2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Identifying and Characterizing Alternative News Media on Facebook

Samuel S. Guimarães*, Julio C. S. Reis*‡, Lucas Lima*, Filipe N. Ribeiro†,
Marisa Vasconcelos§, Jisun An¶, Haewoon Kwak¶, Fabrício Benevenuto*
*Universidade Federal de Minas Gerais (UFMG), Brazil, †Universidade Federal de Ouro Preto (UFOP), Brazil
‡Universidade FUMEC, Brazil, §IBM Research, Brazil, ¶Singapore Management University, Singapore
{samuelsg, julio.reis, lucaslima, fabricio}@dcc.ufmg.br, filipe.ribeiro@ufop.edu.br,
marisaav@br.ibm.com, {jisun.an, haewoon}@acm.org

***Abstract*—As Internet users increasingly rely on social media sites to receive news, they are faced with a bewildering number of news media choices. For example, thousands of Facebook pages today are registered and categorized as some form of news media outlets. This situation boosted the so-called independent journalism, also known as alternative news media. Identifying and characterizing all the news pages that play an important role in news dissemination is key for understanding the news ecosystems of a country. In this work, we propose a graph-based semi-supervised method to measure the political bias of pages on most countries and show the political split of the alternative media, mainstream media, and public figures pages. We validate our method using the publicly available U.S. dataset and then apply it to Brazilian pages, where we found a larger number of right-wing pages in general, except for alternative news media.**

***Index Terms*—Facebook, Social Media, Public Figures, Alternative Media, Mainstream Media, Semi-supervised Learning**

I. Introduction

Today many people rely on social media to satisfy their daily news diet. Nearly 68% of the U.S. adults get informed about the news primarily on social media websites, according to a recent survey of Pew Research Center [1]. This situation creates a shift in how news is consumed and produced, lowering the barrier to entry, and therefore promoting independent journalism, such as citizen journalism [2]. That shift was indirectly measured by a recent work that counted 20,448 self-reported pages of U.S. news outlets located on Facebook [3].

This independent journalism, also called **alternative news media**, still generates some debates about firm definitions, at times challenging the definition of journalism [2]. One article conceptualized key dimensions where this journalism differs from the traditional one of the so-called **mainstream media**, which are *the producers*, *the content*, *the media organizations* formed, and *the media systems* where it lives [4]. A unique type of these news *producers* are **public figures** and political entities, that can also replace traditional news, similar to U.S. President Donald Trump's use of Twitter. Alternative media are gaining considerable space and power in the last years, sometimes using **political bias** as fuel [5, 6].

Understanding alternative media pages and their audiences can help us find trends in how they affect different communities and assessing their societal impacts. Despite their importance, studying alternative media on social media platforms is still challenging as it requires extensive manual efforts in identifying them in the first place. For example, while Facebook is the most used platform for news reading in Brazil, studying its ecosystem is limited, where case studies of existing groups predominate [6, 7].

IEEE/ACM ASONAM 2020, December 7-10, 2020
978-1-7281-1056-1/20/$31.00 © 2020 IEEE

In Brazil, several alternative media outlets have emerged during both left and right-wing governments, mostly mixing activism and reporting, raising questions about their political accountability and compromise with the truth. In this context, our work presents a two-fold contribution: 1) creation and validation of a methodology to identify and measure the political bias of Facebook pages for a given country, and 2) a characterization of the bias of the three actors cited: mainstream media, alternative media, and public figures, in Brazilian Facebook pages.

We use the Facebook Marketing API to identify both news outlets or politics related pages. Later, to classify the pages in mainstream or alternative, we use the dimensions of *producers* and *media organizations* proposed by [4]. We consider a page as alternative media if it does not represent an outlet registered in any official press organizations. That is, it is alternative if it cannot be confirmed as mainstream media or public figure. Finally, we characterize the pages by generating an ideological bias score based on a graph-based semi-supervised learning.

Altogether, this work presents a methodology to identify the political bias of Facebook pages, which shows comparable performance with existing methods, and a case study of the current Brazilian media ecosystem.

II. Related Work

We review related work along three distinct dimensions: (i) identifying online news pages, (ii) polarization and relationship graphs, and (iii) political bias measurement.

Working to find online news pages, some studies found alternative news outlets by screening the most popular links on Facebook groups and pages [8], while others search for alternative media common narratives [9]. Especially, Ribeiro et al. [3] used recommendations from Facebook Marketing API to create a snowball process collecting all recommendations for U.S. pages. We have a similar approach using a different tool from the Facebook Ads platform.

Just as important, researches considering political polarization and relationship graphs mostly focus on Twitter [9, 10]. Conover et al. [11] is one example, having compared different methods for measuring the political alignment of Twitter users, including text, hashtags, and label propagation analysis using both mentions and retweets graphs. This retweet graph-based approach inspired part of our graph approach.

For assessing ideological bias for news outlets, we introduce four studies that were compared with our method. First, the Pew Research Center [12] analyzed the audiences from **36** news outlets by interviewing 2,901 people and asking them what media they know, which one they read, their political self-identification, and their trust in the media. This analysis allowed them to make a diagnostic about how different political leanings affected the perception of the news.

Budak et al. [13] used content analysis to identify the overall ideological bias of **15** major U.S. news outlets by compiling 803,146 published stories over an entire year, and latter using 749 human judges to classify 10,502 specific political articles. The overall leaning was measured using these articles, and the results showed little difference in the coverage by outlets of different bias, except in scandals.

Bakshy et al. [8] also used Facebook data to calculate the ideology from **500** websites with links shared on the platform. It used the ideological affiliations from 10.1 million users that declared their bias to classify 226,000 URLs from an initial seven million shared by them over six months. They showed that the content on social media could cross ideological lines and reach people from the opposite perspective.

Finally, Ribeiro et al. [3] used the Facebook Marketing API[1] to get information on the proportion of users identified within different parts of the political spectrum, then calculated a bias score for **20,448** American media outlets. They provide a demographic analysis of the U.S. audience, especially using the demographic division of the users in Very Conservative, Conservative, Moderate, Liberal, and Very Liberal, to generate the bias score. However, their approach cannot be exploited for countries other than the U.S. because the political leaning of a pages' audience is not readily available.

As the above studies show, calculating the ideology of pages by both the audience or relations between them is useful and accurate. We followed this trend, proposing a new method that can be extrapolated to any country with sufficient adherence to Facebook. A few articles analyzed the Brazilian media ecosystem with minor conclusions about the news consumption in the country [6, 7, 14]. However, to the best of our knowledge, there is no previous work that compared alternative media, mainstream media, and public figures in the current context.

III. Methodology

A. Selecting Facebook Pages

The first step of our process was finding all the relevant Facebook pages to analyze. We focus on pages that are related to politics, which includes Brazilian public figures and news outlets reporting political news. One main challenge here is the lack of ground-truth political bias for Brazilian media.

Inspired by the method proposed by Ribeiro et al. [3], we make use of the tools from the Facebook Ads platform to classify the political leaning of a given Facebook page. First, we use the Facebook Marketing API that allows the creation and management of ads on Facebook by specifying the target audience with attributes such as age, location, gender, and *interests*. These interests are an extensive set of topics that Facebook infers from the engagement of the users. However, as the political demographic feature is only available for U.S., we use the Audience Insights[2] that suggests related pages of a given interest through the "Page Likes" menu.

By using these two features, we iteratively collect a list of relevant pages as follows: 1) compile a small number of "seed" pages that were, preferentially, manually curated; 2) get the associated interests for each page, 3) use these returned interests to find related pages, and 4) go back to (2) until no new page is suggested. In step 3, we consider pages that are one of 13 relevant categories[3]: Public Figures, Politicians, Government Officials, Authors, Political Organizations, Political Parties, News & Media Websites, Media/News Companies, Broadcasting & Media Production Companies, Magazines, Journalists, TV Programs (News related) and Newspapers.

B. Inferring Political Leaning

Our proposed method assesses the political bias of a Facebook page by utilizing audience interaction information, not being limited to U.S. pages like previous work [3].

For example, consider two pages with associated interests, a and b. Given an interest in page a, the Audience Insights tool provides a list of associated pages, including b. For each related page, the tool provides three metrics: Monthly Active People (MAP), Audience, and Affinity score, shown in Figure 1. In our example, MAP is the number of monthly active users of page b. The audience is the number of users who are active on page b, given the interest in page a. Then, the affinity score measures how likely a user with the interest in a is to like page b compared to a random user.

Fig. 1: Facebook Audience Insights tool.

While the affinity score allows us to compare related pages of one interest, it is not straightforward to compare in between

[1] developers.facebook.com/docs/marketing-apis

[2] https://www.facebook.com/ads/audience-insights [3] When a page is created, a pre-defined category can be assigned to that page.

interests. We thus propose a new normalized affinity score, $\mathcal{A}$, between pages (e.g., for pages a and b), that is calculated based on the MAP and Audience as:

$$\mathcal{A}_{a,b} = (\frac{Audience_{a,b} + Audience_{b,a}}{MAP_a + MAP_b})$$

Basically, $\mathcal{A}$ is the sum of the number of people who are interested in one page and like the other, and vice-versa, divided by the sum of the number of active users of both pages. It is important to note that all the audience of one interest is not equal to the MAP of the related page. For instance, a person may like a fan page of one celebrity but not the official page. That person is still counted as interested in that celebrity.

With this new affinity score, we compute the political leaning. For that, we construct a graph using the score and use a semi-supervised learning (SSL) method to propagate the ideological bias of some known pages to all others in the following steps:

1) For each page found as interest, we calculate our new affinity score $\mathcal{A}$;
2) We create an undirected weighted graph whose nodes are pages and edges are established when one page was found as related to the other on the Audience Insights, with edge weight as the complement of affinity: $w(u, v) = 1 - \mathcal{A}_{u,v}$;
3) We apply the Floyd-Warshall algorithm [15] to find the distance between all pairs of nodes of that graph;
4) We verify which pages from the selected 13 categories can be identified as right-wing or left-wing and then label them as such;
5) We use graph-based SSL method to classify the remaining pages as left or right, passing the graph as a parameter;
6) We define the ideological leaning as the probability of a page being classified as right-wing minus the likelihood of the page being left-wing, giving a skew between -1 (left) to 1 (right). As we use cross-validation, we actually take the average of this bias on all folds.

For step (5), we experiment with three existing graph-based SSL algorithms: classic label propagation (LP) [16], label propagation with smooth function classes (Smooth LP) [17], and spectral graph transducer (SGT) [18]. As the baseline method, we use the K-nearest neighbors algorithm (KNN) [19] using only the known part of the graph in supervised learning. We perform 10-fold cross-validation and report the area under the ROC curve (AUC) for all instances. In the subsequent validations, we analyze our methodology as a whole, using the ideological leaning of U.S. pages calculated from the average result of the ten folds for each algorithm and comparing our results to the other four related works [3, 8, 12, 13].

IV. Evaluation of our approach

A. Comparing Graph-based SSL Algorithms

To identify the best graph-based SSL algorithm to use in Step (5) of our methodology, we compare four well-known graph-based SSL methods in the task of classifying Facebook pages as either left or right. Later, we use the best algorithms to calculate an actual political bias score and compare the results to four baseline data sets of U.S. news outlets.

For both tests, we created a U.S. Facebook page data set to compare our method with previous work, as they were focused in the United States. First, we compiled a list of seed pages for step (1) of our method. We use the list of 15 news outlets created in [13] as our starting list. After ten iterations of the snowball described, we found 832 pages that had an interest in Facebook. Among them, we identified 136 public figures and political entities, almost evenly split into 65 left-wing and 71 right-wing pages by their political self-identification. We reserved ten test sets of 83 pages for each fold and proceeded with a 10-fold cross-validation. Table I shows the results of each tested model, for both training and test sets.

TABLE I: AUC scores for different SSL methods with 95% confidence intervals.

	AUC (Train)	AUC (Test)
LP	0.9546 [0.9414-0.9679]	0.8440 [0.8091-0.8790]
Smooth LP	0.9509 [0.9298-0.9719]	0.8926 [0.8718-0.9133]
SGT	0.9615 [0.9462-0.9768]	**0.9482** [0.9290-0.9674]
KNN	**1.0000** [1.0000-1.0000]	0.9122 [0.8806-0.9437]

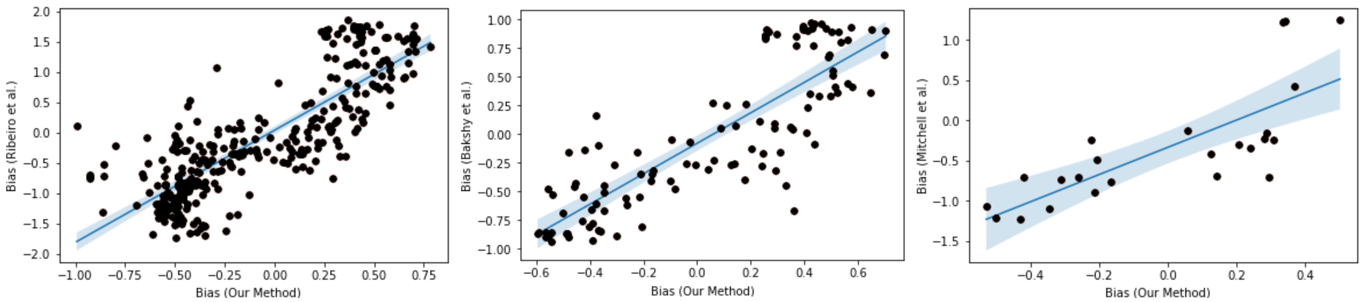
We find that SGT has the best result on average for the test set, beating the KNN baseline. Smooth LP came in second, being statistically equivalent to the baseline, while LP was the worst. In the training set, KNN was better than all others. However, as the training set for KNN was only composed of labeled data, it effortlessly learned the classes, making it a less meaningful comparison than the test data.

B. Comparing our Proposed Method with Previous Work

We now take the three algorithms that were satisfactory in the previous section (SGT, Smooth LP, and the KNN Baseline) to compare the results of using them as step (5) of our methodology with four well-known data sets of political bias. The data set used, created by our snowball process, includes 24 pages from [12] (75% out of total 32), 111 (22.2%) from [8], 14 (93.33%) from [13] and 302 (1.48%) from [3].

1) Comparing Algorithms for the Complete Task: To compare the results from the graph-based SSL methods with the four ground-truth data sets, we use the Pearson correlation coefficients, shown in Table II. We observe that almost all methods had statistically equivalent results, with statistically significant differences only on Ribeiro et al. [3] data. In this case, Smooth LP and KNN have the highest correlations, and SGT is worse than all other options. As KNN and Smooth LP were also satisfactory in the classification task, the SGT advantage becomes less relevant as it only won in step (5) alone. Additionally, considering that the KNN uses only the labeled data, we can deem the **Smooth LP** the best method, as it uses semi-supervised learning, training with most of the graph.

2) Comparing our Method to other Methodology: After establishing the best algorithm, we now compare the results



(a) Ribeiro et al. [3] (b) Bakshy et al. [8] (c) Mitchel et al. [12]

Fig. 2: Comparison of our method to the baseline political leanings from methods based on audience metrics.

TABLE II: Pearson's r for each combination of political leaning baseline and graph-based SSL method.

	Smooth LP	SGT	KNN
Mitchell et al. [12]	**0.7642** [0.5216-0.8925]	**0.8193** [0.6212-0.9190]	**0.8091** [0.6022-0.9141]
Bakshy et al. [8]	**0.8353** [0.7686-0.8841]	**0.8483** [0.7863-0.8934]	**0.8204** [0.7485-0.8733]
Budak et al. [13]	**0.6267** [0.1440-0.8685]	**0.6616** [0.2019-0.8824]	**0.6414** [0.1681-0.8744]
Ribeiro et al. [3]	**0.8225** [0.7823-0.8559]	0.6266 [0.5528-0.6906]	**0.8263** [0.7868-0.8590]

of our entire methodology with the ground truth data sets. Figure 2 and 3 depict how similar our bias scores computed with the Smooth LP algorithm are to the ground truth data sets. We see that the results of our method are highly correlated with the results of those three audience-based data sets (Person's $r = 0.8$ on average). Notably, the sets that also used Facebook data [3, 8] had narrower confidence intervals. Meanwhile, Mitchell et al. [12] had a slightly worse confidence interval, with the lower bound of its correlation being as low as 0.5216, probably because audience bias was assessed using a survey instead of Facebook data, being less comparable to our strategy. Following that trend, the data set that measured the political bias of news stories by showing them to Amazon Mechanical Turk human judges instead of using Facebook [13] had the largest confidence intervals and the worst correlation with the results of our method, even after using their list of pages for our starting list in step (1) of our method. This problem was equally damaging for all algorithms. We theorize that the number of outlets, together with the content-based scores from human labeling, generated this lower performance. Unfortunately, the paper does not provide any rater reliability metric (e.g., Kappa score) which makes it harder to analyze other possible causes of that discrepancy.

Nonetheless, the fact that our method performed well with data from audience analysis is an indication of how our method is reliable compared to other similar methods.

V. The Brazilian Alternative News Landscape

A. Our Brazilian Dataset

With our method validated, we now apply the same methodology to analyze Brazilian pages. As seed pages, we used

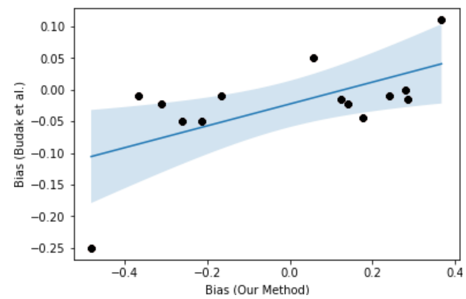


Fig. 3: Comparison of our method to the baseline political leanings from Budak et al. [13], based on media content.

the 21 pages of Brazilian news outlets with diverse political bias from Moretto and Ortelado [7] to collect Brazilian pages. In total, we found 156 pages[4], with 36 public figures and political entities, which are identifiable as left-wing (19) or right-wing (17) pages. We used these 36 pages as our labeled data. Using the Smooth LP algorithm as the step (5) of our method, described in Section III-B, the test set AUC and its 95% confidence interval was 0.9875 [0.9686-1.0]. With this classification, we correctly identified all the ideological leanings of the "seed" pages from the original article [7].

B. Calculating Political Polarization of Brazilian Pages

To further analyze our calculated bias, we divided the range of [-1,1] of our score into three parts to represent **Left**, **Center**, and **Right** political leanings. To accomplish this, we used the standard deviation (δ) of the bias from the ten folds of the cross-validation, assigning data to the correct positions based on the sign of average score adding and subtracting δ. If it is negative in both cases, we labeled it **Left**. If it stays positive, we labeled it **Right**. It is **Center** otherwise. To better understand how the alternative news media differ from other types of news media, we also label the collected pages by three types: public figure, mainstream media, and alternative news media. We grouped all politicians and public figures in the **public figure** category, and we classified the Journals, Websites, TV, Radio, and Magazines as **mainstream media** if

[4] The list of all pages and other additional material is available in https://homepages.dcc.ufmg.br/~samuel.guimaraes/ASONAM2020

they had a registry in any Brazilian official press organization[5]. If there was no registry, we considered them as **alternative news media**.

Table III shows the distribution of pages by their political bias and types. We see that most alternative media outlets are classified as left-wing, while mainstream media outlets are primarily right-wing. A possible explanation for this polarization is that our method collects pages that also have *interests* in the Facebook Ads platform. As pages are added as *interests* based on user interactions, most alternative media we found are from the time of the previous left-wing governments. Meanwhile, some big mainstream outlets have more center-right positions, as they exist since the right-wing Brazilian military government, and are pro-business [20]. Also, using the other audience metrics from Facebook Ads for the interests related to the pages, we found audience attribute trends similar to previous work [6, 7, 14], with public figures having an older and more male audience following, and left-wing pages attracting more people self-described as having higher levels of education, again reinforcing the correctness of our measuring.

TABLE III: Overview of Brazilian Facebook pages data.

	Left	Center	Right	**Total by Type**
Alt. Media	17	5	6	28
Main. Media	14	16	29	59
Pub. Figures	30	8	31	69
Total by Bias	61	29	66	**All Pages:156**

VI. Concluding Discussion

Social media platforms have changed news consumption patterns. Alternative Media proliferates in this new environment, and together with public figures official pages, they affect public perception of affairs, sometimes having politically biased coverage. Especially, Brazil sees a surge in the usage of social networks as news-gathering tools, with great focus on Facebook. Still, little is known about their political leanings and audiences. To bridge this gap, we present a two-fold contribution: (1) a novel graph-based semi-supervised learning method to estimate the ideological bias of news pages using data from Facebook and (2) a characterization of the bias of these pages. Our methodology has an advantage in its applicability, which can be applied to any other country where Facebook marketing API is available. We tested four different learning algorithms and compared them with multiple ground-truth data sets of ideological leaning [3, 8, 12, 13]. This test showed a high correlation of our method with most of them, particularly with other Facebook-based data [3, 8].

VII. Acknowledgments

This work was partially supported by the Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 grant, by the Ministério Público de Minas Gerais (MPMG), project Analytical Capabilities, as well as grants from CNPq, CAPES, and Fapemig.

[5] We use data from the National Association of Journals (ANJ), the National Association of Magazine Editors (ANER), and the National Agency of Telecommunications (ANATEL).

References

[1] E. Shearer and K. E. Matsa, "News use across social media platforms 2018," https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/, September 2018.

[2] N. Newman, "Mainstream media and the distribution of news in the age of social media," 2011.

[3] F. Ribeiro, L. Henrique, F. Benevenuto, A. Chakraborty, J. Kulshrestha, M. Babaei, and K. Gummadi, "Media bias monitor: Quantifying biases of social media news outlets at large-scale," in *ICWSM*, 2018.

[4] K. Holt, T. Ustad, and L. Frischlich, "Key dimensions of alternative news media," *Digital Journalism*, vol. 7, no. 7, pp. 860–869, 2019.

[5] S. Kelkar, "Post-truth and the search for objectivity: political polarization and the remaking of knowledge production," *Engaging Science, Technology, and Society*, vol. 5, pp. 86–106, 2019.

[6] N. Newman, R. Fletcher, A. Kalogeropoulos, and R. Nielsen, *Reuters institute digital news report 2019*. Reuters Institute for the Study of Journalism, 2019, vol. 2019.

[7] M. Moretto and P. Ortellado, "Quanto mais velhos, mais polarizados: Perfil dos usuários que interagem com páginas de notícias no facebook," Monitor do Debate Político no Meio Digital, Tech. Rep. 1, 2018.

[8] E. Bakshy, S. Messing, and L. A. Adamic, "Exposure to ideologically diverse news and opinion on facebook," *Science*, vol. 348, no. 6239, pp. 1130–1132, 2015.

[9] K. Starbird, "Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on twitter," in *Proc. of the ICWSM*, 2017.

[10] A. Makazhanov and D. Rafiei, "Predicting political preference of twitter users," in *Proc. of the ASONAM*, 2013.

[11] M. D. Conover, B. Gonçalves, J. Ratkiewicz, A. Flammini, and F. Menczer, "Predicting the political alignment of twitter users," in *Proc. of the PASSAT and SocialCom*. IEEE, 2011.

[12] A. Mitchell, *Political polarization & media habits*. Pew Research Center, 2014.

[13] C. Budak, S. Goel, and J. M. Rao, "Fair and balanced? quantifying media bias through crowdsourced content analysis," *Public Opinion Quarterly*, vol. 80, no. S1, pp. 250–271, 2016.

[14] R. Romancini and F. Castilho, "Strange fruit: The rise of brazil's 'new right-wing'and the non-partisan school movement," *Journal of Alternative and Community Media*, vol. 4, no. 1, pp. 7–22, 2019.

[15] R. W. Floyd, "Algorithm 97: shortest path," *Communications of the ACM*, vol. 5, no. 6, p. 345, 1962.

[16] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Proc. of the NIPS*, 2004.

[17] X. Zhu, Z. Ghahramani, and J. D. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in *Proc. of the ICML*, 2003.

[18] T. Joachims, "Transductive learning via spectral graph partitioning," in *Proc. of the ICML*, 2003.

[19] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.

[20] S. A. Ganter and F. O. Paulino, "Between attack and resilience: The ongoing institutionalization of independent digital journalism in brazil," *Digital Journalism*, pp. 1–20, 2020.