DÍGITOS
Revista de Comunicación Digital

Visualizing an image network without rendering files: A method to combine user hashtags with computer vision labels

Visualización de una red de imágenes sin renderizar archivos: Un método para combinar hashtags de usuarios con etiquetas de visión computacional

Giulia Tucci
giuliatucci@ufrj.br
Universidade Federal do Rio de Janeiro

Received: 13/01/2022
Acepted: 10/05/2022

ABSTRACT This article presents a method for visualizing networks of geolocated images without rendering the image files on the network. The path I followed to develop this method is the result of an intensive "data sprint" which took place during the University of Amsterdam Digital Methods Initiative Summer School 2021. During the data sprint, I developed a methodological framework to generate a network of Twitter geolocated images combining the hashtags twitted with the images and the Google Cloud Vision API best single expression to describe each image (*best guess label*). Considering the limitations of working with a massive amount of image data and the computational memory required to generate network visualizations, the possibility of using description tags to create image networks is promising. The images analyzed during this study were extracted from Twitter filtering for the #deepfakes and #deepfake and tagged with country code location. Thus, the hashtags included in the tweets by Twitter users provide the context and the user description of the image. This information was combined in a bipartite network with a computer vision entity, the computer vision description of the image, to generate a networked description of the whole image set. I point that this method can be considered in exploratory research when working with large sets of images.

KEYWORDS Digital Methods, Cloud Vision API, Network Analysis, Twitter hashtags, Deepfakes

RESUMEN Este artículo presenta una metodología de visualización de redes de imágenes geolocalizadas sin renderizar los archivos de imagen en la red. El desarrollo del método, aquí descrito, resulta del "sprint de datos" intensivo logrado durante la *Digital Methods Initiative Summer School* 2021 de la University of Amsterdam. Durante el *data sprint*, fue desarrollado un marco metodológico para generar una red de imágenes geolocalizadas de Twitter que combina los hashtags twitteados con las imágenes y la mejor expresión unitaria de Google Cloud Vision API a describir cada imagen (*best guess label*). Considerando las limitaciones al trabajar cantidades masivas de datos de imágenes y la memoria computacional requerida para generar visualizaciones de redes, el método promete la posibilidad de usar etiquetas de descripción para la creación de redes de imágenes. Las imágenes analizadas durante este estudio se extrajeron del filtrado de Twitter para #deepfakes y #deepfake, etiquetados a la ubicación del código de país. Así, hashtags incluidos en los tuits de usuarios de Twitter aportan el contexto y la descripción de la imagen por parte del usuario. Esta información se combinó en una red bipartita con una entidad de Computer Vision y la descripción de Computer Vision de la imagen, al fin de generar una descripción en red de todo el conjunto de imágenes. Señaló que el método es considerable a la investigación exploratoria en casos de grandes conjuntos de imágenes.

PALABRAS CLAVE Métodos digitales, API Cloud Vision, análisis de redes, hashtags, Twitter, Deepfakes

How to cite this article:
Tucci, G. (2022).Visualizing an image network without rendering files: A method to combine user hashtags with computer vision labels. *Dígitos. Revista de Comunicación Digital,* 8: 109-126. DOI: 10.7203/drdcd.v1i8.237

Visualizing an image network without rendering files: a method to combine user hashtags with computer vision labels

1. Introduction

In this article, I propose an exploratory method for the visualization of a network of geolocated images without the necessity to render the image files in the graph. To interpret the images without plotting the files, I built upon a recipe proposed by Chao and Omena (2021) to generate and analyze image networks without rendering the images within the network. My strategy was to combine the Computer Vision API output label that defines each image, as indicated in the recipe, with the hashtags posted by Twitter users to describe and contextualize the images. Therefore, I created a bipartite network where the edges that connect the hashtags nodes and the computer vision labels nodes represent the images. In situations where there is a small set of images to be processed and rendered as nodes within a network, there is no requirement for massive dedicated computer memory, and the execution of the task is relatively easy. However, as the volume of images comprising a study dataset increases, the process of showing images on a network becomes more difficult and computer memory-demanding. Although I developed this method during a study conducted with a small set of images, it is recommended to test it, validate it, and apply it in network analyses of a large set of image files.

The crescent profusion of Internet data produced by digital platforms users and of information generated by artificial intelligence tools associated with these platforms, as well as their application in multidisciplinary academic research, gave rise to an extensive demand for the development of research strategies, tools, and methods to extract, process, and visualize information of (sometimes) massive sets of data. That demand intensified a not recent debate concerning the choice for qualitative, quantitative, or mixed-methods approaches. In addition, the increasing “societal relevance of applications” raised the argument that computer science “needs to define itself as a socio-technical discipline that contributes to the solution of social problems in context” (Stevens et al., 2018: 23).

Digital Methods, a path of study for doing Internet research advanced by Richard Rogers (2009), repurposes traditional social sciences and humanities research methods and emphasizes data collection and analysis methods developed within the Internet medium and for Internet data research. In other words, the path pioneered by Rogers suggests that, instead of adapting traditional research methods that focus on collecting information from users (interview, survey, etc.), we study and learn from the features, affordances, and methods developed by the leading platforms (Google, Facebook, Twitter, Telegram, etc.). Digital Methods “seeks to move Internet research beyond the study of online culture and beyond the study of the users of ICTs [Information and

Communication Technologies] only" (Rogers, 2013: 17) and surpasses the qualitative or quantitative research dilemma as a research practice that reworks digital platforms data, web services mechanisms, and platforms' structures to do Social Sciences research (Omena, 2019).

Every year, the University of Amsterdam's Digital Methods Initiative (DMI) holds winter and summer data sprints. Participating in a data sprint is part of an intense and fast-paced learning experience. After being trained to apply research tools and techniques related to a project, the goal is to spend a couple of days using the data and the resources to identify relevant findings and help answer the proposed research questions.

Working with a small amount of time generates a need to be immersed in the data, testing different ways to explore the dataset and create visualizations. This process outcome is motivating and, in my case, was the inspiration for combining geolocation data, hashtags, and computer vision outputs as nodes within a network. Presenting the results and participating in group discussions oriented by experienced mentors were fundamental steps in perceiving that I could systematize the analyses I was conducting into an image analysis technique.

The 2021 Digital Methods Initiative Summer School happened virtually during the coronavirus pandemic from 5 to 16 July 2021. The DMI Summer School theme was Fake everything: Social media's struggle with inauthentic activities, and the project I worked on during the first half of the program was Mapping deepfakes with digital methods and visual analytics, facilitated by Richard Rogers, Lucia Bainotti, Sarah Burkhardt, Gabriele Colombo, Janna Joceli Omena, and Jason Chao. We aimed to interpret the sociological and cultural understandings of deepfakes by using visual analyses and Digital Methods.

Designated as the "Photoshop of videos" by Bimber and Gil de Zúñiga (2020), deepfake technology was the subject of an alerting announcement created and published by Buzzfeed (Mack 2018) in which a fabricated Barack Obama makes a pronouncement voiced by an actor. The Buzzfeed piece that warned about the new technology danger circulated widely on social media (Vaccari and Chadwick 2020), generating concern in the general public regarding new ways of manufacturing content and contributing to an environment of uncertainty.

Initially employed by pornographic content creators to overlap celebrities' faces in sexual content videos (Cole, 2017), deepfake is a technique to algorithmically create artificial videos by swapping faces starting from two people video footages (Westerlund, 2019). When this new technology emerged, it created a fuss concerning the falsification of discourses, sabotage activities, and national security matters (Chesney, Citron, 2018; Westerlund, 2019) since deepfake is a tool capable of distorting the boundaries of what seems real for disseminating disinformation.

The rise of deepfakes disquieted scholars resulting in works focused on deepfakes effects and ways of mitigating them. Kietzmann et al. (2020) created a framework for managing deepfakes risk and diminishing the technology's dangerous effects. Vaccari and Chadwick (2020) conducted an experiment to check if the Buzzfeed Obama deepfake video had the potential to delude individuals and alter their perception of its authenticity. The researchers did not find evidence suggesting that the manipulated video misled the participants. However, they advised that unchecked deepfake content can increase the levels of distrust in the online informational environment.

My group project goal was to understand whether the conversation on the Internet about deepfakes is related to threats on the informational environment or, despite the deceptive and damaging potential of this new technology for society, it focuses on describing the techniques used and on debunking the hype surrounding deepfakes. Thus we designed a research project to identify the related topics and the context of the conversation about deepfakes from 2017 to 2021.

The project was divided into three sub-projects since there were three main paths to cover: Discourse mapping, Image vernaculars and trends, and Computer vision. The Computer Vision team was divided again into three groups to examine the discourse related to deepfakes using Twitter and Google Cloud Vision data. The three sub-groups project titles were: Google Vision web entities for Google images over time, Sites of image circulation: Google images and Google Vision's fully matching images, and Geographical mapping: Twitter hashtags and Google Vision web entities.

During the first data sprint week, my working team was responsible for the Geographical mapping with Twitter hashtags and Google Vision entities study and formulated two research questions in order to geographically map the images used on conversations about deepfakes on Twitter: (i) Where do #deepfake or #deepfakes images circulate geographically on Twitter over time? and (ii) Are there country-specific discourses around deepfake? If yes, what are the themes found in these country-specific discourses?

The first research question was methodologically addressed by Dr. Martin Roth, with the development of a choropleth world map indicating deepfakes circulation across location and time (see Tucci, Roth, Saxler, 2021). I was responsible for exploring the data and conducting network analyses to answer the second research question. While discussing with my subgroup data sprint mentor, Janna Joceli Omena, about the partial results of the network analysis graphs I created, she enlightened me about the sophisticated way I had just encountered to represent image networks without the need to dedicate all computer memory to Gephi process image files. The solution I unfolded to define the images was to merge Twitter hashtags chosen by who tweeted to identify the image and a Computer Vision API outputs.

Gephi, a traditional software for network analysis, demands high dedicated computer memory to run. On Gephi's GitHub page, there are popular issues (Chezsick 2017; Ghost 2016; KallyopeBio 2018) related to memory crashes after users follow the software installation guide's specific instruction to increase memory default settings. Accordingly, the larger the number of elements contained in a network (nodes and edges), the more complex the tasks of generating, configuring, and saving the visualizations produced in Gephi. When plotting images in a network, depending on the number of files to be rendered as nodes, the work can be difficult to execute (or even be undoable). Doing visual research with digital platforms data requires specific methodologies, considering the importance of "approaching these collections of images as data" and not as content, as proposed by Niederer and Colombo (2019) in an article that describes the process of creating visual tools for digital research.

To address these issues, there is a demand for establishing methodological frameworks and strategies that treat image content as data and require less execution time and less dedicated computer memory. Thus, this paper aims to outline the methodological procedure that I have developed to generate a bipartite network that describes a set of images without the necessity to render the image files in the graph. In the following section, the dataset creation and the methods applied to run the analyses are presented, followed by the results and discussion introduced in the third section. In the last section, final considerations and a research agenda are discussed.

2. Materials and Methods: geographical mapping of Twitter data & computer vision labels

This section first describes how to curate a dataset of Twitter hashtags. Then, it explains how to identify geolocated tweets containing images. Subsequently, the section demonstrates how digital methods recipes can serve as inspirational tools for new methodological experiments.

2.1. Dataset description

Before the beginning of the data sprint, the original dataset was collected via Twitter API by filtering tweets containing image files and the hashtags #deepfake or #deepfakes, from January 1st, 2017 to June 1st, 2021, and excluding retweets and replies. The acquired dataset comprises 98.831 tweets that originated 19.713 unique image URLs.

Hashtags are an established resource created on Twitter and used to tag and group content related to specific topics, events, entities, etc. Posting a hashtag on Twitter generates a hyperlink that points to an aggregate of tweets containing the tag, grouping content on that topic. Computer Vision APIs are image recognition technologies that allow researchers to process images using machine learning algorithms and OCR technology to extract information regarding the content of the analyzed image. In this work, I use Google Cloud Vision API as the source of a computer vision description to the image files in my dataset. The feature I chose to work with was the best guess label, which selects a single expression to describe an image, “the service’s best guess as to the topic of the request image. Inferred from similar images on the open web”, as Google (2020) defined it.

Considering the aim to study how conversations about deepfakes travel around the globe, the working dataset was created by combining the 459 geolocated tweets, the 459 associated image files, and the best guess label results for 416 images since the vision API did not process 43 video thumbnail files. The Google Cloud Vision API feature was obtained via Memespector GUI (Chao, 2021).

To explore the Twitter dataset, I summarized the hashtags related to the conversation about deepfakes by using the R package rtweet (Kearney, 2017). Since the data was collected by filtering for the #deepfakes and #deepfake, they are the most frequent hashtags (234 and 232 occurrences, respectively), followed by #ai (63), #fakenews (36), and #protectmyimage (28). The ranking of most frequent hashtags indicate that the majority of images about deepfakes posted with geolocation on Twitter comprises conversations on fake news and artificial intelligence as well as discussions on the protection of social media profile pictures and other shared photos protection against manipulations and stealing (Graham et al., 2014).

2.2.Network analysis with computer vision

Based on a DMI recipe (Chao, Omena, 2021) on how to work with the results obtained after processing the images with a computer vision tool - the MemeEspector-GUI outputs - I followed this procedure to create the networks: I imported raw data to Table 2 Net tool, selected the type of nodes and respective attributions, and imported the network files generated into Gephi. Then, I processed the data to create each network visualization presented in this article, according to the respective objective of the analysis. The specific parameters used to generate each graph are described below.

3. Results and Discussion

3.1. Analysing a network of Twitter hashtags and images

In order to interpret and visualize the dataset of images, the first step was to plot a bipartite network graph of Twitter hashtags and images (Figure 1). To understand how users combine images and hashtags in a single tweet and how this information correlates to specific countries' discourses, I applied a color code to identify countries in the network. Since all 459 images are unique and tweeted from specific countries, I colored the image nodes and their output edges by the tweet country code. Bearing in mind the Twitter hashtags can repeat throughout the dataset and that every tweet contains the #deepfake and/or #deepfakes, hashtags nodes were colored white. In addition, the .jpg files corresponding to the images were rendered as nodes by using the ImagePreview Gephi Plugin (Xue, 2012).

According to Figure 1 there are hyperconnected hashtags shared by Twitter users from diverse countries (i.e. #fakenews, #ai, #deeplearning, and #journalism) and these tags tend to indicate a conversation about tools, methods and subjects used for the creation, characterization and/or debunking of deepfakes.

By zooming into the US (Figure 2) as it is the country with the most locational data (168 images) and the country where dominant discourses appear, whereas there are some images that refer to conferences and academic presentations, the more dominant trend related to the discourse of deepfakes are pop culture and celebrities and commercial brands.

Aiming to explore images' visual elements and to add another visual component to the study of the US discourse surrounding deepfakes, I conducted a visual analysis sorting the 168 US images by color (Figure 3) using PicArrange software (Jung, 2021). The analysis of the US tweeted images associated with #deepfakes or #deepfake provides a sense of image trends and vernaculars (Rogers, 2021). Although it is not possible to identify

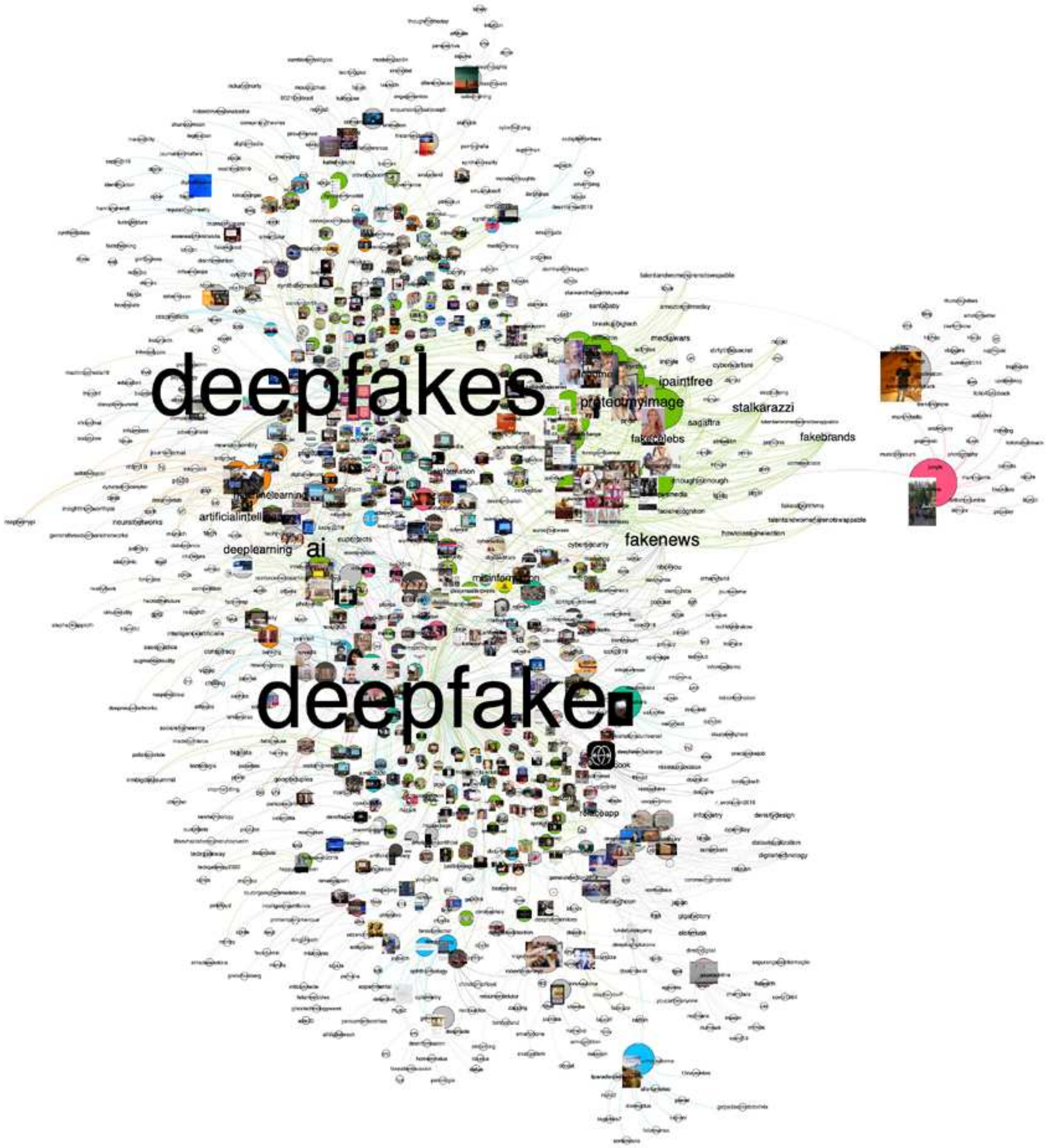

Figure 1: Bipartite network graph of 459 tweeted images and 710 Twitter hashtags in tweets containing #deepfakes and/or #deepfake represented as 1,169 nodes and 1,914 undirected edges. Network visualization was created using Gephi's Force Atlas 2 layout, images nodes were sized by degree, and nodes and edges were colored by country (United States = green, Great Britain = light blue, France = brown, Germany = orange, Canada = pink, and other countries = gray) and the respective .jpg files were rendered with Gephi ImagePreview Plugin. Hashtags' nodes were colored white and the respective labels were sized by frequency.

a dominant pattern of images in the entire dataset, and admitting it consists of unique images, this visual analysis shows the prevalent imagery of celebrities, pop culture, and academic presentations (or conferences) on the US located tweets. The image wall (Figure 3) provides the opportunity to identify images that relate to exemplifying and describing the creation of deepfakes content. For example, pictures comparing human faces side to side (the original versus the deepfake altered content).

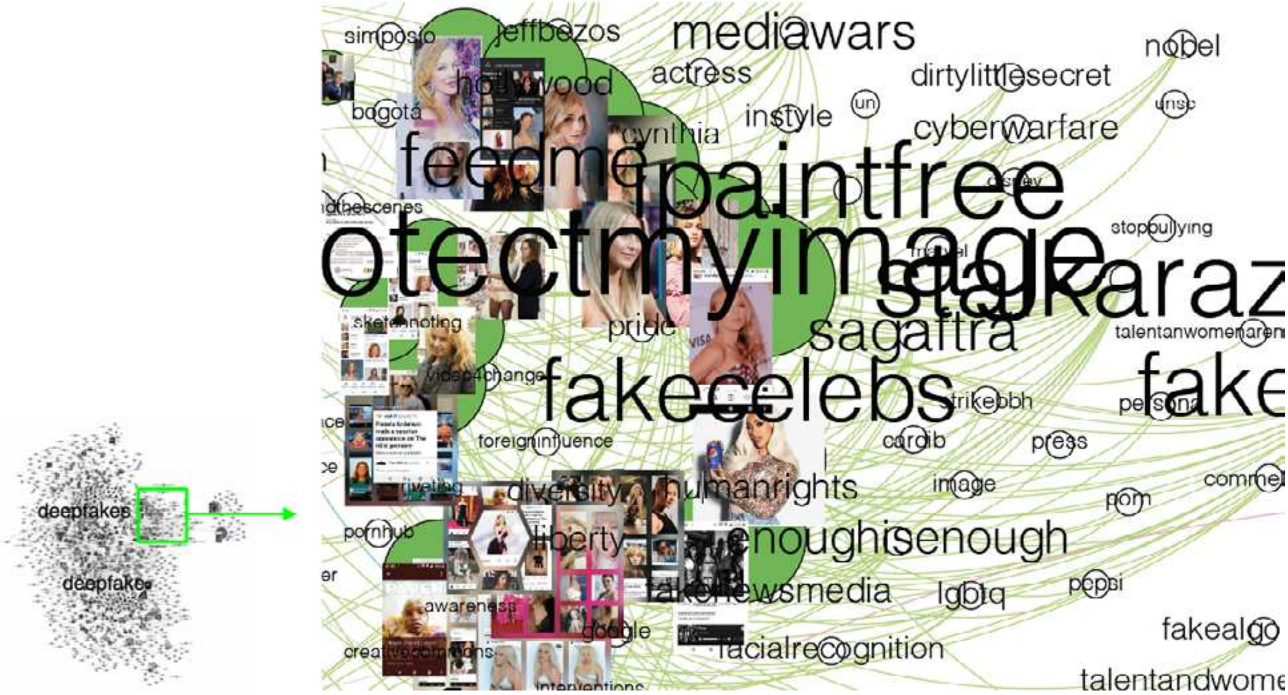

Figure 2: Zoom of the bipartite network of images and Twitter hashtags in tweets containing #deepfakes and/or #deepfake (green nodes represent images shared on Twitter in the United States). The majority of images show faces of celebrities and the pop culture and celebrities related hashtags #hollywood, #cardib, #press, #instyle, #mediawars, #actress, #fakecelebs, #pepsi, #jeffbezos.

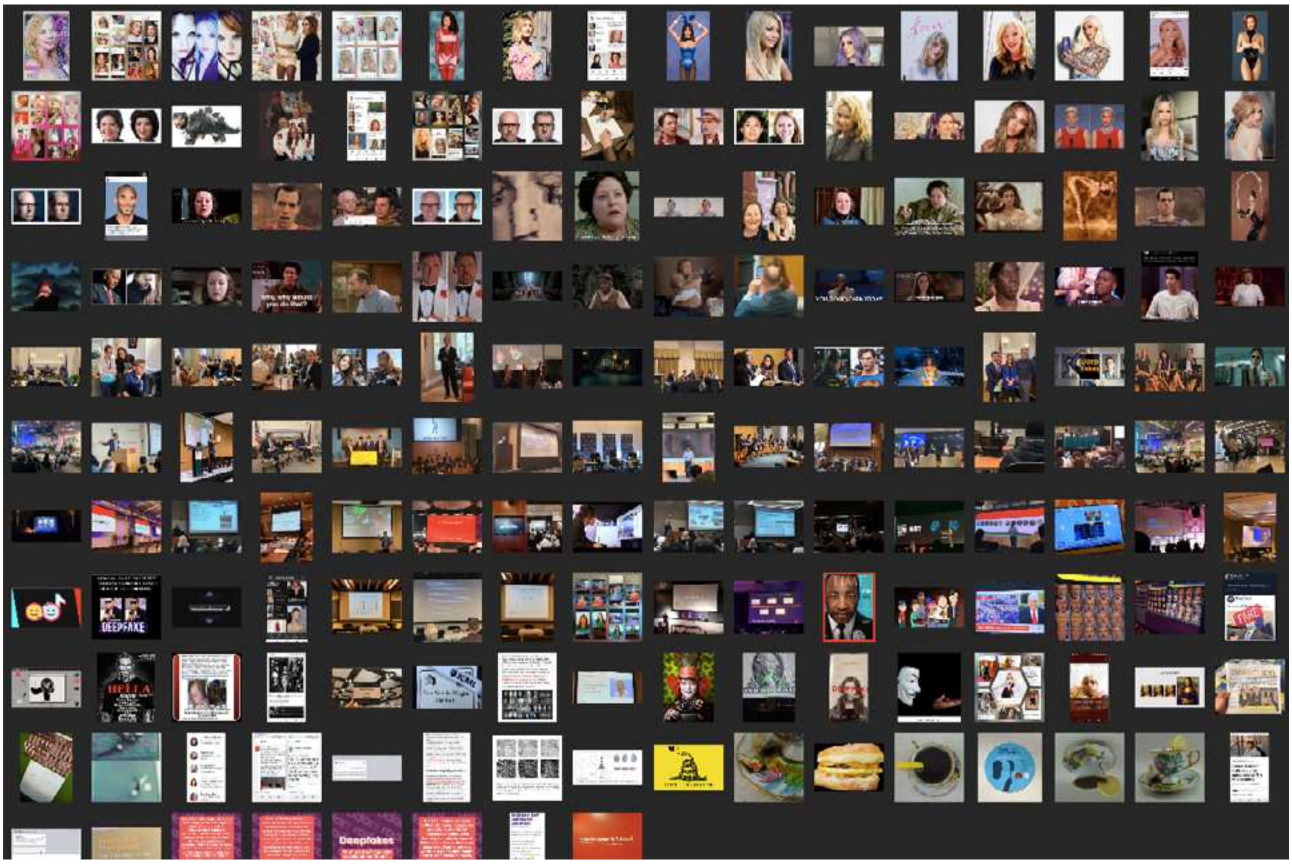

Figure 3: Image sorting analysis of 168 US images related to the discourse around deepfakes on Twitter. This visualization was created using the PicArrange software (Jung, 2021).

3.2. Analysing a network of Twitter geolocated images and best guess labels

To test the existence of country-specific Google vision vernaculars related to the deepfake debate on Twitter, the strategy I have chosen was to link tweeted images' geolocation data with Google Cloud Vision's automated description of the images. Hence the second step of the research process was to generate a bipartite network to relate the Cloud Vision API feature of each image to the country code returned by Twitter API (Figure 4).

In Figure 4, nodes that represent best guess labels were colored 'light blue', and country flags images were rendered as the respective country node. The edges represent the images. All nodes were sized by occurrence number, and the labels were scaled proportionally. In addition, I have colored the nodes and the edges of the three most frequent countries in our data: United States (168 images, color = royal blue), Great Britain (45 images, color = green), and France (32 images, color = red). The graph was generated with country flags files obtained in a GitHub flags repository (Borgos, 2021).

The network connecting the computer vision best guess label and the image geolocation shows specific clusters of Google Cloud Vision labels for different countries, demonstrating that the images tweeted from diverse locations were tagged by Cloud Vision API with specific groups of visual entities. In other words, the majority of Google visual vernaculars for #deepkafes and #deepkafes are country-specific. Taking into account that platforms vernaculars are shaped by platforms mediated practices and users habits of communication (Gibbs et al., 2015) and that visual vernaculars are patterns of images used to articulate and communicate an issue (Pearce, Colombo, 2019), Figure 4 indicates that Twitter users from different countries have particular ways of communicating about the deepfakes issue. However, some best guess labels appear in the content tweeted in more than one country, being the most frequent Cloud Vision labels in this data: presentation, communication, display device, and seminar.

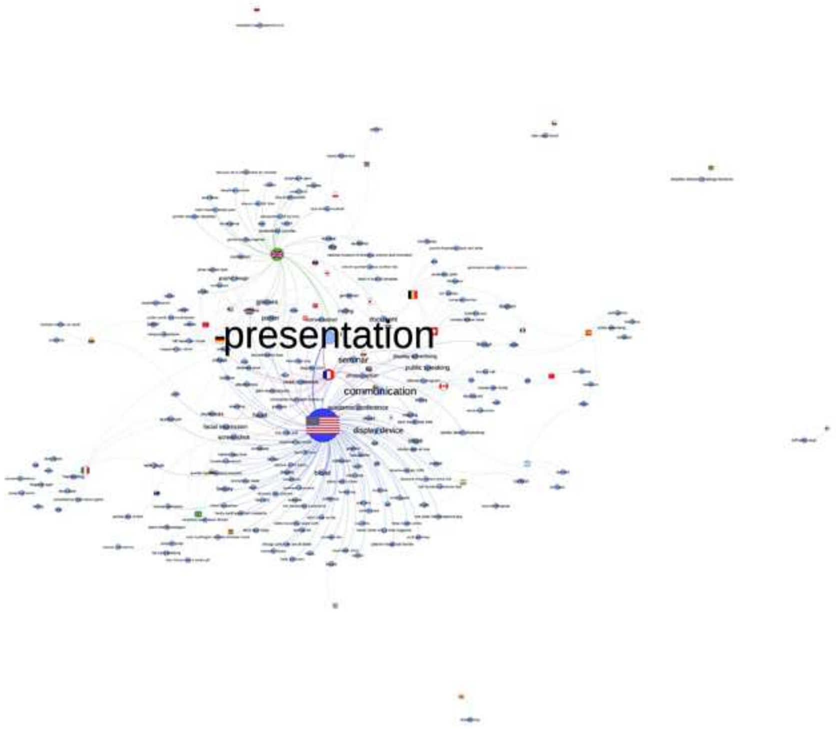

Figure 4: Bipartite network graph of Cloud Vision best guess label of 416 images and 41 countries where the images were tweeted, represented as 246 nodes and 328 directed edges. Network visualization was created using Gephi's Force Atlas 2 layout, nodes were sized by degree; Cloud Vision labels were sized by frequency; country nodes and edges were colored (US = royal blue, France = red, Great Britain = green) and the respective county flag .jpg file was rendered with Gephi ImagePreview Plugin. Best guess label nodes were colored light blue.

By zooming into the United States cluster (Figure 5), the Cloud Vision labels that describe the images tweeted in the country are expressions that can describe pictures related to pop and celebrity culture, which aligns with the Twitter hashtags - images network visualization zoom into the US region (Figure 2).

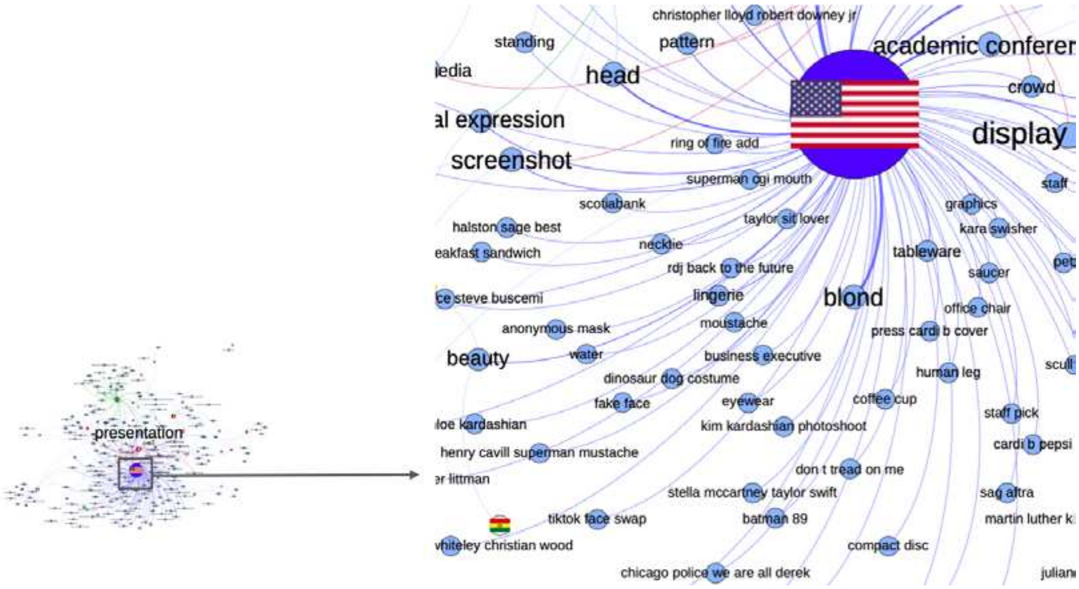

Figure 5: Zoom into the bipartite network of Cloud Vision best guess label and countries where the images were tweeted from. The zoom is centered on the United States node. The figure shows labels related to brands, celebrities and pop culture images' description, for example: cardi b pepsi, batman 89, blond, cardi b cover, beauty, lingerie, kim kardashian photoshoot, stella mccartney taylor swift etc.

3.3 How to read image networks without rendering the image files?

Considering the limitation of working with a small dataset of geolocated images, I decided again to combine data collected from different sources. For this, I have compiled a table containing information on how Twitter users described and contextualized the images (hashtags) and from how Google Cloud Vision API defines the images (best guess labels). Then, I generated a bipartite network graph of Twitter hashtags and Cloud Vision best guess labels (Figure 6). The strategy employed was to consider the images' country information as a fixed variable to explore the combination of the hashtags users posted to tag the image on Twitter and the Cloud Vision API best guess label outputs, considering the geographic perspective.

The Zoom into the network, considering the United States network region (Figure 7), shows the relations between the hashtags that US Twitter users combined with the #deepfakes and #deepfake and the best guess label attributed by Google Cloud Vision API, providing a more contextualized image description. This is a way for reading the image files. The #deepfakes and #deepfake labels were removed from the network for better visualization since they both relate to every image represented.

The comparison between the zoom into the network of Twitter hashtags and images (Figure 2) and the zoom into the network of Twitter hashtags and Cloud Vision best guess labels (Figure 7) indicates that it is possible to achieve a similar network interpretation result by plotting the labels attributed by the Cloud Vision API instead of rendering the image files in the graph. For example, the observation of Figure 7 the less specific Cloud Vision labels (i.e. beauty, facial expression, head, collage, and blond) gives a good notion of the images that appear in Figure 2. In addition, there are more detailed images descriptions that indicate precisely the image content (i.e. 'cardi b pepsi' and 'stella mccartney taylor swift').

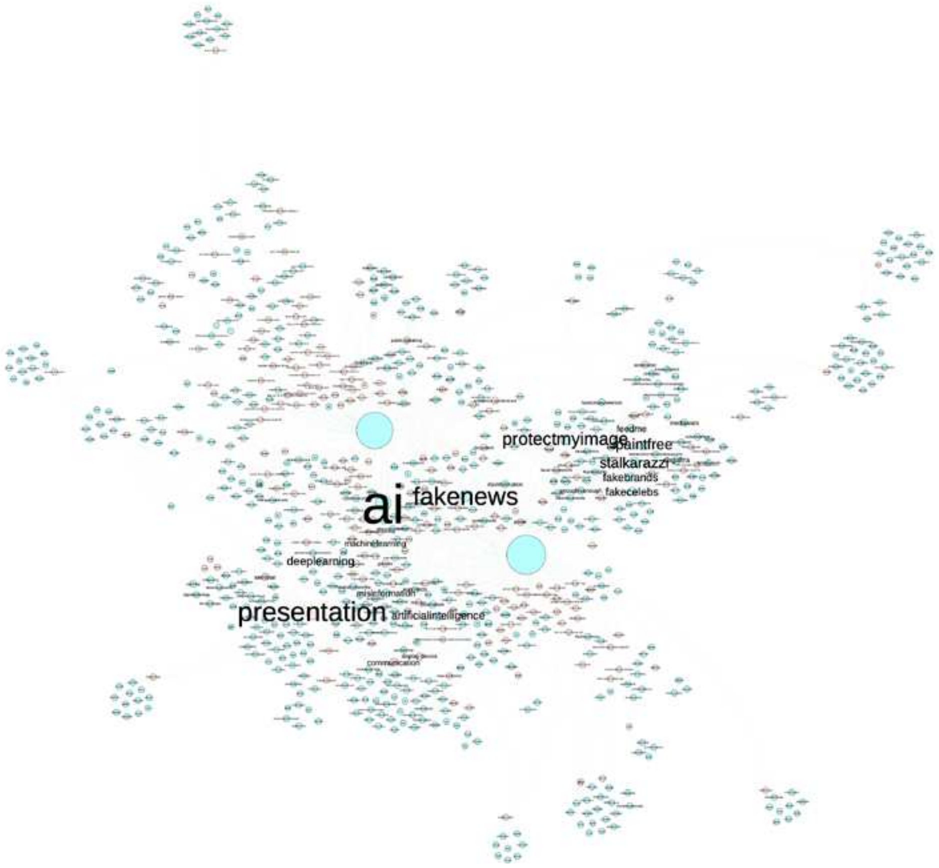

Figure 6: Bipartite network graph of Google Cloud Vision best guess label of 416 tweeted images and 710 Twitter hashtags, represented as 846 nodes and 1472 directed edges. Network visualization was created using Gephi's Force Atlas 2 layout, the two types of nodes were sized by degree and the labels were sized by frequency. For better visualization, the #deepfake and #deepfakes labels were removed, since at least one of the hashtags are in all tweets of the working dataset. Twitter hashtags nodes were colored light blue and Google Cloud Vision API best guess label nodes were colored light pink. The network edges represent the images. The #deepfake and #deepfakes labels were erased from the graph for better visualization.

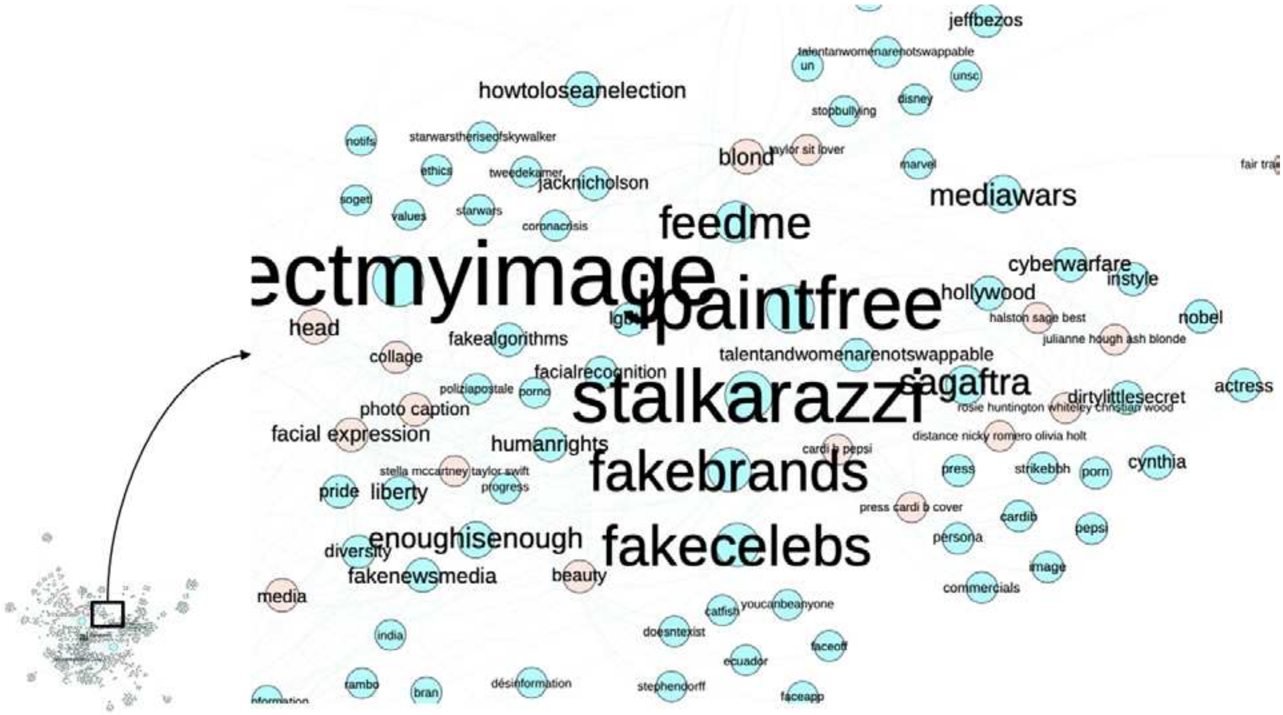

Figure 7: Zoom into the US region of the bipartite network of Twitter hashtags and Cloud Vision best guess label where the images represent the edges. This figure shows the connections between the hashtags tweeted to characterize images by Twitter users and the Google Cloud Vision description of the image.

An important consideration about the relevance of the context provided by the Twitter hashtags, is that without this information and based only on the Cloud Vision tags, this data would not be understood as a group of images shared under the topic #deepfakes. Even with the removal of the #deepfake and #deepfakes, the other hashtags tweeted in combination provide context locating the images under the 'deepfake' topic. For example, #protectmyimage and #fakecelebs indicate that the theme relates to fake content about celebrities and a campaign on the protection of social media profile pictures and other shared photos. In Figure 7 there are other hashtags that give context, like #fakebrands, #fakealgorithms, and #facialrecognition, among others. Although there are specific platform vernaculars for different platforms, the Twitter hashtags can be considered an example of a form of expression established by convention through the community (Gibbs et al., 2015) and they are an appropriate fit to give context to images.

Analyzing the bipartite network graphs (Figures 1-3, and 5-7), it is possible to point out that the discourse constructed surrounding deepfakes has a similar structure in computer vision API images' description and in users' choice of hashtags. In this way and based on the existence of specific vernaculars for images shared on Twitter in specific countries, I have noticed that it is possible to adopt this methodological strategy when trying to understand a (very large) set of images that are geographically located. In other words, these findings indicate that computer vision APIs can be used to describe a set and extract relational structures and common topics of these images without having to render and visualize them. Google Cloud Vision best guess labels give meaning to images while hashtags provide context to the pictures (Cloud Vision API did not return any web entity result relating the images to deepfake).

3.4. Using Twitter's geolocated images and Google Vision best guess label as networks: a method protocol

I have proposed a methodological strategy (shown in Figure 8) to use network analysis in the description of geolocated images utilizing user description data (Twitter hashtag) and computer vision API data (best guess label Cloud Vision API classification).

The interpretation of the images-hashtags geolocated network should be carried out by mixing social and computational elements. As Marres (2020) indicates, in the study of situations in computational settings, the researcher should consider that social processes happening in digital environments are influenced by the technical specificities of the medium. Therefore, I propose that the methodological strategy schematized in Figure 8 applies to interpreting and situating geolocated Twitter images. Combining image artificial intelligence descriptions with Twitter hashtags contemplates "both the visual elements of an image and the contextual elements encoded through the hashtag practices of networked publics" (Geboers and Van De Wiele 2020). From a Digital methods perspective, researchers developing visual methodologies to extract information from digital platforms data shall create tools that facilitate the interpretation of this data and produce "visualizations that make complex information legible and ready for further analysis" (Niederer and Colombo 2019).

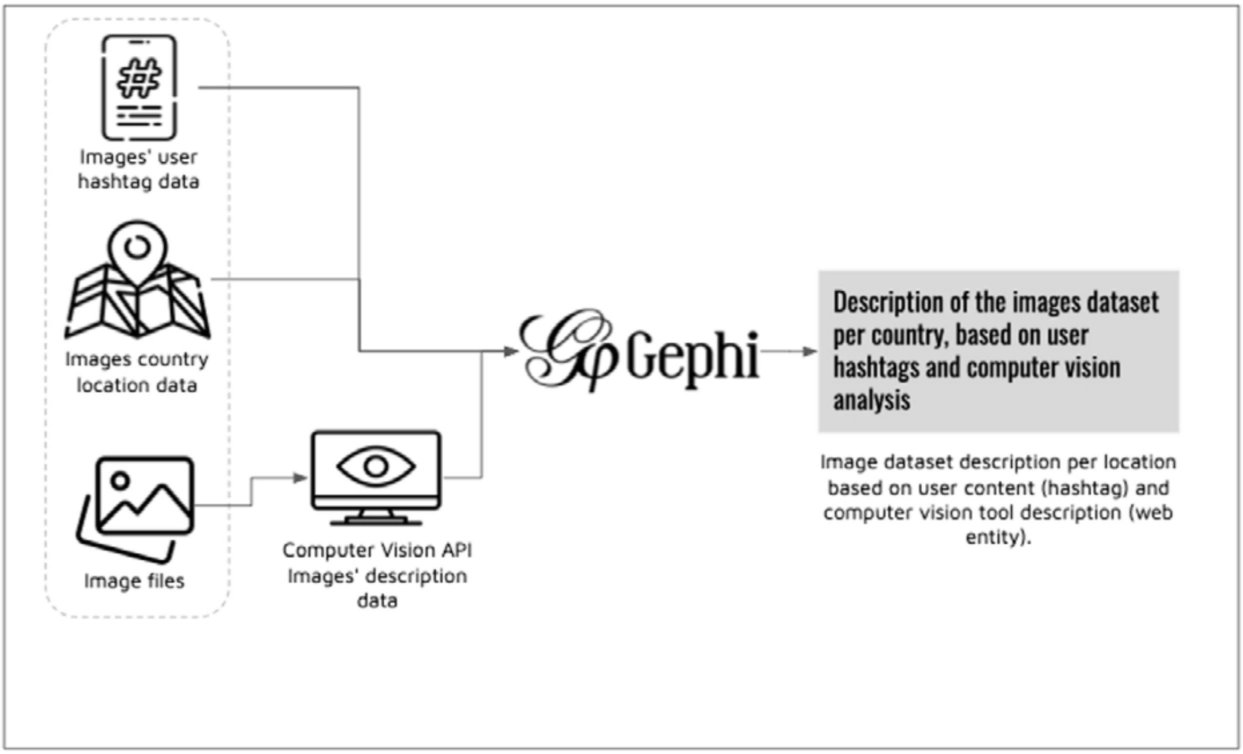
Images' user hashtag data
Images country location data
Image files
Computer Vision API Images' description data
Gephi
Description of the images dataset per country, based on user hashtags and computer vision analysis
Image dataset description per location based on user content (hashtag) and computer vision tool description (web entity).

Figure 8: Schematic methodological strategy developed to use network analysis in the description of geolocated images utilizing user description data (Twitter hashtag) and computer vision API data (best guess label Cloud Vision API classification).

4. Limitations

During the development of this work, I have considered several limitations. I generated the network analyses presented in this article starting from a dataset that contained scarce geolocated data related to the debate about deepfakes on Twitter. Considering a total of 19.713 unique images extracted from the platform, only 459 images were tagged with their respective country code metadata. Although the small amount of data available was the motivation for the development of the method presented in this article, a small set of images could be analyzed qualitatively for a more robust interpretation and to give more precise answers to the research questions.

In addition, since our sample was narrowed to unique country-coded tweeted images, the working dataset comprises a small number of images shared during four years (2017 - 2021). This lack of robust data can bias the results of what type of discourses were happening about deepfakes in every location and over these years. The absence of geographical information on the majority of tweets, nevertheless, is an important disadvantage since “the attributes of language and location are crucial for understanding the geographies of online flows of information” (Graham et al., 2014). For the dataset used in this study, only 2,3 % of tweets contained country information and it is unlikely that they form a representative sample of the broader universe of content.

Another reason for interpreting the findings related to specific country-vernaculars with attention is that the dataset is biased towards particular language spaces, potentially related to the search terms #deepfakes and #deepfake. When analyzing Twitter hashtags it is important to consider that the data can contain slang and abbreviations (Kim et al., 2013) and/or hashtags not related to the topic (i.e. call-for-action hashtags) since these are elements that are often present on user-generated content.

Furthermore, despite the visual consistency provided by computer vision (Pearce, Gaetano, 2019), an aspect of adopting artificial vision techniques is the loss of specificity in the description of the image versus the qualitative approach, which (maybe) should be more appropriate to analyze a small dataset. By admitting that reading images via different visual APIs can generate diverse outcomes, Mintz and Silva (2019) tested Google, Microsoft, and IBM computer vision services. They verified that Google Vision API tends to be more specific when describing details. However, Google Cloud Vision failed to annotate relevant elements in a war context, for example labeling a picture of a human corpse as a 'zombie' (Geboers and Van De Wiele 2020). The existence of bias in image interpretation by Google Cloud Vision or every other Computer Vision service must be taken into account. These automated annotation tools can (and probably do) reproduce cultural stereotypes and generalizations, as shown by Silva et al. (2020).

Technical limitations to be considered when using Google's automated image reading are the inconsistency of Cloud Vision API results when analyzing noisy images (Hosseini, Xiao, and Poovendran 2017) and the susceptibility of the service to black-box attacks (Brunner et al. 2019; Ilyas et al. 2018), ways of altering neural network-based classifiers by insistently querying the system with a manipulated input information.

Addressing the socio-cultural bias tendency and the technical related issues is desirable to achieve more consistent automated image reading results. A possibility is to use a combination of computer vision services to interpret the results. Finally, there is the need to use larger datasets to apply and validate this method of reading images in bipartite networks without rendering the files.

5. Final considerations and future agenda

The environment of intensive hands-on research maintained in data sprints provides the opportunity for the development of new research strategies and for adapting and studying new forms of applying data extraction and analysis tools and techniques. The ideas exchanges, the interdisciplinary discussions, and collaborative efforts stimulate discoveries and the creation of knowledge.

The research developed by the sub-group I participated in during the DMI Summer School 2021 data sprint aimed to combine Twitter and computer vision data to observe how the debate about deepfakes on Twitter traveled the world over time. So, after following the methodological steps described in this paper and analyzing the obtained results, we were able to infer that deepfakes related content has spread around the world and is concentrated in the United States. The approach we adopted was able to expose location-based - and potentially cultural and/or language-based - differences and similarities in the discourses surrounding deepfakes. Even considering the loss of specificity and nuances as well as the vagueness of semantic interpretation when relying on computer vision techniques, the combination of Cloud Vision labels with user-created data and Twitter metadata (geolocation) added a semantic layer to our analysis.

We have found that the images representing the conversation about the deepfake topic on Twitter link specific themes and actors, considering a geographic location. When analyzing the US results, the ensemble formed by pictures of celebrities and hashtags like #fakecelebs, #fakenewsmedia, and #stalkarazzi emerge, which is consistent with other research. Westerlund (2019) argues that celebrities are easy targets for fabricated content since they have an abundance of free video, image, and audio content circulating over the internet and that content availability facilitates deepfakes creation. Dasilva et al. (2021) analyzed the debate around deepfakes on a Twitter network. They have shown that journalists and media are bridging information nodes and that celebrities figures are among the most referenced and virialized users in the networks.

The Cloud Vision labels and countries' bipartite network analysis indicates users tweet images related to deepfakes that were described by the computer vision tool as specific vernaculars for different geographical locations. Coming out of this conclusion, I developed a methodological framework for reading networks of images without rendering the files in the graph. In other words, I propose that combining the image description and context given by a digital platform user (Twitter hashtags) and the image definition given by a computer vision API (Google Cloud Vision best guess label) is an efficient way to characterize geolocated images. Applying this method can serve as a solution for computational memory issues while creating networks of massive sets of images since the network analysis software (in my case, Gephi) can experience computer memory-related difficulties managing large volumes of data. This methodological path can be thought out for exploratory research since it could generate insights and ideas serving as a starting point to additional analyses.

As I mentioned before, further testing and validation of this methodological path with larger datasets are required. Another future agenda raised by this study is to look beyond Twitter and apply this method with data extracted from other platforms since "the affordances and performances that constitute a vernacular are not necessarily specific to a platform" (Gibbs et al., 2015). For example, an application for this method could be the analysis of a large set of pictures shared on Instagram in a country-specific context by combining hashtags users post as a description of the images with computer vision retrieved image information. Finally, since some techniques allow combining computer vision services interpretations in network analysis (Silva et al. 2020), future experiments merging images annotations from multiple sources are encouraging to achieve more robust results and reduce bias.

Acknowledgments

I am grateful to all those who have contributed to the development of this work. I thank Professor Richard Rogers for all the excellent work developed by the Digital Methods Initiative, for conducting this research experience memorably, and for sparking my interest in this topic; Lucia Bainotti, Sarah Burkhardt, Gabriele Colombo, Janna Joceli

Omena, and Jason Chao for facilitating the Mapping deepfakes with digital methods and visual analytics project, for their assistance during every step of the work, and for stimulating rich discussions and debates during the DMI Summer School; Martin Roth and Flavia Saxler for accepting the challenge of participating in the Geographical mapping with Twitter Hashtags and Google Vision entities sub-group, for all the knowledge exchanged during the process, and for all the hard work we have done together. In addition, I would like to thank the two anonymous reviewers as well as this special edition editors for their valuable suggestions on the manuscript.

References

Bimber, B.; Gil de Zúñiga, H. (2020). The Unedited Public Sphere. *New Media & Society, 22*(4):700–715. doi: 10.1177/1461444819893980.

Borgos, J. B. (2021). country-flags Available at: https://github.com/hampusborgos/country-flags (Accessed: 08 Jul. 2021).

Brunner, T.; Diehl, F.; Truong Le, M.; Knoll, A. (2019). "Guessing Smart: Biased Sampling for Efficient Black-Box Adversarial Attacks". Paper presented at the *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, South Korea. doi: 10.1109/ICCV.2019.00506.

Chao, T. H. J. (2021). *Memespector GUI: Graphical User Interface Client for Computer Vision APIs (Version 0.2)* [Software]. Available from https://github.com/jason-chao/memespector-gui. (Accessed: 20 May 2022).

Chao, J.; Omena, J.J. (2021). *Networks of Image Description*. Available at: https://docs.google.com/document/d/1ccBH691tW-6jVTrjQtMjd2aDTalZolIbRBtMqIGISIg/ (Accessed: 05 Dec. 2021).

Chesney, R.; Citron, D.K. (2018). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *SSRN Social Science Research Network*. Rochester, NY. doi: /10.2139/ssrn.3213954.

Chezsick (2017). "Gephi 0.9.1 Can Never Increase the Memory · Issue #1705 · Gephi/Gephi." *GitHub*. Retrieved April 8, 2022 (https://github.com/gephi/gephi/issues/1705).

Cole, S. (2017). AI-Assisted Fake Porn Is Here and We're All Fucked. Vice. Available at: https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn (Accesse:d 05 Dec. 2019).

Dasilva, J. P.; Ayerdi,K. M.; Galdospin, T. M. (2021). Deepfakes on Twitter: Which Actors Control Their Spread?. *Media and Communication, 9*(1): 301–312.

Geboers, M. A.; Van De Wiele, C. D. (2020). Machine Vision and Social Media Images: Why Hashtags Matter. *Social Media + Society, 6*(2). doi: 10.1177/2056305120928485.

Gibbs, M.; Meese, J.; Arnold, M.; Nansen, B.; Carter, M. (2015). #Funeral and Instagram: death, social media, and platform vernacular". *Information, Communication & Society*, 18(3), 255–268. doi: 10.1080/1369118X.2014.987152.

Ghost (2016). "Gephi Memory Issue · Issue #1609 · Gephi/Gephi." *GitHub*. Retrieved April 8, 2022 (https://github.com/gephi/gephi/issues/1609).

Google (2020). *Google.Apis.Vision.v1.Data.WebDetection Class Reference*. Available at: https://developers.google.com/resources/api-libraries/documentation/vision/v1/csharp/latest/classGoogle_1_1Apis_1_1Vision_1_1v1_1_1Data_1_1WebDetection.html#afddf64d68931805b23d9c863dd47b9c7 (Accessed: 05 Dec. 2021).

Graham M.; Hale S. A.; Gaffney D. (2014). Where in the World Are You? Geolocation and Language Identification in Twitter. *The Professional Geographer, 66*(4): 568-578, doi: 10.1080/00330124.2014.907699.

Ilyas, A.; Engstrom, L.; Athalye, A.; Lin, J. (2018). "Black-Box Adversarial Attacks with Limited Queries and Information." In *Proceedings of the 35th International Conference on Machine Learning* (pp. 2137–46). PMLR. https://proceedings.mlr.press/v80/ilyas18a.html.

Jung, K. (2021). PicArrange -- Visually Sort, Search, and Explore Private Images on a Mac. arXiv:2111.13363

KallyopeBio (2018). "Gephi UI Nearly Unresponsive with Large Network (100% CPU Utilization on One Core) · Issue #1947 · Gephi/Gephi." *GitHub*. Retrieved April 8, 2022 (https://github.com/gephi/gephi/issues/1947).

Kearney, M.W. (2017). rtweet. Available at: https://github.com/mkearney/rtweet.download (Accessed: 23 Mar. 2018)

Kietzmann, J.; Lee, L.W.; McCarthy, I. P.; Kietzmann, T. C. (2020). Deepfakes: Trick or Treat?. *Business Horizons, 63*(2):135–46. doi: 10.1016/j.bushor.2019.11.006.

Kim, A.E.; Hansen, H.M.; Murphy, J.; Richards, A.K., Duke, J.; Allen, J.A. (2013). Methodological Considerations in Analyzing Twitter Data. *JNCI Monographs*, 47: 140–146. doi: 10.1093/jncimonographs/lgt026.

Mack, D. (2018, April 17). "This PSA About Fake News From Barack Obama Is Not What It Appears." *BuzzFeed News*. Retrieved April 8, 2022 (https://www.buzzfeednews.com/article/davidmack/obama-fake-news-jordan-peele-psa-video-buzzfeed).

Marres, N. (2020). For a Situational Analytics: An Interpretative Methodology for the Study of Situations in Computational Settings. *Big Data & Society, 7*(2). doi: 10.1177/2053951720949571.

Mintz, A.; Silva, T. (2019). *Interrogating Vision APIs*. Lisboa: Universidade Nova de Lisboa | NOVA FCSH | iNOVA Media Lab.

Niederer, S.; Colombo, G. (2019). Visual Methodologies for Networked Images: Designing Visualizations for Collaborative Research, Cross-Platform Analysis, and Public Participation. *Diseña, 14*:40–67. doi: 10.7764/disena.14.40-67.

Omena, J.J. (2019). Introdução: O que são Métodos Digitais? In Janna Joceli Omena (Ed.), *Métodos Digitais: Teoria-Prática-Crítica*. Lisboa: ICNOVA.

Pearce, W.; Colombo, G.; De Gaetano, C. (2019). Using computer vision to see Google's visual vernacular of climate change (2008-19). https://research.hva.nl/en/publications/using-computer-vision-to-see-googles-visual-vernacular-of-climate.

Rogers, R. (2009). *The end of the virtual: Digital Methods*. Inaugural lecture, University of Amsterdam. doi: 10.5117/9789056295936.

Rogers, R. (2021). Visual media analysis for Instagram and other online platforms. *Big Data & Society*. doi: 10.1177/20539517211022370.

Silva, T.; Mintz, A.; Omena, J.J.; Gobbo, B.; Oliveira, T.; Takamitsu, H. T.; Pilipets, E.; Azhar, H. (2020). Investigando mediações algorítmicas a partir de estudo de bancos de imagens. *Logos, 27*(52):30. doi: https://doi.org/10.12957/logos.2020.51523.

Stevens, G.; Rohde, M.; Korn, M.; Wulf, V. (2018). Grounded Design: A Research Paradigm in Practice-Based Computing. In V. Wulf; V. Pipek; D. Randall, M. Rohde; K. Schmidt; G. Stevens (Eds): *Socio Informatics –A Practice-based Perspective on the Design and Use of IT Artefacts*. Oxford University Press, Oxford. doi: 10.1093/oso/9780198733249.003.0002.

Tucci, G.; Roth, M.; Saxler, F. (2021). Geographical mapping: Twitter hashtags and Google Vision web entities. *Digital Methods Initiative*. Available at: https://wiki.digitalmethods.net/Dmi/WinterSchool2021Deepfakes#A_5.3.3._Geographical_mapping:_Twitter_hashtags_and_Google_Vision_web_entities (Accessed: 05 Dec. 2021).

Vaccari, C.; Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society, 6*(1):205630512090340. doi: 10.1177/2056305120903408.

Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review, 9*(11): 40–53. doi: 10.22215/timreview/1282.

Xue, C. (2012). gephi-plugin-image-preview. Available at: https://github.com/chrisxue815/gephi-plugin-image-preview (Accessed: 08 Jul. 2021).