

APIs de Visão Computacional: investigando mediações algorítmicas a partir de estudo de bancos de imagens

*Computer Vision APIs: interrogating algorithmic mediations
through stock photos research*

TARCÍZIO SILVA^[1]

Universidade Federal do ABC

ANDRÉ MINTZ^[2]

Universidade Federal de Minas Gerais

JANNA JOCELI OMENA^[3]

Universidade Nova de Lisboa,
iNOVA Media Lab – ICNOVA

BEATRICE GOBBO^[4]

Politecnico di Milano –
DensityDesign Lab

TAÍS OLIVEIRA^[5]

Universidade Federal do ABC

HELEN TATIANA TAKAMITSU^[6]

Universidade Estadual Paulista

ELENA PILIPETS^[7]

Universidade de Klagenfurt

HAMDAN AZHAR^[8]

PRISMOJI

RESUMO

O artigo apresenta resultados de estudo sobre Interfaces de Programação de Aplicações (API, na sigla em inglês) de visão computacional e sua interpretação de representações em bancos de imagens. A visão computacional é um campo das ciências da computação dedicado a desenvolver algoritmos e heurísticas para interpretar dados visuais, mas são ainda incipientes os métodos para sua aplicação ou investigação críticas. O estudo investigou três APIs de visão computacional por meio de sua reapropriação na análise de 16.000 imagens relacionadas a brasileiros, nigerianos, austríacos e portugueses em dois dos maiores bancos de imagens do ocidente. Identificamos que: a) cada API apresenta diferentes modos de etiquetamento das imagens; b) bancos de imagens representam visualidades nacionais com temas recorrentes, mostrando-se úteis para descrever figurações típicas emergentes; c) APIs de visão computacional apresentam diferentes graus de sensibilidade e modos de tratamento de imagens culturalmente específicas.

Palavras-chave: bancos de imagens; métodos digitais; visão computacional

ABSTRACT

The paper presents the results of a study of computer vision Application Programming Interfaces (APIs) and their interpretation of representations in stock images. Computer vision is a field of computer sciences dedicated to the development of algorithms and heuristics for visual data interpretation, but the methods for its critical application or investigation are still incipient. The study inquired into three computer vision APIs through their repurposing in the analysis of 16.000 images related to Brazilians, Nigerians, Austrians and Portuguese, through the search for their demonyms in two of the main Western stock image sites. It indicates that: a) each API presents highly different modes of image labeling; b) stock image sites represent national visualities with recurrent themes, being useful resources for describing typical emergent depictions; c) computer vision APIs present different degrees of sensitivity and modes of treatment of culturally specific images.

Keywords: stock images; digital methods; computer vision

INTRODUÇÃO

O atual panorama da internet e comunicação é caracterizado por uma dualidade entre multiplicação de tecnologias de mídia, que se faz presente no campo da audiência e também da produção (NAPOLI, 2008) e a crescente e controversa plataformização da web e, conseqüentemente, da economia da atenção digital. Se a ideia de liberação do pólo emissor ou relativa descentralização da produção de conteúdo predominou durante a disseminação das tecnologias de informação e comunicação (CASTELLS, 2002; LEMOS, 2002), hoje o panorama comercial demonstra considerável concentração do digital em poucas empresas. O acrônimo GAFAM cita as corporações em torno do Google, Amazon, Facebook, Apple e Microsoft cujo poder de oligopólio é não apenas comercial, mas também sobre a economia da atenção e até mesmo a interpretação da realidade social, uma vez que concentram boa parte dos dados gerados cotidianamente por pessoas e empresas, bem como dos recursos técnicos para interpretá-los.

Helmond (2015) observou a plataformização da web pelas organizações de mídias sociais como um esforço efetivo de gerenciar os fluxos de dados e economia da web na criação e reconfiguração de valores a partir dos rastros digitais e dados. Esta tendência promoveu a integração e interprogramabilidade entre ambientes de maneira hierárquica entre os ambientes web que não podem ser vistos em isolamento, aprofundando a inadequação de métodos de pesquisa tradicionais para seu entendimento (PEARCE et al, 2018). Plataformas como Facebook, por exemplo, avançaram para outros ambientes digitais ao fornecer recursos de autenticação fácil ou comentários distribuídos. A Amazon, por sua vez, foi um dos primeiros casos de e-commerce a aplicar a ideia de *marketplace*: o site se torna um intermediário entre consumidores e lojistas, de diversas escalas, que precisam do website para acessar seus clientes. Mas o Facebook é muito mais do que um site de rede social e a Amazon muito mais do que um site de e-commerce. Estas organizações se tornaram conglomerados em concorrência por áreas de ponta da tecnologia, como robótica e inteligência artificial (IA), impulsionados pelos dados que passam por seus servidores nas transações e comunicações que abrigam. Srnicek (2017) explica a vantagem destas corporações sobre modelos tradicionais de negócio, uma vez que plataformas se posicionam entre usuários e fornecedores e, ao mesmo tempo, dominam o espaço onde as transações ocorrem, tendo acesso privilegiado não só aos dados de empresas que as usam, mas também dados exclusivos que surgem da posição privilegiada de possuir dados de vários concorrentes.

Neste panorama, plataformas de IA surgem tanto dessas novas grandes corporações do mercado digital quanto de empresas que nasceram do mercado de computação tradicional (como Microsoft e IBM). Entre os diferentes serviços ofertados, como o processamento de linguagem natural e recomendações baseadas em padrão de consumo, encontra-se a interpretação

computacional de imagens, uma das fronteiras da IA e uma demanda fundamental do panorama midiático contemporâneo. Nas mídias sociais, por exemplo, imagens têm se tornado cada vez mais centrais na comunicação. A literatura do campo diz, a esse respeito, de uma “virada visual” das plataformas digitais (FAULKNER, VIS & D’ORAZIO, 2018), cujas publicações são cada vez mais visuais, inclusive com mídias sociais voltadas especificamente para este tipo de conteúdo (tais como Instagram, Snapchat e Tiktok).

Este desenvolvimento, contudo, não tem sido acompanhado de forma proporcional pela pesquisa acadêmica, que ainda encontra dificuldades na investigação dessa realidade. A pesquisa centrada em imagens tem sido tipicamente voltada a pequenos estudos qualitativos (LAESTADIUS, 2017), mas o estudo de dados visuais em mídias sociais está se direcionando consideravelmente para abordagens multidimensionais que vêm “dados qualitativos em escala quantitativa” (D’ORAZIO, 2014), devido ao volume, as características da circulação e suas complexidades. Estudar imagens em mídias sociais é especialmente desafiador devido à complexidade do processamento computacional de dados visuais – ao qual se dedica a subdisciplina das Ciências da Computação denominada Visão Computacional. No entanto, trata-se de uma demanda crescente. Algumas plataformas de mídia social, por exemplo, processam através de visão computacional todas as imagens publicadas no site para fins como moderação, segmentação de anúncios e acessibilidade a pessoas com deficiências visuais.

Em certo sentido, muito do trabalho do pesquisador acadêmico interessado em entender a circulação de imagens pode ser realizado e expandido através do uso justamente de etiquetamento automático de conteúdo semântico nas imagens, disponível de forma “empacotada” em bibliotecas de programação ou fornecedores de visão computacional. A maioria das grandes empresas de tecnologia e mídia digital lançaram suas próprias plataformas de IA, que servem a estas finalidades, tais como: IBM Watson, Amazon Web Services, Microsoft Azure e Google Cloud Platform. Entretanto, o conhecimento sobre estas plataformas e seu modo de operação é largamente inexplorado, o que demanda que se reflita não apenas sobre seu potencial para aplicações em pesquisas como, também, que se investigue como elas interpretam os dados visuais processados.

Portanto, o objetivo deste artigo é perseguir esta tarefa dupla de explorar as possibilidades analíticas de APIs de visão computacional – como as supracitadas – ao mesmo passo em que a própria constituição destas plataformas é interrogada^[9]. Nessa abordagem, o estudo se alinha aos “métodos digitais”, segundo proposta de Richard Rogers (2013, 2015), que se define, entre outros aspectos, pela reapropriação (*repurposing*) crítica e reflexiva de instâncias operativas da *web* (objetos nativos do digital) e dos dados gerados pelas plataformas em uma investigação social sensível aos *meios* (cf. OMENA, 2019). Em particular, indagamo-nos acerca dos vieses

inerentes às APIs de visão computacional, que serão avaliados na prática analítica e investigativa. Por tal enquadramento, o estudo mobiliza também elementos da crítica tecnológica de base decolonial que enfatiza que as tecnologias não são apenas coisas neutras mas artefatos culturais incorporados de relações de poder e representação enquanto artefatos mediadores (HAAS, 2003).

■ VISÃO COMPUTACIONAL E O ESTUDO DE IMAGENS

No campo das ciências da computação, denomina-se Visão Computacional a subdisciplina voltada especificamente para a interpretação computacional de dados visuais. Historicamente, trata-se de um dos primeiros problemas propostos, nos anos 1950, para o desenvolvimento da IA, então como uma derivação da cibernética (CARDON, COINETET & MAZIÈRES, 2018). Entre seus primeiros desenvolvimentos encontram-se programas para a reconstituição computacional de espaços e objetos tridimensionais a partir de imagens fotográficas (MANOVICH, 1993; ROBERTS, 1963). Com a elaboração de modelos algorítmicos para a interpretação de imagens, a visão computacional permite incorporar fotografias e vídeos – entre outros tipos de registro – como entrada de dados para a navegação de robôs, ciência forense e sistemas de gestão da informação além, claro, das aplicações bélicas e de vigilância.

De modo importante, contudo, o problema fundamental da visão computacional corresponde ao que, no jargão da área, refere-se como um “problema mal posto”, isto é: um problema para o qual é impossível alcançar uma solução única ou ótima, mas apenas aproximações probabilísticas (SMEULDERS et al., 2000). Longe de definitiva, qualquer interpretação computacional de uma imagem sempre será *uma* interpretação que dirá tanto da imagem analisada quanto – e de modo fundamental – sobre o programa que a produziu.

No contexto de plataformização da *web*, em especial com a crescente centralidade dos conteúdos visuais nas práticas de uso das plataformas, programas de reconhecimento de imagens advindos da visão computacional possuem um papel fundamental. Tendo em vista o caráter constituinte da *datificação* e das *mediações algorítmicas* na própria definição das plataformas de mídia social (D’ANDRÉA, 2018; VAN DIJCK, 2013), os programas de reconhecimento de imagens permitem a integração dos conteúdos visuais à operação das plataformas. Também, é justamente a massiva disponibilidade destes conteúdos que possibilita o desenvolvimento contemporâneo da visão computacional sob o paradigma do aprendizado de máquinas (ALPAYDIN, 2016). Este paradigma se caracteriza pela natureza indutiva de sua operação, em que o algoritmo não é

explicitamente elaborado no programa mas, sim, inferido pelo próprio programa a partir de um grande volume de dados de treinamento (MACKENZIE, 2017).

No entanto, vários problemas emergem desta configuração relativamente recente. Um contra-argumento típico dirigido a defensores do paradigma do aprendizado de máquinas diz respeito a ininteligibilidade do sistema de decisões produzido (CARDON et al., 2018). Após o processo de treinamento, os modelos resultantes oferecem previsões probabilísticas com base nos dados de treinamento, mas torna-se difícil aferir e (mais ainda) intervir na arquitetura do sistema de modo a corrigir eventuais impertinências e vieses. Segundo relatam Cardon et al., este problema da ininteligibilidade foi contrabalançado na controvérsia científica em torno do aprendizado de máquinas pela relativa eficácia então verificada nas aplicações dos sistemas. Esta se devia ao volume de dados utilizado no treinamento que, em teoria, levaria a uma redução de erros e vieses. Porém, como boa parte da crítica contemporânea tem salientado, o volume quantitativo dos dados não reflete necessariamente em sua qualidade para inferir modelos gerais. Dados colhidos da *web* – “*in the wild*” – tendem a reproduzir esquemas de invisibilização e opressão que, por seu caráter sistêmico, não são compensados pelo grande volume de dados usados no treinamento. Pelo contrário, verifica-se que os modelos inferidos a partir destes dados reificam e, ao fazê-lo, reforçam as dinâmicas de invisibilização e marginalização.

Estes programas também têm se tornado ferramentas e objetos de pesquisa nas ciências humanas e sociais, em aplicações com grau crítico variável. Trabalhos recentes os têm mobilizado como ferramentas para compreender tendências culturais e comportamentais em grandes bases de dados e, também, para mapear as diferenças entre sistemas de reconhecimento de imagens, como vieses e preconceitos cristalizados nas bases de dados, modelos e algoritmos. Considerando o primeiro caso, exemplos incluem o mapeamento semântico de cidades (RICCI et al., 2017; RYKOV et al., 2016); estudos comparativos de culturas dos *selfies* (TIFENTALE & MANOVICH, 2015); análise de persuasão visual (HUSSAIN, 2017; JOO et al., 2014); descrição de *home styles* na comunicação política (ANASTASOPOULOUS et al., 2016); classificação de propaganda política (QI et. al, 2016); estudo de modos de engajamento político *online* (OMENA, RABELLO & MINTZ, 2017); bem como de dinâmicas de circulação de imagens (OMENA et al, 2019; D’ANDREA & MINTZ, 2019; SILVA, BARCIELA & MEIRELLES, 2018). Já para o estudo de vieses e diferenças entre sistemas de aprendizado de máquinas, abordagens voltam-se a suas implicações de gênero (HENDRICKS et al., 2018) e raça (BUOLAMWINI, 2017). Alguns destes estudos têm resultado em informações de impacto como guias de boas práticas (OSOBA & WELSER IV, 2017) ou uma proposta para bases de dados cuidadosamente elaboradas para aprendizado de máquina (BUOLAMWINI & GEBRU, 2018).

■ APIS DE VISÃO COMPUTACIONAL: O QUE SÃO?

Interface de Programação de Aplicações (API na sigla em inglês), denomina um modo de estruturação de programas computacionais que permite sua interoperabilidade com outros sistemas. Por meio de uma API, um programa de computador pode ser projetado de modo a empacotar determinadas funções e recursos de dados para que sejam acessados por um programa externo. No contexto da computação distribuída e da plataforma da *web*, APIs permitem a disponibilidade pública ou comercial de serviços computacionais e dados. Na prática, significa que milhares de indivíduos ou empresas podem usar de forma automática e padronizada, geralmente por pagamento baseado em demanda, os serviços digitais e dados de uma empresa fornecedora.

Na área do fornecimento de recursos de Visão Computacional, se destacam empresas como IBM, Microsoft, Google, Amazon e Clarifai, merecendo destaque também os recursos de nicho TinEye e Kairos. Entre as funções desempenhadas, oferece-se, por exemplo, a classificação de imagens, o reconhecimento facial e reconhecimento óptico de caracteres. Disponibilizadas nos últimos anos por grandes empresas do Vale do Silício, estas APIs constituem hoje um modelo plataforma de visão computacional que tem se difundido com aplicações em áreas diversas – inclusive na pesquisa de plataformas de mídia social^[10].

Para exemplificar um recurso típico destas ferramentas, podemos ver na Figura 1 uma demonstração do recurso do IBM Watson Visual Recognition. No modelo pré-treinado do fornecedor, o sistema consegue identificar objetos e categorias como tecido (*fabric*), cor cinza (*gray color*), tipos de tecido e outras entidades com diferentes graus de confiança. Com dezenas de milhares de etiquetas e classes já pré-treinadas, o recurso pode ser aplicado imediatamente por desenvolvedores para diferentes fins. Entre os estudos de caso fornecidos pela IBM está o uso do recurso para identificar rapidamente tipos de danos em veículos, por exemplo.

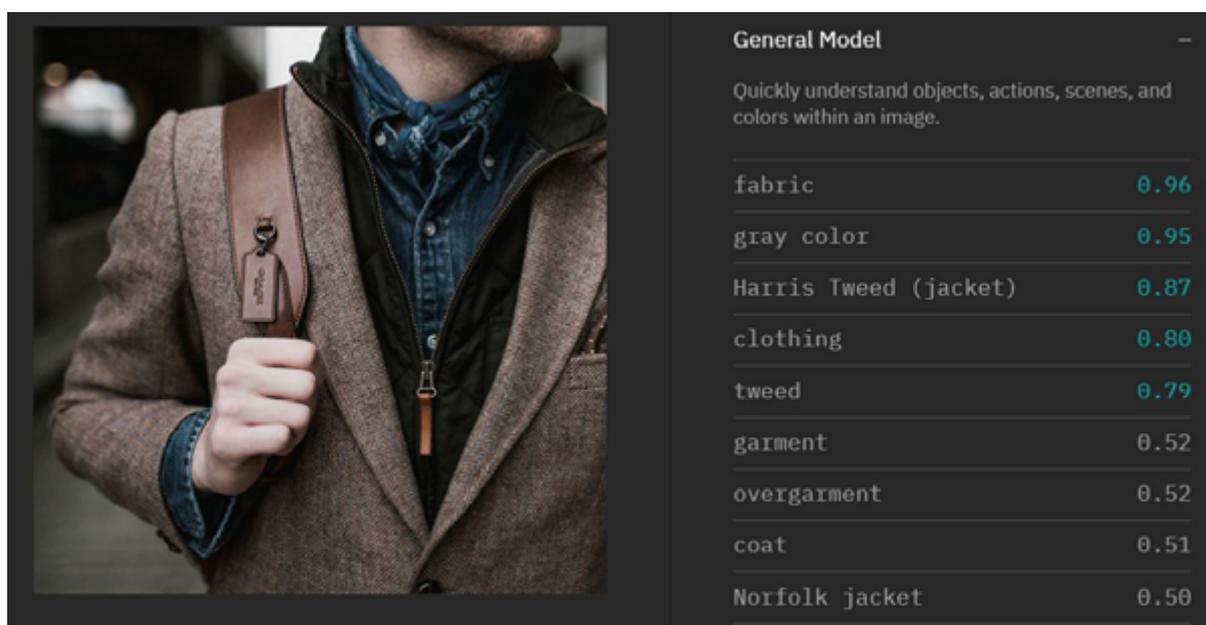


FIGURA 1: Captura de tela da demo do Watson Visual Recognition no site oficial da empresa

As APIs de visão computacional chamam a identificação destas entidades de etiquetas, tags ou classes, a depender da organização de cada uma delas. Mas este é apenas um dos grupos de recursos oferecidos. São crescentes as opções disponíveis, mas podemos destacar cinco outros grupos: a) reconhecimento automatizado do texto nas imagens, transformando o recurso visual em textual; b) recursos ligados à identificação de pessoas e suas características, como idade, gênero e expressões faciais; c) descoberta de imagens equivalentes ou similares na web, assim como extração de informações relacionadas da web semântica; d) modelos verticais, como reconhecimento de celebridades, pontos turísticos ou tipos de comida; e e) detecção de conteúdo explícito, como violência e pornografia, aplicado para fins de moderação e filtragem de conteúdo.

Resumimos na Tabela 1 os principais recursos disponíveis de forma explícita pelas APIs da Google, IBM e Microsoft, fornecedores com os quais nos engajamos mais diretamente neste estudo. Como é possível ver abaixo, alguns recursos são exclusivos – a Google, por exemplo, integra seus recursos com os buscadores, permitindo tanto busca reversa da imagem na web quanto a extração de informações ligadas a estas imagens pela varredura dos websites onde são empregadas.

TABELA 1: Comparação de alguns dos recursos de APIs de Visão Computacional

| | GOOGLE | IBM | MICROSOFT |
|-------------------------------|--------|-----|-----------|
| Etiquetas / Tags / Classes | Sim | Sim | Sim |
| Entidades de Web Semântica | Sim | Não | Não |
| Classes de comida | Não | Sim | Não |
| Legendas automáticas | Não | Não | Sim |
| Deteção de conteúdo explícito | Sim | Sim | Sim |
| Deteção de rostos | Sim | Não | Sim |
| Expressões faciais | Sim | Não | Não |
| Celebridades | Não | Não | Sim |
| Pontos turísticos / Locais | Sim | Não | Sim |
| Gênero | Não | Sim | Sim |
| Idade | Não | Sim | Sim |
| Reconhecimento de texto | Sim | Não | Sim |
| Idioma do texto | Sim | Não | Não |
| Busca reversa na web | Sim | Não | Não |

Os recursos citados são amplamente disponíveis e empregados por organizações e setores comerciais ou governamentais. Entre estas aplicações, as colaborações de grandes empresas de tecnologia com projetos bélicos e repressivos estadunidenses renderam protestos por parte de seus funcionários. Em 2018, funcionários da Google realizaram paralisações contra o uso das tecnologias de IA da empresa na otimização de ataques por drones^[11]. Em acontecimento indicativo do combate à discussão sobre ética da inteligência artificial, uma das funcionárias engajadas no protesto, Meredith Whittaker, deixou a empresa após ser pressionada a se afastar da pesquisa em ética e tecnologia na Universidade de Nova York^[12]. Em 2019, funcionários da Amazon também estão buscando impedir que a empresa apoie o setor de imigração responsável por rastrear imigrantes não-documentados nos EUA e pelo seu aprisionamento em instituições análogas a campos de concentração na região da fronteira^[13].

Na interface com os métodos digitais de pesquisa, APIs de visão computacional têm sido empregados em prol de desenhos criativos de investigação e também para a própria crítica reflexiva destes recursos. Podemos compreender estas iniciativas em relação ao proposto por Farida Vis (2013) de questionamento sobre outros modos possíveis de pensar os dados coletados ou processados através de recursos computacionais como IA e Big Data, para imaginar outras aplicações possíveis.

INTERROGANDO APIS E BANCOS DE IMAGENS: AUSÊNCIAS E HIPER-VISIBILIDADES

Retomando o caráter complexo dos ecossistemas de mídia citados na introdução, é preciso destacar o papel dos bancos comerciais de imagens. Empresas de bancos de imagens existem desde os anos 1920s (BRUHN, 2003), fornecendo fotografias para empresas de mídia, editoras, agências de publicidade e afins sendo usadas por clientes produtores em diferentes formatos e suportes como revistas, jornais, mídia *out-of-home* e embalagens. No século XXI, com a internet e o barateamento da produção fotográfica, o formato *microstock* se popularizou, alargando o tamanho das redes de consumidores e produtores. O chamado *microstock* é um modelo através do qual as empresas de bancos de imagens tornam-se uma interface entre fotógrafos e estúdios profissionais ou amadores que podem vender suas fotografias individualmente e a baixo custo para todos os tipos de clientes.

A análise cultural do papel representacional exercido pelos bancos de imagens é relativamente recente. Frosh (2001) discute os diferentes significados e camadas de referências para produtores, compradores e consumidores finais de mídia. Um conceito crucial evocado pelo autor é a invisibilidade: a maior parte dos consumidores de mídia não conhece o conceito de bancos de imagens, apenas consome e interpreta a visualidade em sua estética e conteúdos representativos nos contextos finais.

Essa invisibilidade também se apresenta parcialmente no campo acadêmico. Trabalhos científicos sobre a observação e análise crítica de repositórios de imagens são escassos, especialmente quando falamos sobre métodos digitais e computação social. Entre as investigações mais recentes sobre produção e circulação de bancos de imagens, podemos citar o trabalho de Pritchard e Whiting (2015) sobre gênero e envelhecimento; artigo de Giorgia Aiello e Anna Woodhouse (2016) sobre estereotipização de gênero; relatórios de *sprints* coordenado por Aiello

(2016, 2017) sobre bancos de imagens em relação com fotojornalismo e representações de gênero; estudo de West sobre estereotipização de gênero e invisibilidade de classe em imagens representativas de ambientes de trabalho (WEST, 2018); e método para incrementar equidade em sistemas de recomendação de imagens (KARAKO & MANGGALA, 2018). Nossa proposta é contribuir a este conjunto de experimentos sobre bancos de imagens, através de lentes dos métodos digitais e APIs de visão computacional – ao mesmo tempo em que investigamos criticamente estas ferramentas.

O projeto buscou investigar APIs de Visão Computacional como mecanismos de pesquisa enquanto os aplica experimentalmente ao estudo de representação visual de diferentes países através das imagens resultantes da busca por seus gentílicos (especificamente: brasileiros, portugueses, nigerianos e austríacos). Essa abordagem dual é baseada na inflexão proposta por Noortje Marres e David Moats (2015) para o princípio fundante da *simetria* no âmbito dos Estudos Sociais de Ciência e Tecnologia (STS, na sigla em inglês). Em considerações voltadas para o estudo de controvérsias *online*, os autores defendem a necessidade de uma abordagem simétrica do 'conteúdo' das controvérsias e também dos meios em que elas se manifestam. No caso deste estudo, a simetria se dá entre as representações dos bancos de imagem e as mediações algorítmicas das imagens em redes digitais, mobilizadas também como ferramenta metodológica da pesquisa. Para este fim, o estudo é baseado na abordagem de grandes *datasets* visuais coletados de websites de bancos de imagens (Shutterstock e Adobe Stock), com o objetivo de derivar descrições e identificar padrões de comparação e, ao mesmo tempo, desenvolver considerações sobre os potenciais e limites dessa aplicação metodológica.

Deste modo, a delimitação de questões de pesquisa pertinentes à tecnologia da visão computacional e, ao mesmo tempo, aos bancos de imagens, convergiu nas seguintes perguntas:

- Quais as diferenças entre os provedores de APIs de visão computacional?
- Como APIs de visão computacional "entendem" as mesmas fotos?
- Como se distinguem as ontologias das etiquetas de cada API?
- Podemos investigar representações nacionais com ferramentas de visão computacional?
- Como bancos de imagens representam a visualidade dos países analisados?
- Como as especificidades culturais são tornadas visíveis através do uso de APIs de visão computacional?

Para responder a estas questões, elaboramos um protocolo e análise baseado em métodos mistos. Um primeiro passo envolveu a coleta de dados nos bancos de imagens, percorreu o processamento destas imagens por diferentes APIs de visão computacional e, finalmente, desembocou na análise destes dados por meio de uma variedade de abordagens e técnicas direcionadas a diferentes aspectos das questões postas. Tais especificidades serão explicitadas ao longo da análise.

O protocolo de pesquisa engendra uma diversidade de pontos de vista que permite o isolamento de variáveis para ao menos três eixos de investigação comparativa: entre APIs de visão computacional, sites de bancos de imagens e nacionalidades (Figura 2). Optamos neste estudo por focar as APIs de visão computacional, como já destacado, e também as diferentes nacionalidades, dado nosso interesse em especificidades culturais das APIs. Diferentes bancos de imagens são interessantes para diversificar os vieses de cada nacionalidade através de diferentes fontes, mas este aspecto não é tão central no escopo do estudo. Portanto, os bancos de imagem são mobilizados de modo indiferenciado nas análises.

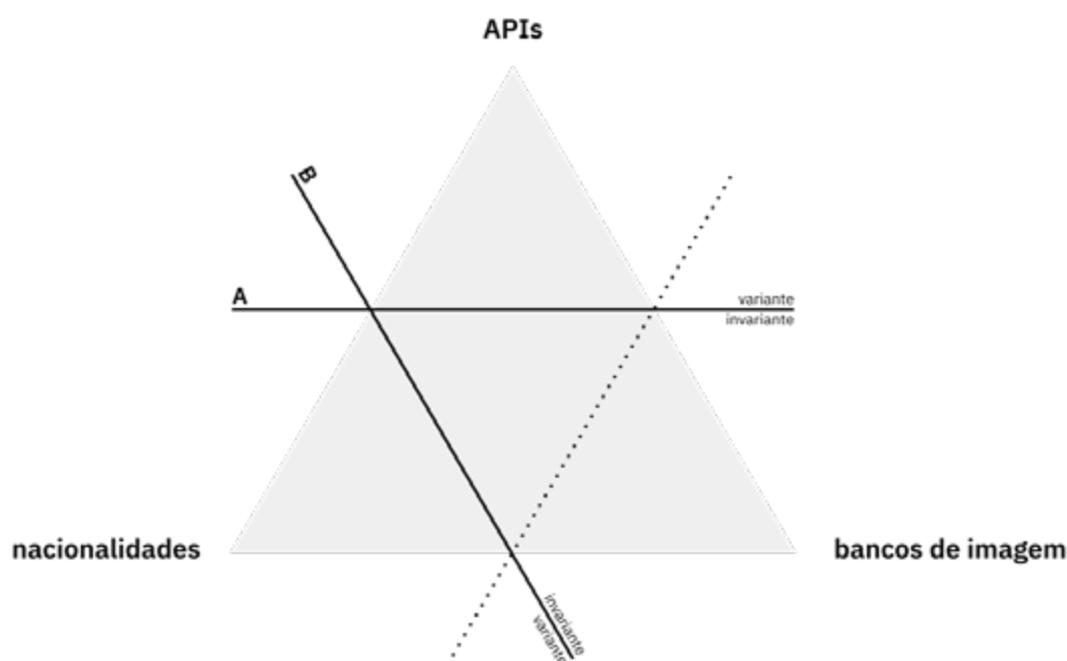


FIGURA 2: Eixos comparativos da análise

A coleta de dados foi realizada através de técnica de raspagem, em que o código-fonte das páginas *web* dos bancos de imagem foi processado a fim de encontrar as URLs das imagens. As imagens em suas versões de pré-visualização (*thumbnail*) foram então carregadas localmente para processamento. Utilizamos um script em Python desenvolvido especificamente com este fim (MINTZ, 2019). Para cada gentílico, coletamos 2000 imagens para cada banco de imagem (Figura 3).

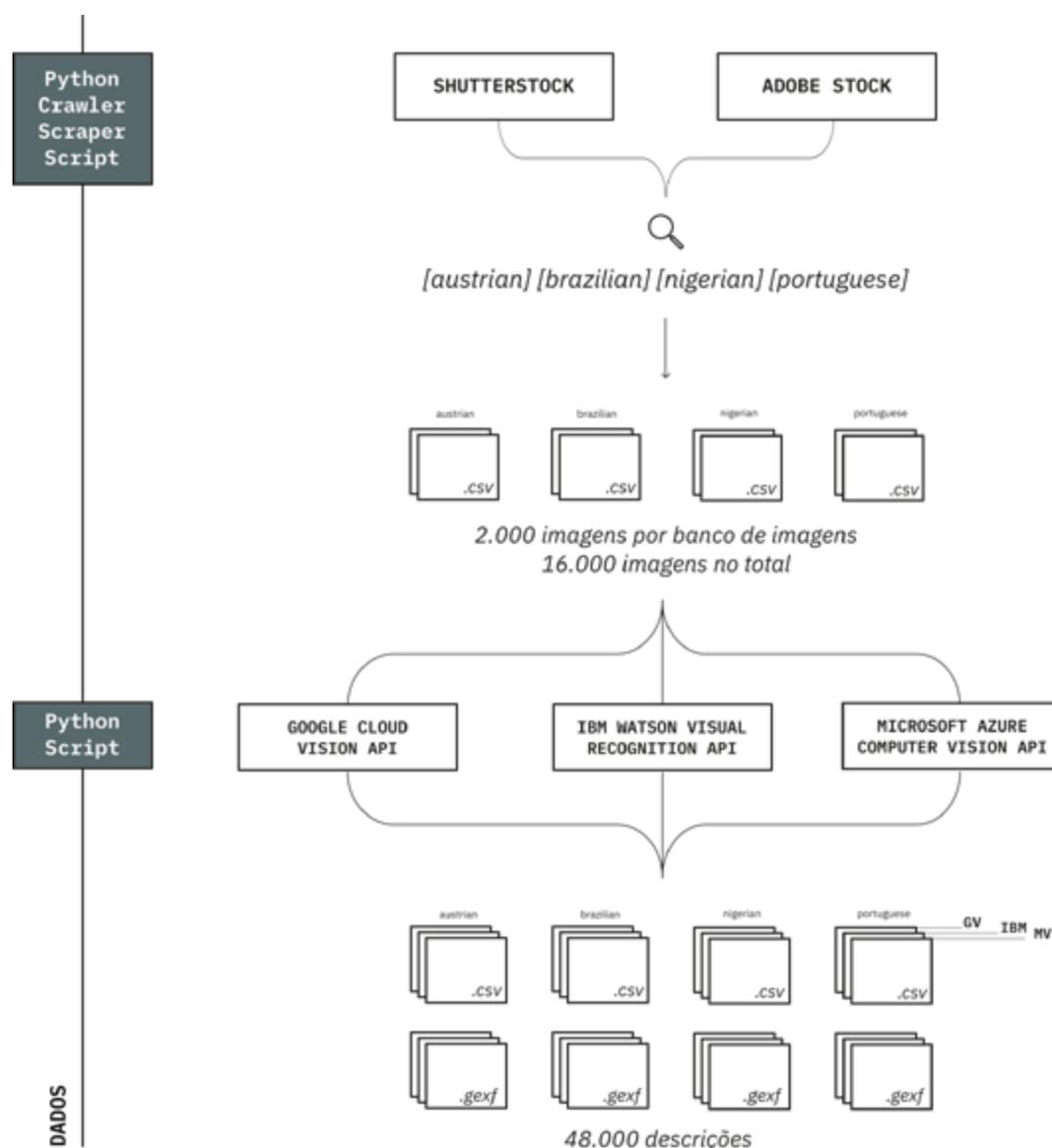


FIGURA 3: Protocolo de coleta de dados

Como citamos na seção anterior, foram selecionados três fornecedores de visão computacional para o estudo: Google Cloud Vision API, Microsoft Azure Computer Vision API e IBM Watson Visual Recognition API (ver Tabela 1). No escopo do estudo, as imagens foram processadas através de seus respectivos módulos que fornecem tags ou etiquetas descritivas – Label Detection, para a Google; Image Classification para a IBM; e Tags da Microsoft.

As imagens foram submetidas em lote às APIs usando um script em *Python* chamado *Data Inspector* (GUERIN JR. & SILVA, 2018), por sua vez baseado no script *Memespector*. O script *Memespector* foi originalmente desenvolvido em PHP por Bernhard Rieder (RIEDER, DEN TEX & MINTZ, 2018) e em seguida adaptado para Python e expandido por André Mintz (2018b). A lógica

de operação é comum a todos estes programas: uma lista de imagens (composta de URLs e outros dados tabulares no formato CSV, *comma-separated values*) é tomada como *input*, e cada imagem é então carregada e processada através das APIs, retornando como resultado os dados tabulares anotados (no formato CSV) e também dados relacionais, ou grafos (no formato GEXF). Deste modo, os conjuntos de dados processados resultantes puderam ser interrogados através de técnicas de pesquisa provenientes da estatística, linguística de corpus, análise de redes sociais e análise visual crítica-exploratória. Entre os resultados, queremos agrupar as observações em dois pontos, respectivamente ligados a práticas de (hiper-)visibilidade e ausência.

GRANULARIDADE E PADRONIZAÇÃO NOS ESPAÇOS SEMÂNTICOS DAS APIS

Como vimos no exemplo da Figura 1, cada imagem processada nos recursos de visão computacional retorna pelo sistema a marcação de termos que referenciam objetos, estados, qualidades ou características. Um primeiro desafio, que diz muito às nossas questões de pesquisa, é o fato de que os fornecedores destes recursos não compartilham a lista completa das etiquetas possíveis. Assim não há a possibilidade de se analisar a capacidade do recurso de entender um determinado grupo semântico antes de se realizar testes empíricos.

TABELA 2: Número de etiquetas atribuídas por cada API de visão computacional a cada um dos datasets estudados (n = 4.000).

| | AUSTRIAN | BRAZILIAN | NIGERIAN | PORTUGUESE |
|-------------------------------------|----------|-----------|----------|------------|
| Microsoft Azure Computer Vision API | 317 | 561 | 485 | 501 |
| IBM Watson Visual Recognition API | 1.632 | 2.044 | 1.846 | 1.991 |
| Google Cloud Vision API | 2.037 | 2.170 | 1.145 | 1.992 |

Considerando que um olhar essencialista é, por definição, padronizador e simplificador, uma primeira questão é: até que ponto os fornecedores de visão computacional oferecem um número efetivamente relevante de tags que se aproximem em determinado grau da complexidade das cenas fotografadas? Na Tabela 2 podemos ver o número de etiquetas únicas atribuídas a cada dataset, comparando fornecedores e conjunto de dados de cada país. A diferença mais significativa está no número de tags da Microsoft, muito atrás de IBM e Google. Para além dessa visada quantitativa, parece relevante também considerar as relações estabelecidas entre as etiquetas e os tipos de agrupamentos formados por sua co-ocorrência.

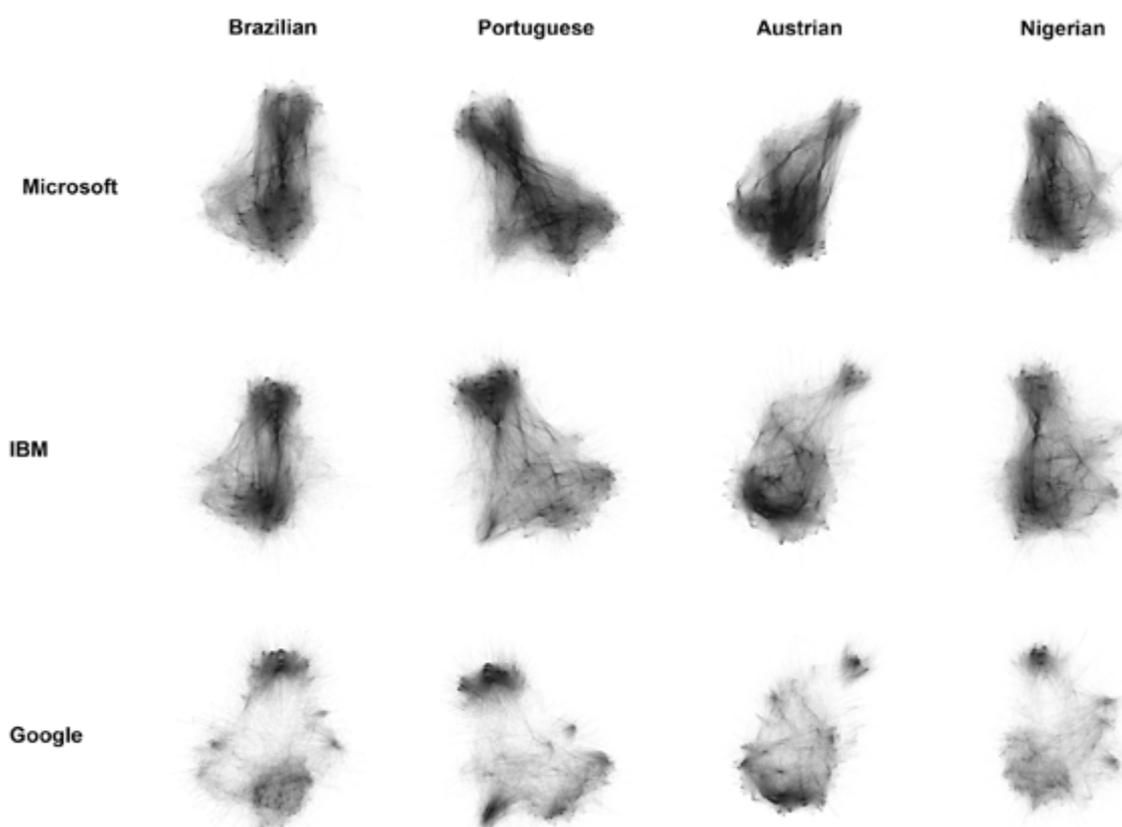


FIGURA 4: Matriz comparativa das redes bimodais imagem-etiqueta para cada país e cada API.

A Figura 4 exibe um conjunto de visualizações que auxilia a compreender a topologia das relações entre etiquetas atribuídas pelas APIs para cada *dataset* analisado. A esta topologia nos referimos por *espaços semânticos*. Na matriz comparativa, as redes relativas a cada *dataset* (colunas) foram inicialmente compostas mesclando as etiquetas das três APIs em uma mesma rede, então submetida ao algoritmo de espacialização ForceAtlas2. Então, para compor a matriz, essa rede combinada foi filtrada para cada API, de modo a ser possível comparar a distribuição das etiquetas e das relações nesta espacialização conjunta. Percebe-se, assim, como o baixo número de etiquetas da API da Microsoft se reflete em redes muito conectadas e, portanto, com pouca diferenciação topológica entre os clusters. Nota-se, também, como embora quantitativamente IBM e Google retornem números comparáveis de etiquetas, a API da IBM traça um espaço semântico mais indiferenciado que a da Google. No caso desta última, a densidade de conexões é distribuída de forma menos difusa pela rede e tende a se condensar em *clusters* específicos, sugerindo um maior grau de especialização do etiquetamento. Veremos na sequência como esta característica geral é performada em casos específicos.

Buscando entender qualitativamente a sensibilidade das APIs a imagens culturalmente específicas, a análise incluiu o exame de pequenos grupos de imagens ou imagens individuais e como as APIs as etiquetam. Enquanto a análise, em última medida, se desdobra na avaliação das etiquetas aplicadas à imagem sob a luz do conhecimento contextual dos países e culturas sendo representadas, esta abordagem envolve o desafio de delimitar, a partir dos grandes conjuntos de dados, estes casos específicos e quantitativamente menores. A filtragem de ocorrências de etiquetas específicas permite analisar pequenos grupos de imagens ligadas por etiquetas relacionadas, através de métodos exploratórios mistos.

As redes bi-modais de imagens e etiquetas (que veremos em detalhe na seção a seguir) compuseram dispositivos exploratórios que permitiram o exame dos padrões visuais emergentes entre os conjuntos de dados assim como possíveis classificações errôneas ou exceções. O software de dados tabulares Google Sheet foi também empregado devido a sua funcionalidade de pré-visualização de imagens com a sua função `IMAGE()`. Tornou-se assim possível filtrar os conjuntos de dados de acordo com etiquetas particulares para escrutinar as imagens nas quais elas foram aplicadas pela API. Este redirecionamento da análise pode ser aproximado à engenharia reversa das APIs, dado que ao tomar as etiquetas como ponto de partida é possível inferir aspectos dos dados de treinamento que alimentaram os algoritmos e seus possíveis vieses.

Identificar alimentos, pratos e culinária é algo que os fornecedores de visão computacional incluem e, como veremos de fato, conseguem realizar em grande medida. Mas será que o reconhecimento de especificidades regionais é suficiente? Como um dos exemplos, podemos falar do Pastel de Nata (Figura 5), um famoso doce português que aparece de modo abundante no dataset relacionado ao país. Considerando as etiquetas atribuídas a um grupo de três imagens que encontramos com pastéis de nata, encontramos um caso que demonstra o baixo grau de especificidade cultural das APIs. Diante dos resultados, fica claro que a Google apresentou os resultados mais acurados, chegando a atribuir a etiqueta "pastel". Além dela, identificou também "custard tart", um nome genérico à receita que também se relaciona a doces similares da Inglaterra e França. A IBM Watson atribuiu etiquetas como "brioche" e "Yorkshire pudding", que não são o mesmo doce, mas que possuem formato similar ao Pastel de Nata. A API da Microsoft usou as descrições "doughnut" e "donut", que são muito diferentes se comparados ao alimento português e tipicamente relacionados à cultura estadunidense.

| Imagens | APIs | Etiquetas |
|---|-----------|---|
|  | Google | Dish, Food, Egg tart , Baked goods, Dessert, Custard tart , Pastel , Pastry, Scotch pie , Finger food, Yorkshire pudding , Custard tart |
|  | IBM | cheese souffle , dish, nutrition, food, food product, cinnamon roll , sweet roll , bun, bread, honey bun , brioche |
|  | Microsoft | plate, food, doughnut , topped, fruit, donut , oranges, cake |

FIGURA 5: Etiquetas atribuídas a três imagens de pasteis de nata pelas APIs. Etiquetas não relacionadas a alimentos foram filtradas para facilitar a leitura.

Outro tema abordado no estudo concerne como APIs de visão computacional etiquetam características fenotípicas não-brancas e acessórios não-ocidentais. Os casos em questão emergem de uma avaliação exploratória da representação de povos indígenas e negros nos conjuntos de dados das fotos sobre brasileiros e nigerianos e seu etiquetamento pelas APIs. Durante a navegação através de fotos de pessoas negras, uma observação emergiu sobre como cabelo e acessórios foram etiquetados, particularmente pela API da Google Cloud Vision.

Descobrimos que a etiqueta “peruca” (“wig”) foi consistentemente atribuída a mulheres negras com cabelo crespo ou usando turbantes, além da atribuição da etiqueta “Dastar” em turbantes de brasileiras da região da Bahia, quando na verdade o Dastar é um turbante especificamente indiano usado por adeptos do Sikhismo. Também sobre turbantes, a etiqueta “acessório de moda” foi frequentemente atribuída a vestimentas no conjunto de dados da Nigéria. Isto, de certa forma, simplifica a carga simbólica religiosa e tradicional contidas nos itens. No caso dos turbantes das baianas brasileiras, as etiquetas “Tradition” e “Ceremony” apareceram mais frequentemente, aproximando das características religiosas.



FIGURA 6: Etiquetas atribuídas a imagens de mulheres negras com cabelo crespo e usando turbantes.

Esta análise nos mostra algumas pistas sobre as limitações das APIs no sentido cultural de suas etiquetas e nos leva a questões sobre, por exemplo, quais das variações de tipos de cabelos são incluídas na API e porque não há uma etiqueta “cabelo crespo” nas fotos de mulheres negras. Essa dualidade entre visibilidade e hiper-visibilidade tem sido explorada por pesquisadoras das implicações de tecnologias de IA e da visão computacional, em particular, para questões raciais e de gênero (BUOLAMWINI, 2017; BUOLAMWINI & GEBRU, 2018; NOBLE, 2018). Um aspecto salientado pelo trabalho destas pesquisadoras é o de como estas tecnologias devem ser compreendidas em vista do contexto em que operam, refletindo não apenas vieses contidos nos *datasets* mas, também, formas de opressão e exclusão que condicionaram a constituição tanto dos *datasets* quanto do software, incluindo aí a pouca diversidade das equipes de desenvolvimento.

A ideia de uma computação pós-colonial é revisada por Ali (2016) ao discorrer como o exame de relações culturais e de poder na computação, interação humano-computador e tecnologias de comunicação e informação tem sido realizada. A virada decolonial (QUIJANO, 2010), porém, assume que apesar dos desenhos tradicionais do colonialismo terem sido suplantados, persiste “uma herança continuada do colonialismo em sociedades contemporâneas na forma da discriminação social” e “práticas e heranças do colonialismo europeu em ordens sociais e

formas de conhecimento”^[14] (ALI, 2016, p. 4). Considerando a abissal lacuna entre, de um lado, corporações como Google, IBM e Microsoft, e desenvolvedores e comunicadores das periferias de outro, no que tange os potenciais de criação de bancos de dados e ontologias de visão computacional, as pressões de custo-benefício, eficácia e efeito de rede tendem a concentrar projetos em torno daqueles fornecedores. As práticas de padronização de conteúdo e ferramentas estariam no cerne das ciências da computação em si, que em sua natureza “obriga a assumir uma perspectiva do paradigma positivista, além de seu enfoque empírico analítico próprio às ciências experimentais”^[15] (PORTILLA, 2013, p. 98).

REDES DE ESPAÇOS SEMÂNTICOS E TIPICIDADE

Com o objetivo de analisar tanto a etiquetagem performada por cada API quanto as características visuais dos dados abordados, o estudo recorreu a visualização de redes bimodais ao processar a atribuição das etiquetas às imagens como dados relacionais. Trata-se de uma abordagem comum em pesquisas similares e que encontra-se programada na versão em Python do script Memespector (Mintz, 2018b). As redes de imagem-etiqueta são geradas ao se considerar as imagens e as etiquetas/tags descritivas como dois tipos de nós. Os laços representam a atribuição de uma etiqueta a uma imagem.

Esta representação dos dados permite seu processamento e visualização através de softwares como o Gephi (GEPHI, 2017), que fornece diversas ferramentas de análise, de algoritmos de layout a módulos estatísticos baseados no framework metodológico de teoria dos grafos. Para este estudo, a análise apoiou-se principalmente em análise visual de redes (GRANDJEAN & JACOMY, 2019; VENTURINI, JACOMY & JENSEN, 2019), que foca em descrever propriedades dos conjuntos de dados de acordo com traços topológicos das redes baseados na posição, tamanho e codificação de cor dos nós. O ForceAtlas2 (JACOMY et al., 2014) foi o principal algoritmo de layout usado na espacialização da rede e o cálculo de modularidade (BLONDEL et al., 2008) foi aplicado para identificar os principais *clusters*. A modularidade divide a rede em seções de acordo com sua possível estrutura de comunidades, atribuindo códigos a cada partição. Dependendo dos dados representados, estas partições podem levar o pesquisador a descobrir agrupamentos significativos, como grupos temáticos, geográficos ou conceitos semânticos gerais. A exploração das visualizações em rede foi realizada em formatos impressos conjugados à sua visualização nas telas para identificação dos clusters. Além da visualização gráfica das redes, utilizamos de versões em que as imagens em si foram plotadas na posição de seu nó correspondente, utilizando o script Image Network Plotter (MINTZ, 2018a). Esta representação facilita o trânsito entre o aspecto visual das imagens e a posição relativa que ocupa na rede, segundo a leitura computacional realizada pelas APIs (Figura 7).

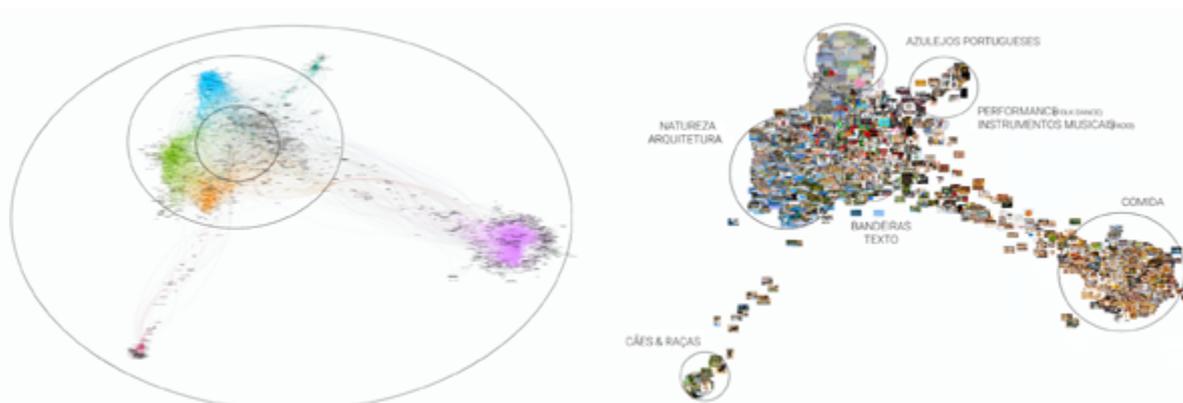


FIGURA 7: Representações da rede imagem-etiqueta gerada a partir do mesmo conjunto de dados: na esquerda, a relação entre as labels/tags e na direita as imagens dispostas espacialmente a partir da relação bi-modal entre a comunalidade das etiquetas.

A análise precisa considerar de forma articulada aspectos dos conjuntos de dados analisados e da API utilizada para a análise. Algoritmos de *layout* baseados em sistemas de força como o ForceAtlas2 funcionam com base na lei de potência e no processo de conexão preferencial (JACOMY et al., 2014), implicando uma lógica bastante particular de leitura do espaço semântico das APIs de visão computacional. De modo importante, as redes bimodais analisadas são formadas pela confluência entre a lógica de etiquetamento e a configuração singular dos conjuntos de dados visuais abordados. Portanto, o modo de distribuição espacial das imagens e etiquetas, do centro à periferia das redes, parece ser resultado de ao menos três fatores: a) a generalidade ou especificidade das etiquetas; b) a variedade de objetos definidas pelo escopo do dataset e sua especificidade tópica; e c) as características topológicas do espaço semântico resultante das associações entre etiquetas em cada serviço. Por exemplo, vemos na Figura 7 agrupamentos bastante particulares posicionados na periferia das redes como resultado do etiquetamento da API da Google: instrumentos musicais, tipos de cachorros e comida. A formação destes agrupamentos é resultado tanto da priorização destes tópicos nos dados quanto do alto grau de granularidade da API da Google ao descrevê-los. Encontramos etiquetas bastante específicas para cachorros, como 'spanish water dog', 'lagotto romagnolo' e 'cão da serra de aires', por exemplo.

Considerando estes parâmetros de análise, as redes entre imagens e etiquetas nos permitiu identificar categorias tópicas dentro de cada conjunto de dados visuais e a relativa predominância de cada uma foi tomada como indicador de tipicidade de cada representação nacional (Figura 8). Ao analisar estas redes lado a lado, foi possível definir categorias tópicas comuns que podem ser observadas nos vários conjuntos de dados, dada sua presença e relativa proeminência em todos os casos. Estas foram: "Natureza", "Comida" e "Pessoas". Adicionalmente, através desta análise

uma categoria única também foi identificada como emergente em cada conjunto: “carnaval”, para Brazilian; “azulejos”, para Portuguese; “cidade” para Austrian; e “dinheiro” para Nigerian.

Como anteriormente observado sobre a Figura 4, as visualizações de rede apresentadas na Figura 8 permitem compreender os diferentes graus de especialização do etiquetamento por cada API. Nesta figura, em particular, percebe-se também como os agrupamentos formados pelo etiquetamento servem como indicadores das visualidades constituídas nos bancos de imagens para cada nacionalidade.

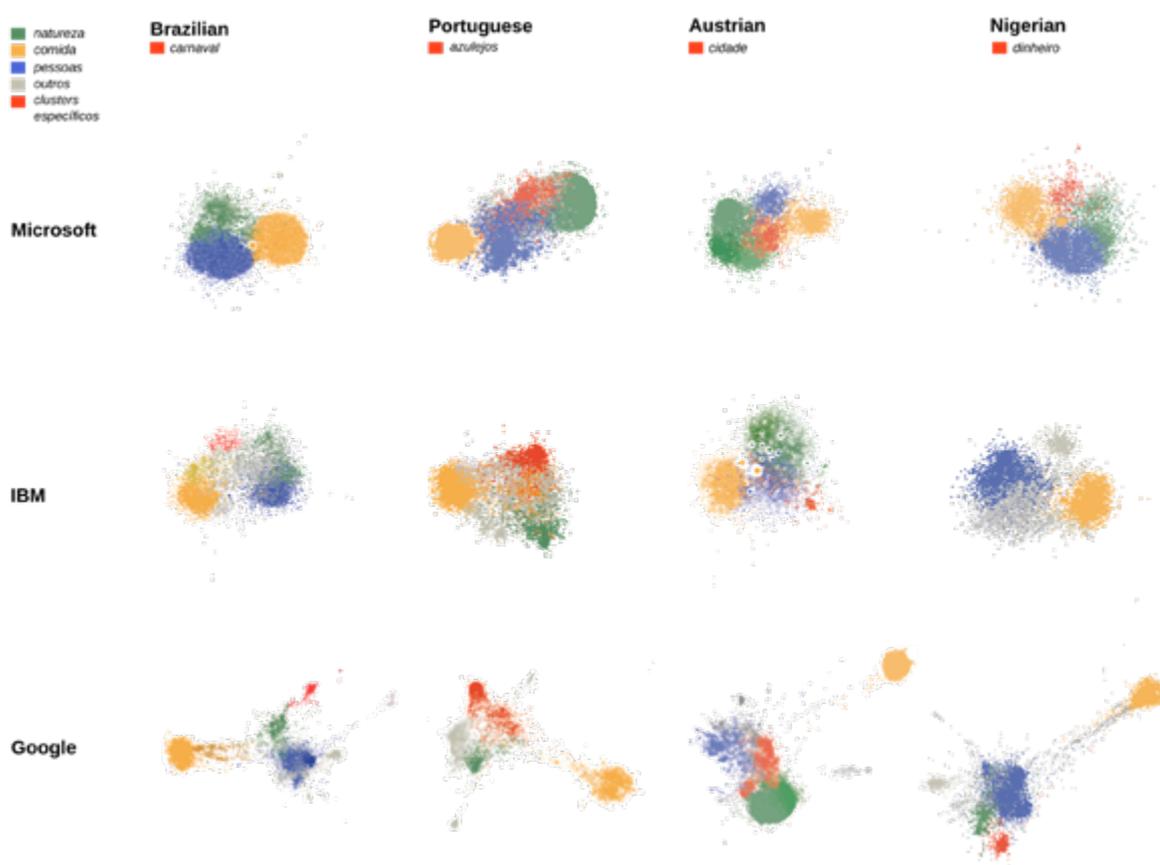


FIGURA 8: Matriz comparativa das redes imagem-etiqueta segundo o gentílico e a API de visão computacional.

As categorias compartilhadas entre os países formam comunidades de dimensões variáveis em cada caso, mas de forma relativamente equilibrada na maior parte deles. Destaca-se o caso austríaco, contudo, cujo cluster relativo a *natureza* é significativamente maior que nos demais. No caso português, o *cluster* relativo a *pessoas* é também comparativamente menor que nos demais casos abordados, enquanto o agrupamento específico, relativo a *azulejos* é bastante pronunciado. Nos casos brasileiro e nigeriano, os agrupamentos relativos a comida e pessoas foram os mais pronunciados.

Indo ao nível das etiquetas em si, algumas destas percepções gerais são especificadas em torno de alguns termos mais frequentes em cada caso. A Figura 9 apresenta gráfico com os 10 termos mais frequentes para os datasets de cada nacionalidade, segundo a API da Google. A frequência em cada caso é comparada com a frequência daquele mesmo termo em outras nacionalidades. Nota-se também a já mencionada excepcionalidade das figurações no dataset *austríaco*, único caso em que as etiquetas mais frequentes não são relacionadas a comida e sim a categorias de paisagem – “*mountain*”, “*natural land*”, “*sky*” etc. O dataset português traz termos relativos às imagens de azulejos após aqueles relativos a comida – “*design*” e “*textile*”. Ambos os datasets brasileiro e nigeriano trazem etiquetas relativas ao agrupamento *peessoas* após aqueles relativos a comida – “*fun*” e “*smile*”.

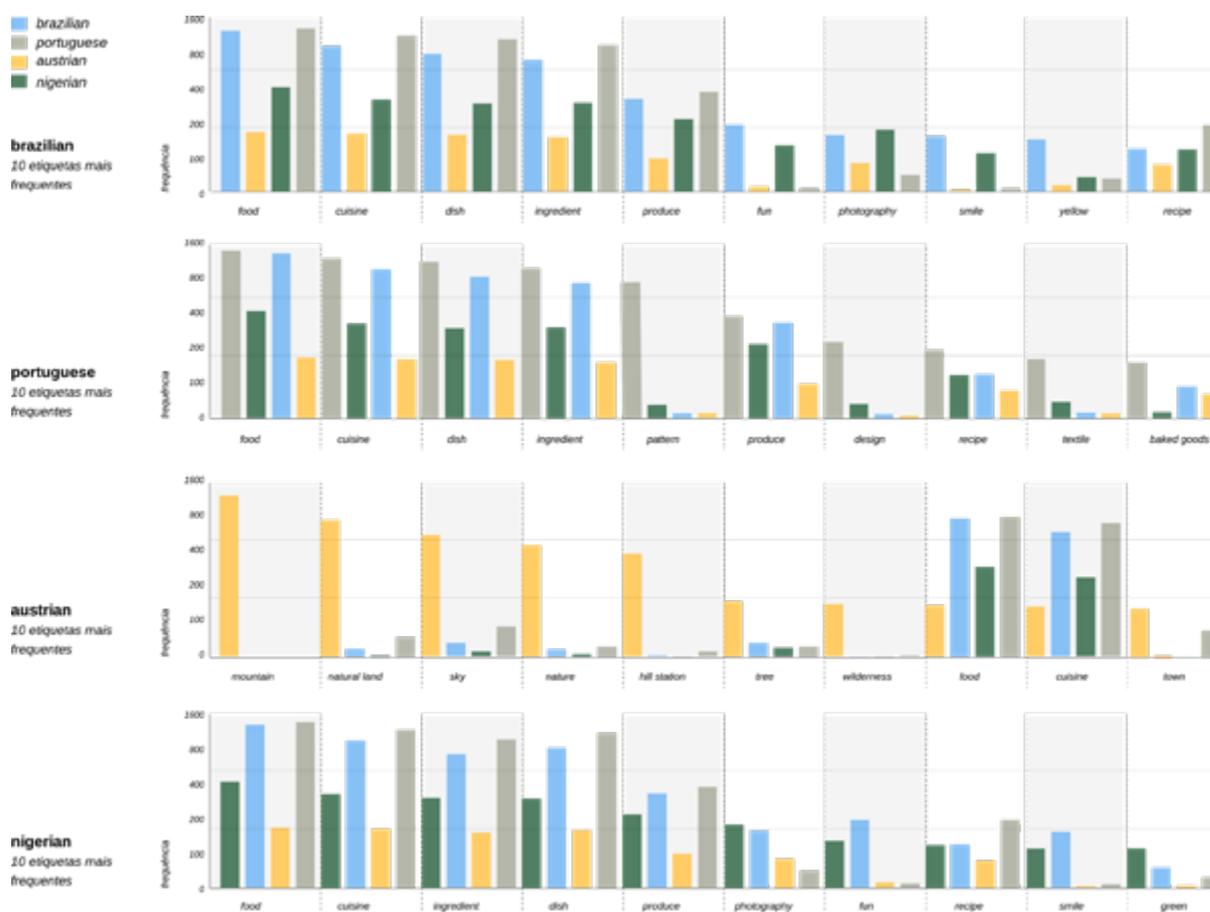


FIGURA 9: Comparativo das 10 etiquetas mais frequentes por gentílico.

No âmbito dos estudos de marketing e consumo, o tema da tipicidade é trabalhado a partir de estudos da psicologia cognitiva como um parâmetro da força de um produto ou marca (LOKEN & WARD, 1990). A noção é compreendida como a força de vinculação de uma instância individual a

uma categoria geral. Ou, posto de outro modo, como a medida de quão exemplar é uma instância para se compreender uma categoria. Para o caso deste estudo, poderíamos dizer da tipicidade de uma figuração quando buscamos compreender em que medida ela seria representativa de uma nacionalidade no âmbito dos bancos de imagem e dos produtos midiáticos que deles se utilizam. A tipicidade, nesse contexto, seria relacionada à frequência com que certas figurações aparecem como resultado da busca pelo gentílico de cada país. Bruhn (2003) destaca o caráter alegórico das imagens de bancos de imagens, que parecem visar a representação de categorias. Contudo, esse fator precisa ser conjugado às mediações algorítmicas mobilizadas na análise e que permitem os agrupamentos destas figurações. Considerando o funcionamento do aprendizado de máquinas, haveria uma dinâmica particular de conformação da tipicidade de uma categoria no processo de treinamento por exemplos. Fatores que contribuem na constituição da categoria envolvem a frequência de vinculação de determinadas figurações e os atributos visuais compartilhados entre estas figurações. Aspectos que, em última medida, são invisibilizados pelo fechamento das tecnologias em APIs e que se relacionam à constituição de suas bases de treinamento e arquitetura de redes neurais.

Entre as indicações de Ali (2016) para a adoção de uma perspectiva decolonial na computação, está, como mínimo, a consideração geopolítica e das políticas dos corpos nos diferentes engajamentos de produção e pensamento sobre computação. Desenhar, construir, pesquisar ou teorizar sobre fenômenos não pode prescindir do olhar sobre as relações de poder inerentes à concentração da posse, desenvolvimento e venda de tecnologias nos EUA e Vale do Silício, por exemplo. Aspectos da tipicidade observada no estudo parecem se relacionar a este aspecto. O enviesamento reproduzido pelas tecnologias em questão vincula-se a posição geopolítica destas empresas e define em larga medida as mediações tecnológicas da comunicação digital contemporânea. A produção de códigos, *datasets* e interfaces segue uma lógica de “espelhamento” através da qual os produtores pensam em usuários similares a si mesmos (HAAS, 2012).

Em linha com a perspectiva dos métodos digitais, como indicamos acima, a mobilização destas ferramentas como dispositivos de análise deve ser combinada com sua consideração crítica contínua, como parte integrante do estudo. O escrutínio crítico e público com sistemas de IA, automação, indexação e categorização de conteúdos pode gerar remediações que minimizam os problemas apresentados (RAJI & BUOLAMWINI, 2019). A centralidade assumida por APIs de visão computacional na interpretação de dados visuais deve ser, assim, considerada criticamente.

CONCLUSÕES

Os estudos de grandes *corpora* em plataformas digitais têm focado em aspectos quantificáveis do ambiente ou das interfaces em si (LAESTADIUS, 2017). Métodos mistos como nossa abordagem de análise de redes visuais capacita os pesquisadores a mudar constantemente entre níveis de visualização e exploração dos dados. Visualizações de rede podem gerar insights únicos através da espacialização e modularidade de dezenas de milhares de imagens. Ao mesmo tempo, estas redes permitem filtrar instâncias específicas relacionadas a um tema ou contexto.

Nós descobrimos padrões sobre o que é comumente relacionado a diferentes países através de seus gentílicos. Comida, Natureza e Pessoas emergem como categorias salientes relacionadas a cada um dos quatro países. Uma vez que todas estas categorias estão relacionadas a conceitos usualmente percebidos como ligados a lugares e culturas, bem como à visualidade de bancos de imagens, consideramos que isto indica adequação dos métodos empregados para a comparação entre os países.

Concomitantemente, as características únicas apontaram para descobertas particulares. Em cada país emergiu um tema específico ligado a sua cultura e estratégias turísticas que, por conseguinte, criam demandas para bancos de imagens. Enquanto estereótipos positivos e negativos exerceram um papel relevante nestas categorias (como “Carnaval” para o Brasil), a oferta de álbuns por fotógrafos e estúdios pode ter enviesado alguns resultados. No caso do Nigéria e Áustria, por exemplo, álbuns prolíficos direcionaram o *dataset* (e suas imagens e etiquetas) a acentuar seus temas. Isto enfatiza uma importante variável não inclusa no estudo: a razão entre a quantidade de imagens sob estudo na plataforma e o número de produtores. Isto é, em nosso projeto, se menos produtores de conteúdo usam uma determinada tag, cada produtor pode influenciar mais a análise, relativamente. Para avançar esta questão, estudos futuros podem comparar países periféricos com países hegemônicos nas indústrias globais de mídia, como EUA, Reino Unido e Japão.

Sobre os fornecedores de visão computacional, o estudo joga luz sobre suas diferenças, limitações e modos de reapropriá-los para explorar a cultura e representações incorporadas em plataformas digitais com métodos digitais. Serviços comerciais como Google Cloud Vision API, Microsoft Azure Computer Vision API e IBM Watson Visual Recognition são considerados “caixas pretas” (BUOLAMWINI & GEBRU, 2018; LATOUR, 2001; PASQUALE, 2016) demandando métodos de auditoria para o aferimento de seu modo de funcionamento, precisão ou cobertura. Nem a lista de etiquetas nem seu número total são informados pelos fornecedores. Esperamos que estudos como o nosso somem ao campo para ajudar outros pesquisadores interessados em usar APIs de computação visual para pesquisa social. Além disto, os diferentes fornecedores

não se sobrepõem ou equivalem no entendimento de complexos dados culturais como imagens, então usá-los acriticamente pode ser um problema metodológico. Alguns dos procedimentos realizados neste estudo (como o dicionário de clustering) mostra a viabilidade de combinar dois ou mais fornecedores mesclando os *datasets* anotados e redes. Futuros experimentos com outros fornecedores são recomendados para ampliar o escopo das descobertas.

Especialistas nos tópicos pesquisados podem direcionar investigações multifacetadas sobre os *datasets*. Comparações internacionais com equipes internacionais podem avançar o entendimento tanto do método (visão computacional) quanto do objeto (bancos de imagens).

Por fim, estereótipos culturais e “tipicidades” sobre um país e suas populações são reproduzidas nos fornecedores de conteúdo, então entender os modos pelos quais isto acontece é importante em prol de ecossistemas de mídia mais justos. Fornecedores de bancos de imagens são importantes para agências de publicidade, relações públicas e empresas editoriais em todo o mundo e o modelo de negócio do *microstock* estende seus impactos também para pequenas e médias empresas e instituições públicas, tornando o entendimento de suas culturas visuais e rotinas produtivas especialmente importantes devido a sua pervasividade.

REFERÊNCIAS

- AIELLO, G., et al. A critical genealogy of the Getty Images Lean In Collection. 2016. Recuperado 19 de agosto de 2019, de <https://wiki.digitalmethods.net/Dmi/WinterSchool2016CriticalGenealogyGettyImagesLeanIn>
- AIELLO, G., et al. *Taking stock: Can news images be generic?* 2017. Recuperado 19 de agosto de 2019, de <https://wiki.digitalmethods.net/Dmi/TakingStock>
- AIELLO, Giorgia; WOODHOUSE, Anna. *When corporations come to define the visual politics of gender: The case of Getty Images*. Journal of Language and Politics, v. 15, n. 3, p. 351-366, 2016.
- ALI, Syed Mustafa. *A brief introduction to decolonial computing*. XRDS: Crossroads, The ACM Magazine for Students, v. 22, n. 4, p. 16-21, 2016.
- ALPAYDIN, Ethem. *Machine learning: the new AI*. MIT press, 2016.
- BLONDEL, Vincent D. et al. *Fast unfolding of communities in large networks*. Journal of statistical mechanics: theory and experiment, v. 2008, n. 10, p. P10008, 2008.
- BRUHN, Matthias. *Visualization services: Stock photography and the picture industry*. Genre: Forms of Discourse and Culture, v. 36, n. 3-4, p. 365-381, 2003.

- BUOLAMWINI, J. *Gender shades: Intersectional phenotypic and demographic evaluation of face datasets and gender classifiers* (Tese de Mestrado, Massachusetts Institute of Technology). 2017. Recuperado de <https://dspace.mit.edu/bitstream/handle/1721.1/114068/1026503582-MIT.pdf>
- BUOLAMWINI, Joy; GEBRU, Timnit. *Gender shades: Intersectional accuracy disparities in commercial gender classification*. In: Conference on fairness, accountability and transparency. 2018. p. 77-91.
- CARDON, Dominique; COINTET, Jean-Philippe; MAZIERES, Antoine. *Neurons spike back: The Invention of Inductive Machines and the Artificial Intelligence Controversy*. 2018.
- CASTELLS, Manuel. *A sociedade em rede*. São Paulo: Paz e Terra, 2002.
- CHOLLET, F. et al. *Keras*. 2015. Recuperado de <https://keras.io>
- D'ANDRÉA, Carlos Frederico de Brito. *Cartografando controvérsias com as plataformas digitais: apontamentos teórico-metodológicos*. Galáxia (São Paulo), n. 38, p. 28-39, 2018.
- D'ANDREA, Carlos; MINTZ, André. *Studying the Live Cross-Platform Circulation of Images With Computer Vision API: An Experiment Based on a Sports Media Event*. International Journal of Communication, v. 13, p. 21, 2019.
- D'ORAZIO, F. *The Future of Social Media Research*. In Woodfield, K. (org.), *Social media in social research: blogs on blurring the boundaries*. NatC1n Social Research, 2014.
- DENG, J. et al. *Imagenet: A large-scale hierarchical image database*. 2009 IEEE conference on computer vision and pattern recognition, 248–255. 2009.
- FAULKNER, S.; VIS, F., & D'ORAZIO, F. *Analysing Social Media Images*. In BURGESS, J.; MARWICK, A.; POELL, T. (Orgs.), *The SAGE Handbook of Social Media* (p. 160–178). 2017.
- FROSH, Paul. *Inside the image factory: stock photography and cultural production*. Media, Culture & Society, v. 23, n. 5, p. 625-646, 2001.
- GEBOERS, M., et al. *Tracing relational affect on social platforms through image recognition*. 2019. Recuperado de Universiteit van Amsterdam website: <https://wiki.digitalmethods.net/Dmi/SummerSchool2019TracingAffect>
- GEPHI Consortium. Gephi (Versão 0.9.2), 2017. Recuperado de <https://gephi.org/>
- GRANDJEAN, Martin; JACOMY, Mathieu. *Translating Networks: Assessing Correspondence Between Network Visualisation and Analytics*. Digital Humanities, 10, 2019.
- HAAS, Angela M. *Race, rhetoric, and technology: A case study of decolonial technical communication theory, methodology, and pedagogy*. Journal of Business and Technical Communication, v. 26, n. 3, p. 277-310, 2012.

- HELMOND, Anne. *The platformization of the web: Making web data platform ready*. Social Media+ Society, v. 1, n. 2, p. 2056305115603080, 2015.
- HENDRICKS, Lisa Anne et al. *Women also snowboard: Overcoming bias in captioning models*. European Conference on Computer Vision. Springer, Cham, 2018. p. 793-811.
- HUSSAIN, Zaeem et al. *Automatic understanding of image and video advertisements*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. p. 1705-1715.
- JACOMY, Mathieu et al. *ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software*. PloS one, v. 9, n. 6, 2014.
- JOO, Jungseock et al. *Visual persuasion: Inferring communicative intents of images*. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014. p. 216-223.
- KARAKO, Chen; MANGGALA, Putra. *Using image fairness representations in diversity-based re-ranking for recommendations*. In: Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization. 2018. p. 23-28.
- LAESTADIUS, L. Instagram. In SLOAN, L.; QUAN-HAASE, A. (Orgs.), *The SAGE handbook of social media research methods* (p. 573–592). Los Angeles ; London: SAGE Publications Ltd, 2017.
- LATOUR, Bruno. *Um coletivo de humanos e não-humanos: no labirinto de Dédalo*. ____ *A esperança de Pandora: ensaios sobre a realidade dos estudos científicos*. São Paulo: EDUSC, p. 201-46, 2001.
- LEMOS, A. *Cibercultura, tecnologia e vida social na cultura contemporânea*. Porto Alegre: Sulina, 2002.
- LOKEN, Barbara; WARD, James. *Alternative approaches to understanding the determinants of typicality*. Journal of Consumer Research, v. 17, n. 2, p. 111-126, 1990.
- MANOVICH, L. *The engineering of vision from constructivism to computers* (Tese (doutorado), University of Rochester), 1993. Recuperado de <http://manovich.net/EV/EV.PDF>
- MARRES, Noortje; MOATS, David. *Mapping controversies with social media: The case for symmetry*. Social Media+ Society, v. 1, n. 2, p. 2056305115604176, 2015.
- MINTZ, A. *Image Network Plotter*, 2018a. Recuperado de <https://github.com/amintz/image-network-plotter>
- MINTZ, A. *Memespector Python*, 2018b. Recuperado de <https://github.com/amintz/memespector-python>
- MINTZ, A. *Stock scraper*, 2019. Recuperado de <https://github.com/amintz/stock-scraper>
- NAPOLI, Philip M. *Toward a model of audience evolution: New technologies and the transformation of media audiences*. McGannon Center Working Paper Series, p. 15, 2008.

- NIEDERER S.; COLOMBO, G. Visual Methodologies for Networked Images: Designing Visualizations for Collaborative Research, Cross-platform Analysis, and Public Participation. *Diseña*, (14), 40–67, 2019.
- NOBLE, Safiya Umoja. *Algorithms of oppression: How search engines reinforce racism*. New York: NYU Press, 2018.
- OMENA, J. J. (ed.). *Métodos digitais: teoria-prática-crítica*. Lisboa: ICNOVA, 2019. Recuperado de: <https://www.icnova.fcsh.unl.pt/metodos-digitais-teoria%E2%80%90pratica%E2%80%90critica/>
- OMENA, J. J.; RABELLO, E.; MINTZ, A. *Visualising hashtag engagement: Imagery of political polarization on Instagram*. 2017. Recuperado de Universiteit van Amsterdam website: <https://wiki.digitalmethods.net/Dmi/InstagramLivenessVisualisingengagement>
- OMENA, J.J. et al. *Bots and the black market of social media engagement*. 2019. Recuperado 20 de agosto de 2019, de <https://wiki.digitalmethods.net/Dmi/SummerSchool2019Botsandtheblackmarket>
- OSOBA, O. A.; WELSER IV, W. *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*. Rand Corporation, 2017.
- PASQUALE, F. *The Black Box Society – The Secret Algorithms That Control Money and Information*. Cambridge, Massachusetts London, England: Harvard University Press, 2016..
- PEARCE et al. *Visual cross-platform analysis: digital methods to research social media images*, 2018. *Information, Communication & Society*, 1-20.
- PORTILLA, Jesús Homero Insuasti. *Ciencias de la computación:¿ un reto para el pensamiento decolonial?.* *Revista Criterios*, v. 20, n. 1, p. 91-99, 2013.
- PRITCHARD, Katrina; WHITING, Rebecca. *Taking stock: A visual analysis of gendered ageing*. *Gender, Work & Organization*, v. 22, n. 5, p. 510-528, 2015.
- QUIJANO, Aníbal. *Colonialidade do poder e classificação social*. In: SANTOS, B. (org). *Epistemologias do Sul*. São Paulo: Cortez, p. 84-130, 2010.
- RAJI, Inioluwa Deborah; BUOLAMWINI, Joy. *Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial ai products*. In: *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. 2019. p. 429-435.
- RICCI, Donato et al. *Designing Digital Methods to monitor and inform Urban Policy. The case of Paris and its Urban Nature initiative*. In: *3rd International Conference on Public Policy (ICPP3)-Panel T10P6 Session 1 Digital Methods for Public Policy*. SGP, 2017. p. 1-37.
- RIEDER, B.; DEN TEX, E., & MINTZ, A. *Memespector*. 2018. Recuperado de <https://github.com/bernorieder/memespector>

ROBERTS, L. G. *Machine perception of three-dimensional solids*. 1963. Recuperado de <http://www.packet.cc/files/mach-per-3D-solids.html>

ROGERS, R. (2013). *Digital methods*. Cambridge, Massachusetts: The MIT Press.

ROGERS, Richard. *Digital methods for web research*. Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource, p. 1-22, 2015.

RYKOV et al. Semantic and geospatial mapping of Instagram Images in Saint-Petersburg. 2016. Proceedings of the AINL FRUCT 2016 Conference Saint-Petersburg, Russia, 10-12 November 2016. Recuperado de <http://ieeexplore.ieee.org/servlet/opac?punumber=7889413>

SILVA, T.; BARCIELA, P.; MEIRELLES, P. *Mapeando Imagens de Desinformação e Fake News Político-Eleitorais com Inteligência Artificial*. 3º CONEC: Congresso Nacional de Estudos Comunicacionais Da PUC Minas Poços de Caldas - Convergência e Monitoramento, 413–427, 2018. Recuperado de <https://conec.pucpcaldas.br/wp-content/uploads/2019/06/anais2018.pdf>

SILVA, T.; MEIRELLES, P.; APOLONIO, B. *Visão Computacional nas Mídias Sociais: Estudando imagens de #Férias no Instagram*. Apresentado em I Encontro Norte e Nordeste da ABCiber, São Luís, 2018.

SMEULDERS et al. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), 1349–1380, 2000.

SRNICEK, N. *Platform Capitalism* (1º ed). Cambridge, UK ; Malden, MA: Polity Press, 2017.

TIFENTALE, A.; MANOVICH, L. Selfiecity: Exploring photography and self-fashioning in social media. In *Postdigital Aesthetics* (p. 109–122). Springer, 2015.

VAN DJICK, J. *The culture of connectivity: A critical history of social media*. Oxford, UK; New York: Oxford University Press, 2013.

VENTURINI, T.; JACOMY, M.; JENSEN, P. *What Do We See When We Look at Networks. An Introduction to Visual Network Analysis and Force-Directed Layouts* (SSRN Scholarly Paper N° ID 3378438), 2019. Recuperado de Social Science Research Network website: <https://papers.ssrn.com/abstract=3378438>

VIS, F. A critical reflection on Big Data: Considering APIs, researchers and tools as data makers. *First Monday*, 18(10), 2013.

WEST, C. *The Lean In Collection: Women, Work, and the Will to Represent*. *Open Cultural Studies*, 2(1), 430-439, 2018.

-
- [1] Tarcízio Silva é Mestre em Comunicação e Cultura Contemporâneas pela UFBA e realiza doutorado em Ciências Humanas e Sociais pela UFABC. Organizou publicações como “Comunidades, Algoritmos e Ativismos Digitais: olhares afrodiáspóricos” (LiteraRUA, 2020) e “Monitoramento e Pesquisa em Mídias Sociais: metodologias, aplicações e inovações” (Uva Limão, 2016). E-mail: eu@tarciziosilva.com.br
- [2] André Mintz é professor adjunto de arte e tecnologia do Departamento de Fotografia e Cinema da Escola de Belas Artes da UFMG e pesquisador do R-EST, grupo de estudos em redes sociotécnicas. Doutor e mestre em Comunicação Social pela mesma instituição, também cursou mestrado em Culturas da Arte Mídia na Aalborg University, Lodz University e Danube University Krems. Sua pesquisa volta-se às relações entre arte, visualidades e infraestruturas computacionais em articulações com métodos digitais e estudos sociais de ciência e tecnologia. E-mail: andregmintz@ufmg.br
- [3] Janna Joceli Omena é investigadora doutoral em Digital Media e professora convidada na Information Management School (NOVA IMS), Universidade Nova de Lisboa. Desenvolve o seu trabalho no iNOVA Media Lab como coordenadora do grupo SMART (Social Media Research Techniques) e líder do SMART Data Sprint. É responsável pela edição do livro “Métodos digitais: teoria-prática-crítica” (2019) e tem particular interesse na tecnicidade apresentada pela especificidade do meio (medium-specificity). Principais linhas de pesquisa: métodos digitais, estudos de plataformas-software e análise visual de redes. Blog de investigação: <https://thesocialplatforms.wordpress.com> Twitter: [@JannaJoceli](https://twitter.com/JannaJoceli). [4] B e a t r i c e Gobbo é estudante de doutorado em Design no Politecnico di Milano desde 2018. Ela é membro do Density Design Lab, um grupo de pesquisa voltado à visualização da informação. Sua pesquisa atual enfoca o papel do design de comunicação no campo das ciências da computação.
- [5] Taís Oliveira é Doutoranda no Programa de Pós-Graduação em Ciências Humanas e Sociais na Universidade Federal do ABC, linha de pesquisa Políticas Públicas, participação social e ação coletiva. E-mail: tais.oliveira@ufabc.edu.br
- [6] Helen Tatiana Takamitsu é doutora em Engenharia de Produção e membro do Grupo de pesquisas “Estudos Métricos em Informação”, do Programa de Pós Graduação da Ciência da Informação da Unesp de Marília. Sua linha de pesquisa é em Marketing Digital, Gestão de Negócios Digitais, Empreendedorismo no Meio Digital e Análise de Redes Sociais.
- [7] Elena Pilipets é pesquisadora pós-doutoral no departamento de Estudos de Comunicação e Mídia da Alpen-Adria-Universität Klagenfurt, Áustria e pesquisadora do SMART (Social Media Research Techniques) do iNOVA Media Lab, Universidade NOVA de Lisboa, Portugal. Ela obteve seu doutorado em Teoria da Mídia e Estudos Culturais e agora trabalha com os temas da circulação em mídias sociais, controle algorítmico e engajamento afetivo. Seus interesses de ensino e pesquisa são estudos culturais da mídia, pesquisa em internet, métodos digitais, afeto e teoria ator-rede.
- [8] Hamdan Azhar é Vice-Presidente de Ciência de Dados em uma empresa estadunidense de gestão de ativos. Anteriormente foi cientista de dados Facebook e é o fundador do laboratório de dados PRISMOJI. Seus interesses de pesquisa estão na interseção de ciência de dados, ciência social computacional, tecnologia e sociedade.

- [9] O artigo é derivado de um estudo realizado coletivamente no âmbito do SMART Datasprint 2019, evento organizado pelo iNOVA Media Lab, da Universidade Nova de Lisboa. O relatório da investigação encontra-se disponível no endereço: <https://smart.inovamedialab.org/smart-2019/project-reports/interrogating-vision-apis/>. Versão deste estudo foi também apresentada no VIII Simpósio de Ciência Tecnologia e Sociedade, da Associação Brasileira de Estudos Sociais das Ciências e das Tecnologias (ESOCITE), em Belo Horizonte.
- [10] Paralelamente a estas APIs comerciais, vale destacar projetos de código aberto e modelos públicos de reconhecimento de imagens como os disponibilizados no projeto Keras (Chollet et al., 2015). Nesse caso, oferece-se uma entrada facilitada ao desenvolvimento de aplicações com modelos de reconhecimento de imagens restritos baseados na tipologia da base de dados Imagenet (Deng et al., 2009), com 1000 categorias pré-treinadas.
- [11] <https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html>
- [12] <https://www.theguardian.com/us-news/2019/apr/22/google-mass-protests-employee-retaliation>
- [13] <https://www.theguardian.com/us-news/2019/jul/11/amazon-ice-protest-immigrant-tech>
- [14] No original: “an ongoing legacy of colonialism in contemporary societies in the form of social discrimination [...] practices and legacies of European colonialism in social orders and forms of knowledge”. Tradução nossa.
- [15] No original: “obliga a asumir una mirada desde el paradigma positivista, dado además su enfoque empírico analítico propio de las ciencias experimentales”. Tradução nossa