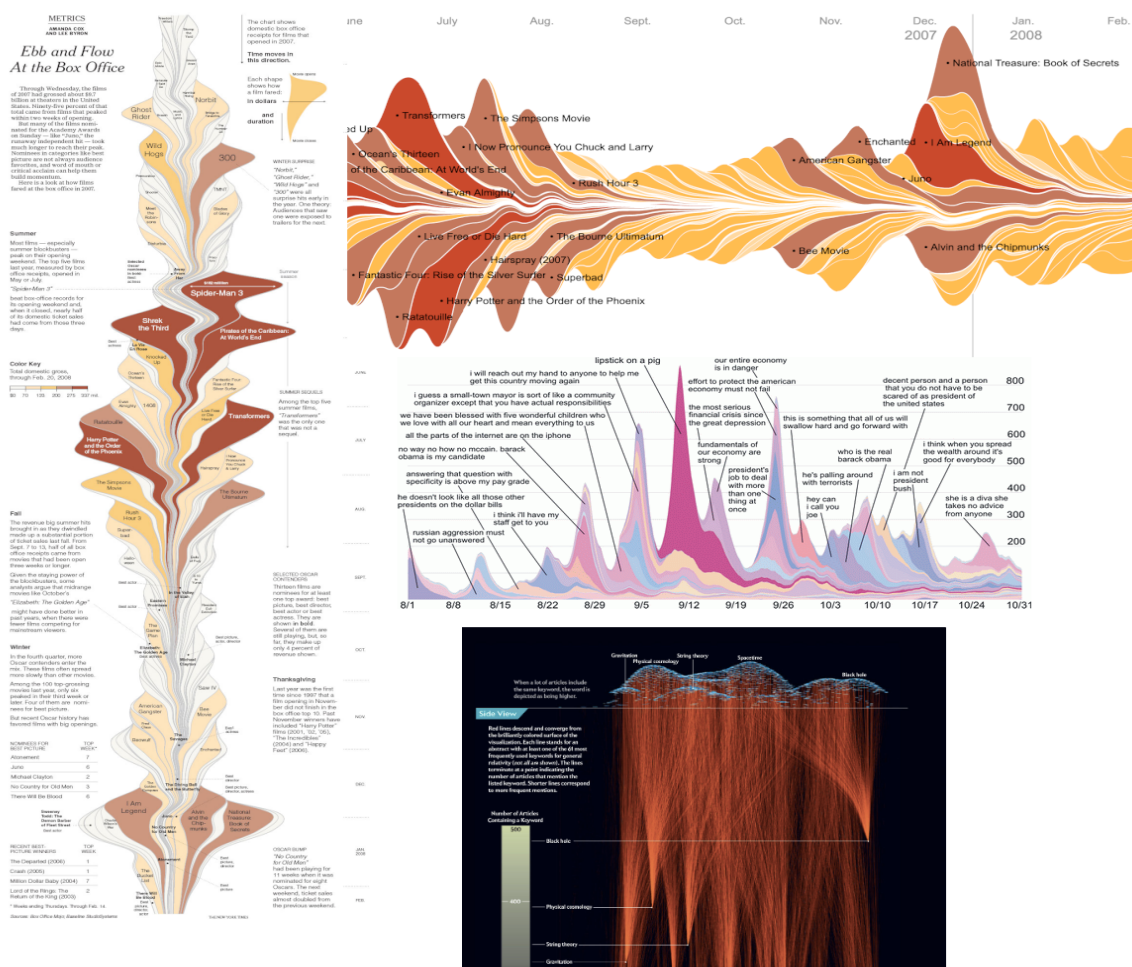
a) Streamgraphs: séries temporais múltiplas e um conjunto de múltiplas variáveis quantitativas que ocorrem no mesmo intervalo. É possível agrupar

séries temporais múltiplas em tempo real a partir de um conjunto de múltiplas variáveis quantitativas que ocorrem no mesmo intervalo de tempo: tempo × comportamento × conteúdo × localização.

A visualização de streamgraphs lembra o movimento de ondas sonoras no decorrer de uma linha do tempo. Esse tipo de visualização de dados é utilizado para fins analíticos em muitos domínios, a exemplo da medicina, finanças e manufatura.

Nos últimos anos, a visualização de streamgraphs tem sido amplamente utilizada para representar padrões de evolução temporal em várias séries de dados, de acordo com Cuenca *et al.* (2018). Seguem exemplos desse tipo de visualização:

Figura 4 - Exemplos de streamgraphs



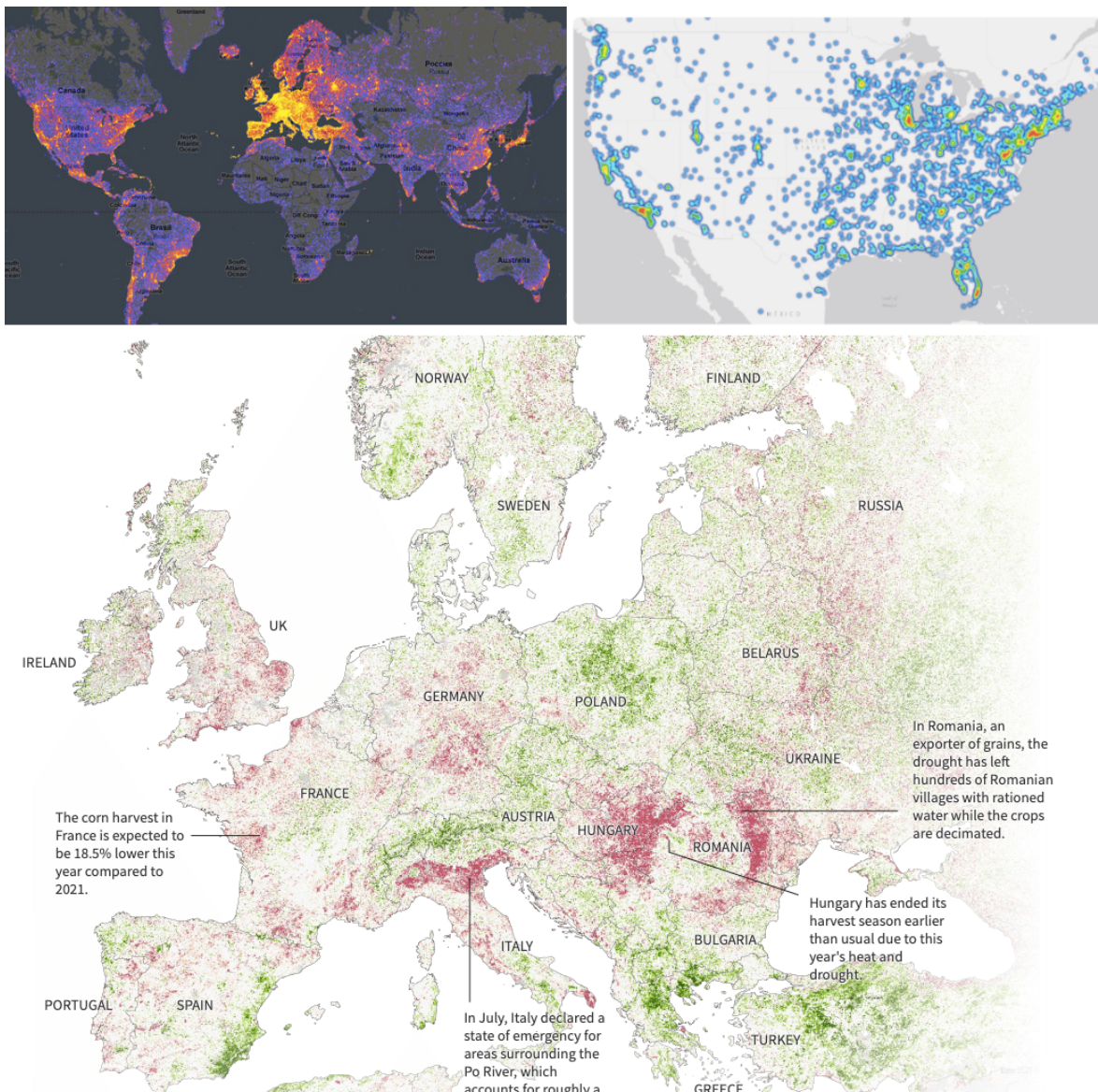
Fonte: Kirk (2010).

b) Heatmaps: um mapa de calor é uma técnica de visualização de dados que mostra a magnitude de um fenômeno. A variação de cor pode ser por matiz ou intensidade, dando pistas visuais óbvias ao leitor sobre como o fenômeno está agrupado ou varia no espaço.

Existem duas categorias fundamentalmente

diferentes de mapas de calor: o mapa de calor de cluster e o mapa de calor espacial. Aplicado aos dados de *streaming*, o heatmaps poderia ser chamado de "Hitmap", ou seja, um mapa que ilustra a intensidade de visualizações de ficções ou blocos temáticos de ficções em um dado espaço geográfico. Exemplos de visualização:

Figura 5 - Exemplos de heatmaps

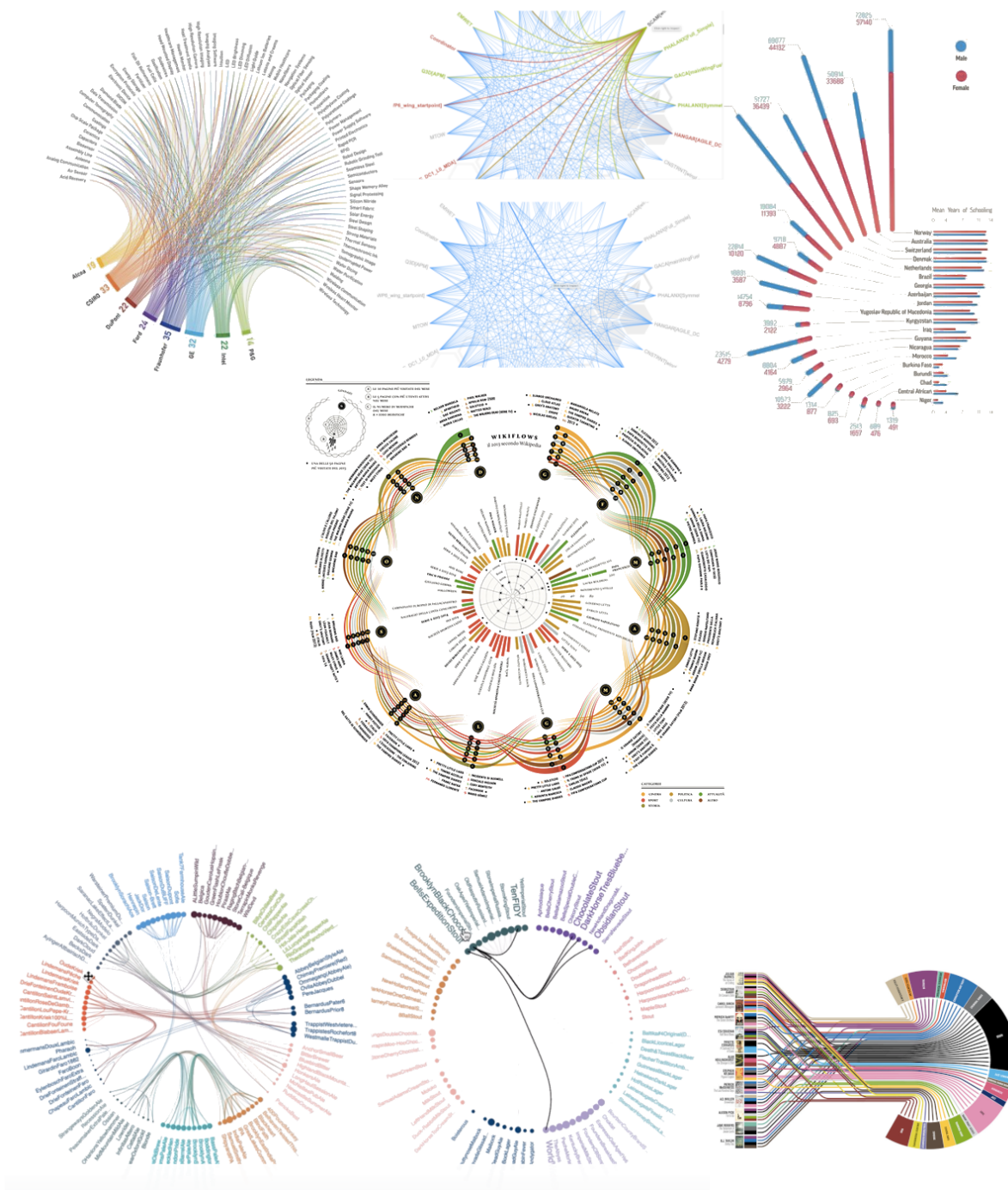


Fonte: A) [BlueMoon.ee](https://www.bluemoon.ee) (c2025). B) Haun e Silvera (2022, p. 3). C) Bhargava e Granados (2022).

c) Wedge Stack Graphs: técnica de visualização que exibe dados hierárquicos em um sistema radial. Pode ilustrar

dados de frequência de vários níveis e também fazer combinados com visualizações de stream graphs. Exemplos:

Figura 6 - Exemplos de wedge stack graphs



Fonte: Ma (2016).

Considerações e conclusões

Aonde vai parar a audiência de *streaming*? A observação sobre o volume de adesões diante da oferta diversificada de serviços parece não ter limites. Mas a popularidade não garante o acesso

completo aos conteúdos ficcionais para assinantes. Um fato pouco considerado nas pesquisas de *streaming* ainda continua a ser as barreiras para a exibição da produção entre países.

O aprimoramento de técnicas que permitem a visualização do comportamento da audiência

constitui-se um desafio diário para as agências de pesquisa, desenvolvimento de soluções e monitoramento.

O tratamento direcionado para a visualização de dados possui grande valor no mercado, contudo, nas produções acadêmicas a visualização parece ainda continuar em segundo plano.

Utilizamos ainda as mesmas ferramentas para apresentar tabelas e indicadores, e parece não haver tempo suficiente para dedicarmos a procurar algo novo e experimentar, testar e ver como fica. É preciso pensar que o trabalho de propor uma visualização de dados torna a pesquisa mais visível nos mecanismos de busca e que o tempo reservado para produzir uma visualização por si já é também uma produção de conhecimento e pode tornar-se um *spin off* relevante das investigações em comunicação.

A proposta de expor os algoritmos e as classificações que determinam as tarefas ou funções na organização dos dados é também um problema atual da Comunicação e deve ser inserida no campo, nas agendas de pesquisa, nas propostas de projetos e disciplinas a fim de trazer à tona aspectos significativos em relação aos algoritmos e como eles atuam no mundo sob a nossa visão de comunicadores.

Referências

ABOUT. *VisualCapitalist*, 2025c. Disponível em: <https://www.visualcapitalist.com/about/>. Acesso em: 28 jul. 2025.

ANDERSON, Chris. The end of theory: the data deluge makes the scientific method obsolete. *Wired*, [s. l.], v. 16, n. 7, p. 16-07, 2008. Disponível em: <https://www.wired.com/2008/06/pb-theory/>. Acesso em: 26 jul. 2022.

BHARGAVA, Aman; GRANADOS, Samuel. Europe's driest summer in 500 years threatens crops, energy production. *Reuters*, 22 ago. 2022. Disponível em: <https://www.reuters.com/graphics/EUROPE-WEATHER/DROUGHT/jnvwnznyvw/>. Acesso em: 11 ago. 23.

BLUEMOON.EE. World map of touristyness. *Information is Beautiful*, c2025. Disponível em: <https://informationisbeautiful.net/2010/world-map-of-touristyness/>. Acesso em: 11 ago. 2023.

BROWNLEE, Jason. *Master machine learning algorithms: discover how they work and implement them from scratch*. Machine Learning Mastery, 2021. Disponível em: <https://machinelearningmastery.com/master-machine-learning-algorithms/>. Acesso em: 28 jul. 2025.

BUDZINSKI, Oliver; GAENSSELE, Sophia; LINDSTÄDT-DREUSICKE, Nadine. The battle of YouTube, TV and Netflix: an empirical analysis of competition in audiovisual media markets. *SN Business & Economics*, [s. l.], v. 1, n. 9, p. 116, 2021. Disponível em: <https://doi.org/10.1007/s43546-021-00122-0>. Acesso em: 28 jul. 2025.

CAIRO, Alberto. The dawn of a philosophy of visualization. In: ENGBRETSSEN, Martin; KENNEDY, Helen. (ed.). *Data visualization in society*. Amsterdã: Amsterdam University Press, 2020. p. 17-18.

CALDAROLA, Enrico G.; RINALDI, Antonio M. *Big data visualization tools: a survey of the new paradigms for methodologies and tools for large data sets visualization*. In: INTERNATIONAL CONFERENCE ON DATA SCIENCE, TECHNOLOGY AND APPLICATIONS, 6., Madrid, 2017. *Proceedings* [...]. Madrid: Scitepress, 2017. p. 296-305. Disponível em: <https://www.scitepress.org/papers/2017/64841/64841.pdf>. Acesso em: 28 jul. 2025.

CONVIVA. *Conviva's State of Streaming Q2 2022*. Foster City: Conviva, setembro de 2022. Disponível em: <https://www.conviva.com/wp-content/uploads/2022/09/Q2>. Acesso em: 4 ago. 2025.

CORMEN, Thomas H. *et al. Introduction to algorithms*. 2. ed. Cambridge: The MIT Press, 2001.

CORRÊA, Elizabeth S.; BERTOCCHI, Daniela. A cena cibercultural do jornalismo contemporâneo: web semântica, algoritmos, aplicativos e curadoria. *Matrizes*, São Paulo, v. 5, p. 123-143, 2012. Disponível em: <https://doi.org/10.11606/issn.1982-8160.v5i2p123-144>. Acesso em: 28 jul. 2025.

CUENCA, Erick *et al.* Multistream: A multiresolution streamgraph approach to explore hierarchical time series. *IEEE Transactions on Visualization and Computer Graphics*, [s. l.], v. 24, n. 12, p. 3160-3173, 2018. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/29994422/>. Acesso em: 28 jul. 2025.

DI BATTISTA, Giuseppe *et al.* Algorithms for drawing graphs: an annotated bibliography. *Computational Geometry*, [s. l.], v. 4, n. 5, p. 235-282, 1994. Disponível em: [https://doi.org/10.1016/0925-7721\(94\)00014-X](https://doi.org/10.1016/0925-7721(94)00014-X). Acesso em: 28 jul. 2025.

DIJKSTRA, Edsger W. A note on two problems in connexion with graphs. *Numerische Mathematik*, [s. l.], v. 1, p. 269-271, 1959. Disponível em: <https://ir.cwi.nl/pub/9256/9256D.pdf>. Acesso em: 4 ago. 2025.

DU, Truman. Netflix vs Disney: who's winning the streaming war? *VisualCapitalist*, 10 out. 2022. Disponível em: <https://www.visualcapitalist.com/cp/netflix-versus-disney-subscribers/>. Acesso em: 2 ago. 2023.

EUROPEAN AUDIOVISUAL OBSERVATORY. *Yearbook 2020/2021 key trends: Television, Cinema, Video and On-Demand Audiovisual Services - The Pan-European Picture*. Strasburg: European Audiovisual Observatory, 2021.

FLOYD, Robert W. Algorithm 97: shortest path. *Communications of the ACM*, [s. l.], jg. 5, n. 6, p. 345-345, 1962. Disponível em: <https://dl.acm.org/doi/10.1145/367766.368168>. Acesso em: 4 ago. 2025.

FRANCIK, Jaroslaw. *Surveillance du flux des données dans l'animation des algorithmes*. Lille: Université des Sciences et Technologies, 1999.

HAUN, Courtney N.; SILVERA, Geoffrey A. A bird in hand: an examination of the influence of nursing school proximal density on hospital quality of care outcomes in U.S. hospitals. *Inquiry*, [s. l.], v. 59, p. 1-9, 2022. Disponível em: <https://doi.org/10.1177/00469580221100166>. Acesso em: 11 ago. 2023.

HILL, Robin K. What an algorithm is. *Philosophy & Technology*, [s. l.], v. 29, n. 1, p. 35-59, 2016. Disponível em: <https://philpapers.org/rec/HILWAA>. Acesso em: 28 jul. 2025.

KEJARIWAL, Arun; KULKARNI, Sanjeev; RAMASAMY, Karthik. Real time analytics: algorithms and systems. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, 41., Kohala, 2015. *Proceedings [...]*. Kohala: VLDB Endowment, 2015. Disponível em: <https://cite-seerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=fb6438dd5255eed62ab00101312aae2058e2e552>. Acesso em: 28 jul. 2025.

KENNEDY, Helen; ENGBRETSSEN, Martin. The relationships between graphs, charts, maps and meanings, feelings, engagements. In: ENGBRETSSEN, Martin; KENNEDY, Helen. (ed.). *Data visualization in society*. Amsterdam: Amsterdam University Press, 2020. p. 19-34.

KENNEDY, Helen; HILL, Rosemary Lucy. The pleasure and pain of visualizing data in times of data power. *Television and New Media*, [s. l.], v. 18, n. 8, p. 769-782, 2017. Disponível em: <https://journals.sagepub.com/doi/10.1177/1527476416667823>. Acesso em: 28 jul. 2025.

KHALID, Zhwan M.; ZEEBAREE, Subhi R. M. Big Data analysis for data visualization: a review. *International Journal of Science and Business*, [s. l.], v. 5, n. 2, p. 64-75, 2021. Disponível em: <https://ijsab.com/wp-content/uploads/671.pdf>. Acesso em: 28 jul. 2025.

KIRK, Andy. Making sense of streamgraphs. *Visualizing Data*, 18 ago. 2010. Disponível em: <https://visualising-data.com/2010/08/making-sense-of-streamgraphs/>. Acesso em: 11 ago. 2023.

KITCHIN, Rob. Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, [s. l.], v. 1, n. 1, p. 1-12, 2014. Disponível em: <https://journals.sagepub.com/doi/10.1177/2053951714528481>. Acesso em: 28 jul. 2025.

KRANEN, Philipp. *Anytime-Algorithmen für Stream-Daten*. 2011. Dissertation (Doktor der Naturwissenschaften) – Fakultät für Mathematik, Informatik und Naturwissenschaften, RWTH Aachen University, Willich, 2011. Disponível em: <https://d-nb.info/1018257942/34>. Acesso em: 15 jul. 2022.

KUCHER, Kostiantyn; KERREN, Andreas. Text visualization techniques: Taxonomy, visual survey, and community insights. In: IEEE PACIFIC VISUALIZATION SYMPOSIUM (PACIFICVIS). *Anais [...]*. Hangzhou: IEEE, 2015. p. 117-121.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. *Nature*, [s. l.], v. 521, n. 7553, p. 436-444, 2015. Disponível em: <https://doi.org/10.1038/nature14539>. Acesso em: 28 jul. 2025.

LOPES, Maria Immacolata Vassalo; ABRÃO, Maria Amélia Paiva. A complexidade da ficção televisiva brasileira: entre o nacional e o internacional. In: LOPES, Maria Immacolata Vassalo; PIÑON, Juan; BURNAY, Catarina D. (coord.). *As produtoras independentes e a internacionalização da ficção televisiva na ibero-américa*. Santiago: Ediciones UC, 2023. p. 75-101.

LOPES, Maria Immacolata Vassalo; PIÑON, Juan; BURNAY, Catarina D. *Obitel 2022: transformações na serialidade da ficção televisiva em tempos de streaming*. Santiago: Universidade Católica de Chile, 2022.

LU, Marcos. Why investors tuned out Netflix. *Visual Capitalist*, 29 abr. 2022. Disponível em: <https://www.visualcapitalist.com/why-investors-tuned-out-netflix/>. Acesso em: 02 ago. 2023.

MA, Joyce. Patent portfolio mapping (data visualization). *Behance*, 7 dez. 2016. Disponível em: [https://www.behance.net/gallery/46088915/Patent-Portfolio-Mapping-\(Data-Visualization\)](https://www.behance.net/gallery/46088915/Patent-Portfolio-Mapping-(Data-Visualization)). Acesso em: 8 ago. 2023.

MADHULATHA, T. Soni. Overview of streaming-data algorithms. *Advanced computing: an international journal (ACIJ)*, [s. l.], v. 2, n. 6, p. 151-160, nov. 2011. Disponível em: <https://arxiv.org/pdf/1203.2000>. Acesso em: 28 jul. 2025.

MANOVICH, Lev. *Cultural analytics*. Cambridge: MIT Press, 2020.

MOSCHOVAKIS, Yiannis N. What is an algorithm? In: ENGQUIST, B.; SCHMID, W. (ed.). *Mathematics unlimited: 2001 and beyond*. Berlin; Heidelberg: Springer, 2001. p. 919-936.

NETFLIX. *What we watched: a netflix engagement report*. Netflix, 12 dez. 2023. Disponível em: <https://about.netflix.com/en/news/what-we-watched-a-netflix-engagement-report>. Acesso em 20 dez. 2023.

NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR (NIC.BR). *Pesquisa sobre o uso da Internet por crianças e adolescentes no Brasil: TIC Kids Online Brasil 2023*. São Paulo: Comitê Gestor da Internet no Brasil, 2024.

OCAMPO, Josh. In Focus, an Actors' Strike and Hollywood's Future. *The New York Times*, 02 ago. 2023. Disponível em: <https://www.nytimes.com/2023/08/02/insider/hollywood-actors-strike.html>. Acesso em: 10 ago. 2023.

OTUKEI, John Richard; BLASCHKE, Thomas. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation*, [s. l.], v. 12, n. 1, p. 27-31, 2010. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0303243409001135>. Acesso em: 28 jul. 2025.

PHAM, Stefan *et al.* Standards-based streaming analytics and its visualization. In: ACM MULTIMEDIA SYSTEMS CONFERENCE, 12., Istanbul, set./out. 2021. *Proceedings [...]*. Istanbul: ACM, 2021. p. 350-355. Disponível em: <https://dl.acm.org/doi/10.1145/3458305.3478438>. Acesso em: 28 jul. 2025.

POUYANFAR, Samira *et al.* A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)*, [s. l.], v. 51, n. 5, p. 1-36, 2018. Disponível em: <https://dl.acm.org/doi/10.1145/3234150>. Acesso em: 28 jul. 2025.

QIN, Xuedi *et al.* Making data visualization more efficient and effective: a survey. *The VLDB Journal*, [s. l.], v. 29, n. 1, p. 93-117, 2020. Disponível em: <https://doi.org/10.1007/s00778-019-00588-3>. Acesso em: 28 jul. 2025.

REGATTIERI, Lorena; MEDEIROS, Jean Maicon; MALINI, Fabio. The use of modularity algorithms as part of the conceptualization of the perspectival form in large networks. Santiago: Datawiz, 2014.

SALLABERRY, Arnaud. *Visualisation de l'information: techniques et solutions visuelles pour explorer les données relationnelles, temporelles et spatiales*. 2020. Tese (Doctorat en Informatique) – Université de Montpellier, Montpellier, 2020. Disponível em: <https://theses.hal.science/tel-03047068>. Acesso em: 28 jul. 2025.

SCHÄFER, Mirko Tobias; VAN ES, Karin. (ed.). *The datafied society: studying culture through data*. Amsterdam: Amsterdam University Press, 2017.

ŠIMOŇÁK, Slavomír. Using algorithm visualizations in computer science education. *Central European Journal of Computer Science*, [s. l.], v. 4, n. 3, p. 183-190, 2014. Disponível em: <https://doi.org/10.2478/s13537-014-0215-4>. Acesso em: 28 jul. 2025.

SHRESTHA, Ajay; MAHMOOD, Ausif. A. Review of deep learning algorithms and architectures. *IEEE Access*, [s. l.], v. 7, p. 53040-53065, 2019. Disponível em: <https://ieeexplore.ieee.org/document/8694781>. Acesso em: 28 jul. 2025.

VALERO MEDINA, José Antonio; ALZATE ATEHORTÚA, Beatriz Elena. Comparison of maximum likelihood, support vector machines, and random forest techniques in satellite images classification. *Tecnura*, v. 23, n. 59, p. 13-26, 2019. Disponível em: <https://doi.org/10.14483/22487638.14826>. Acesso em: 28 jul. 2025.

VAN ES, Karin; WIERINGA, Maranke; SCHÄFER, Mirko Tobias. Tool criticism: From digital methods to digital methodology. In: INTERNATIONAL CONFERENCE ON WEB STUDIES, 2., Paris, 2018. *Proceedings [...]*. Paris: ACM, 2018. p. 24-27. Disponível em: <https://dl.acm.org/doi/10.1145/3240431.3240436>. Acesso em: 28 jul. 2025.

WALLACH, Omri. Which Streaming Service Has the Most Subscriptions? *Visual Capitalist*, march, 3, 2021. Disponível em: <https://www.visualcapitalist.com/which-streaming-service-has-the-most-subscriptions/>. Acesso em 4 ago. 2025.

WU, Aoyu *et al.* AI4VIS: Survey on artificial intelligence approaches for data visualization.

IEEE Transactions on Visualization and Computer Graphics, [s. l.], v. 20, n. 20, p. 1-20, 2021. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/34310306/>. Acesso em: 28 jul. 2025.

YAMAOKA, So *et al.* Cultural analytics in large-scale visualization environments. *IEEE Computer*, [s. l.], v. 44, n. 12, p. 39-48, 2011. Disponível em: <https://www.academia.edu/download/100710259/mc.2011.36320230405-1-ryljq.pdf>. Acesso em: 28 jul. 2025.

Maria Immacolata Vassallo de Lopes

Professora titular sênior da Escola de Comunicações e Artes da Universidade de São Paulo, com Pós-doutorado na Università di Firenze, Itália. Criou em 2005 e coordena o OBITEL (Observatório Ibero-americano da Ficção Televisiva) e OBITEL BRASIL (Rede de pesquisadores brasileiros de ficção televisiva). Coordena o CETVN (Centro de Estudos de Telenovela da ECA-USP). Dirige o periódico MATRIZes (A1). Atua nas áreas de telenovela, ficção televisiva transmídia, *streaming*, de epistemologia e metodologia da comunicação. Tem publicado artigos e livros no país e no exterior em suas especialidades. É pesquisadora 1A do CNPq.

Claudia Freire

Investigadora de Pós-Doutoramento no Centro de Estudos de Comunicação e Cultura - CECC da Universidade Católica Portuguesa de Lisboa. Doutora em Ciências da Comunicação pela Universidade de São Paulo. Mestre em Ciências da Comunicação pela USP. Bacharel em Ciências da Comunicação, Marketing e Propaganda pela Pontifícia Universidade Católica de Minas Gerais. Licenciada em Língua e Literatura Portuguesa pela Universidade Federal de Minas Gerais. Principais Áreas: Epistemologia, Teoria e Metodologia de Investigação em Ciências da Comunicação, Métodos Digitais, Análise de Redes Sociais, Algoritmos e DataViz, Diplomacia Civil.

Endereços para correspondência

Maria Immacolata Vassallo de Lopes

Universidade de São Paulo (USP)
Escola de Comunicações e Artes
Departamento de Comunicações e Artes
Av. Prof. Lúcio Martins Rodrigues, 443
Cidade Universitária, 05508-020
São Paulo, SP, Brasil

Claudia Freire

Universidade Católica Portuguesa
Faculdade de Ciências Humanas
Centro de Estudos de Comunicação e Cultura (CECC)
Palma de Cima, 1649-023
Lisboa, Portugal

Os textos deste artigo foram revisados pela Texto Certo Assessoria Linguística e submetidos para validação dos autores antes da publicação.