

RESEARCH

Open Access



Geographic impressions in Facebook political ads

Adina Gitomer^{1*}, Pavel V. Oleinikov², Laura M. Baum³, Erika Franklin Fowler⁴ and Saray Shai⁵

*Correspondence:
adinagitomer15@gmail.com
¹ Network Science Institute,
Northeastern University,
Boston, MA, USA
Full list of author information
is available at the end of the
article

Abstract

Online political advertising is becoming increasingly popular as political campaigns recognize the utility of social network platforms, like Facebook, for reaching and engaging with voters. Yet, contrary to the wealth of information about campaign advertising on TV, little is known about advertising online, as comprehensive data only recently became available to scholars. Moreover, the newly available data is often aggregated, incomplete, and imprecise. Here, we present an analysis of Facebook political ad data, supplemented with funding-related meta-data obtained through human coding and a partnership with the Center for Responsive Politics. Through computational tools—namely, network analysis—we aim to use this data to describe and categorize political ad funding behavior on Facebook. Specifically, we focus on the geographic concentration of ads, and discover that most ads reach an audience in a single geographic region (i.e., U.S. state) or in a wide range of regions, and very few reach an audience spanning a small number of regions. We use this observation to partition funding entities into three groups based on their relationships to regionally-concentrated ads. We then examine the differences between these groups via bipartite networks connecting funding entities to their geographic audiences, as well as content they support. Our findings reveal that geographic impressions play an important role in online political advertising, and can be used to classify funding entities. As a result, this study represents a step toward ensuring political funding transparency and demystifying online political advertising more broadly.

Keywords: Political advertising, Online campaigns, Bipartite networks, Transparency in politics

Introduction

The online political advertising terrain remains largely underdeveloped, especially in comparison to the large body of work describing television advertising (Ansolabehere and Iyengar 1995; Geer 2006; Franz et al. 2008; Mattes and Redlawsk 2015; Fowler et al. 2016). What is certain, however, is that online political advertising is increasingly prevalent among campaigns and the public alike. It has been reported that digital ads have swollen, and continue to swell, in both number and spend; for instance, between the 2014 and 2018 U.S. midterm elections, spending on digital media campaigns rose by 2400%.¹ Political advertisers have thereby demonstrated a clear interest in taking their

¹ Digital spending hits record in us midterms despite restrictions. <https://www.straitstimes.com/world/united-states/digital-spending-hits-record-in-us-midterms-despite-restrictions>, November 2018. Accessed 24 Aug 2020

messages online. Meanwhile, the Internet's massive user base suggests that the reach of these ads is indeed vast (Dommett and Power 2019). For example, Facebook—the focus of this paper—reported 2.27 billion users in 2018 alone, and that figure has only gone up since (Clement 2020). Incidentally, a number of studies juxtaposing social media with more conventional media (e.g., television and newspaper) and advertising have consistently found differences with respect to both implementation strategies and public effects (Varol 2018; Kaid 2002; Kim et al. 2016; Morris and Ogan 1996). Thus, given online political advertising's growing salience and apparent distinction, there is a strong incentive to analyze its general form and effects as a phenomenon of its own.

More specifically, concerns around transparency in political ad *funding* are heightened in the online environment. Within the last decade, political advertising activity has shifted away from political action committees, and toward super PACs, 501c4 organizations, and 527s (Ridout et al. 2015). The latter set of interest groups stand out because they are permitted to spend unlimited sums of money on political advertising and can often evade disclosing their contributing donors (Dowling and Wichowsky 2013; Fowler et al. 2020a). In addition, the rise of multi-issue non-membership groups (Franz et al. 2015) has meant that there are many more new, amorphous groups with unclear relationships to official campaigns (Victor and Reinhardt 2018), whose names do not convey any information about where their money comes from. This has a variety of implications. For one, studies have shown that group-sponsored ads—and in particular, *unknown group*-sponsored ads—are extremely effective (Brooks and Murov 2012; Dowling and Wichowsky 2015; Ridout et al. 2015; Weber et al. 2012). If we consider the effectiveness of an ad, as many studies do, to be its persuasiveness (the extent to which it raises favorability toward the issue in question) minus its backlash (the extent to which it lowers favorability toward the issue in question), this may be because group-sponsored ads can garner persuasiveness with less risk of backlash (Brooks and Murov 2012; Dowling and Wichowsky 2015). Further, if a funding group is anonymous, or simply new and unknown, consumers cannot be influenced by their preexisting perceptions of the group when evaluating their ads, which otherwise tends to be a major predictor of success in advertising. Ultimately, anonymous funding can be seen as a technique to obscure political incentives behind ads without sacrificing effectiveness, but *enhancing* it. This is a problem on TV, but it is an even bigger problem online, where there are even more ways that ad sponsors can present information—for example, spending across multiple pages with different names. Put differently, as we move online, the effort to make political advertising more transparent must escalate.

The arrival of online ad libraries in the summer of 2018 were a welcome development in the effort toward transparency (Leathern 2018). Citing a need for more accountability among advertisers and the platform itself, Facebook launched an archive in 2018 that stores all of the active political ads published on the site, with some additional information about them, including funder identification.² This mainly implies that access to Facebook's political ad data has only very recently become available to researchers. Although inaccuracies in the archive persist (Oleinikov 2020), and the set of active

² About the ad library. <https://www.facebook.com/business/help/2405092116183307>. Accessed 24 Aug 2020.

ads alone hardly tells the whole story, the newly provided data represent an important opportunity to begin investigating the Facebook advertising world, and consolidating the struggle for greater transparency.

In light of all of the above, we present a data-driven approach to categorizing and characterizing *funding behavior* behind political ads on Facebook. With the overarching goal of increased donor transparency, we aim to make the set of funding entities more navigable by identifying meaningful ways to define and group them. Namely, we discover that describing ads according to their geographic reach provides a salient framework for thinking about the universe of advertisements, as well as the set of entities that back them.

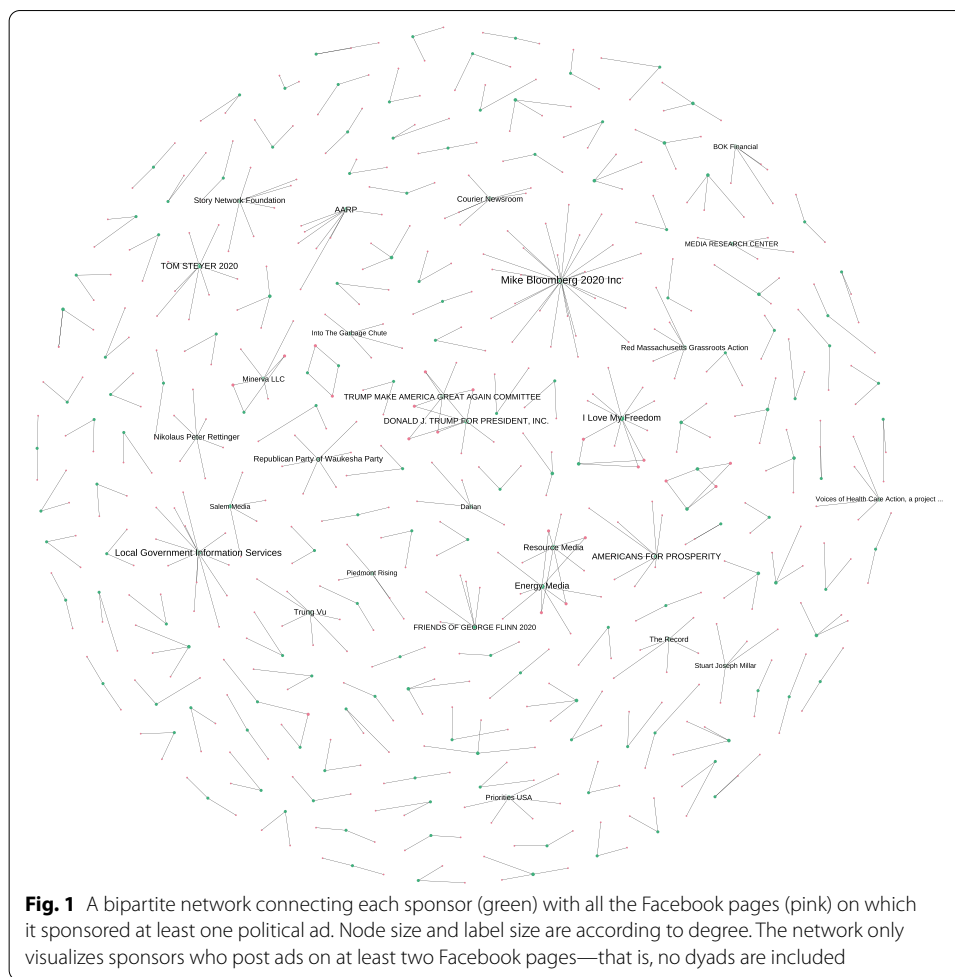
Data and methods

Our core dataset comes from a collaboration with the Wesleyan Media Project—the leading source for tracking and analyzing digital ads and spending in real time.³ It includes nearly all of the presidential and senatorial election-related political ads posted on Facebook between October 7th, 2019 and May 18th, 2020. The ads were collected through robust keyword searches, where keywords were generated based on candidates running for senate or the presidency. A unique advertisement is represented by a numerical “ad ID,” and is associated with about 20 informational fields from the AWS archive hosted by Facebook, in addition to the search terms responsible for procuring the ad.

Two fields from the Facebook archive that are of particular importance to us are the “sponsor” (or disclaimer) field—a string naming who funded the given ad (e.g., “BERNIE 2020”), and the “page id” field—a numerical code uniquely identifying the Facebook page upon which the ad was posted. In this analysis, we consider a unique “funding entity” (or just “entity”) to be the pair (“sponsor”, “page id”). That is, a funding entity is a specific actor *publishing on a specific page*; if a single actor posts on multiple pages, they become associated with multiple funding entities, which allows us to preserve a high level of nuance when characterizing funding behavior. To illustrate this, Fig. 1 shows a bipartite network connecting sponsors (green) to all of the Facebook pages (pink) upon which they sponsored political ads. The network is restricted to those sponsors who post ads on at least two Facebook pages. More specifically, out of 5001 sponsors in our data, in Fig. 1 we show only those 168 sponsors who post ads on multiple pages, since the rest of the sponsors would just appear as disconnected dyads. We see that the network mostly consists of disconnected components of unique actors supporting ads on multiple pages (e.g. the ‘Mike Bloomberg 2020 Inc’ component at the top), and very few instances of more than one actor publishing on a common page. Some examples of actors who do publish ads on common pages include ‘TRUMP MAKE AMERICA GREAT AGAIN COMMITTEE’ and ‘DONALD J. TRUMP FOR PRESIDENT, INC’, ‘Resource Media’ and ‘Energy Media’, ‘I Love My Freedom’ and ‘Making Web LLC’.

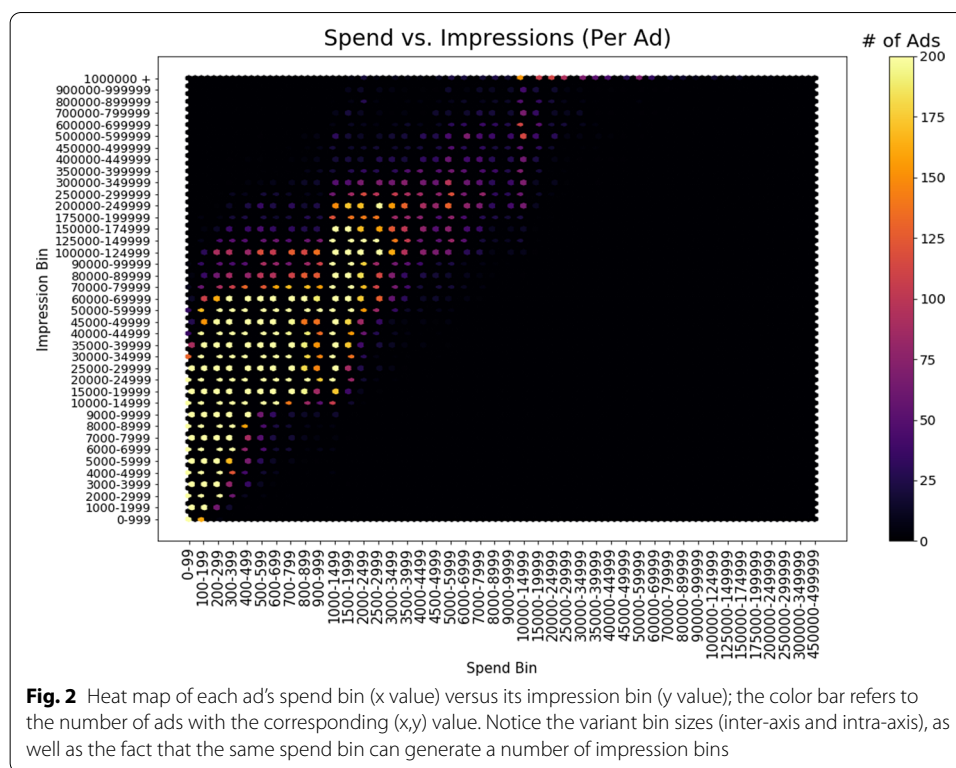
It is worth noting that the sponsor/disclaimer field is a free-form string; therefore, it is subject to typos and variation across instances (Edelson et al. 2019). Moreover, other key fields, such as the total impressions and total spend of an ad, are given as bins, rather

³ Home: Wesleyan media project. <https://mediaproject.wesleyan.edu>.



than as precise values. In Fig. 2, we show the impression bin versus the spend bin for each ad. Firstly, Fig. 2 demonstrates how the bins are not only inexact, but also incongruent—as the numbers rise, so, too, do the bin sizes. In this work, we always use the lower bounds of these bins as the safest estimate; still, summing estimated values (e.g., to calculate a funding entity’s total spend) can result in large errors, especially when an entity funds a large number of low-priced ads. Figure 2 also allows us to observe that the same ad price can generate a range of impressions. This implies that other ad attributes, such as content and demographic reach, may play a role in determining how many impressions each dollar buys (Ali et al. 2019).

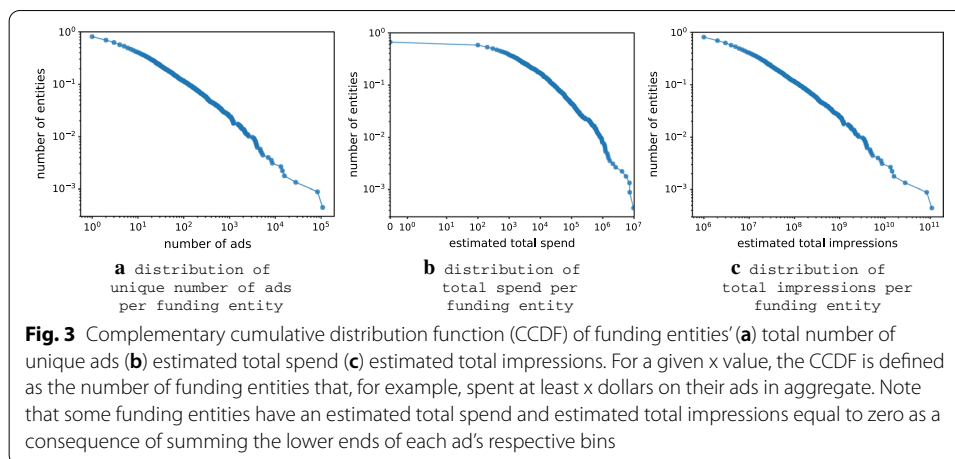
Another valuable attribute provided for each ad is the distribution of geographic regions in which users were shown the ad. More specifically, each ad is associated with a vector that maps a region to the non-zero percentage of impressions (i.e., views) that came from that region, so that the sum over all of the percentages given is 100. If a region is responsible for 0% of an ad’s impressions, it does not appear in the vector; therefore, the vector length indicates the number of regions hit by the ad. For our purposes, “region” is interchangeable with “U.S. state or Washington D.C.” The presence of non-state regions in our data is insignificant, but there are the occasional impressions



from Puerto Rico and Canada. When discussing the region vector, it is important to distinguish between intentional targeting and the set of users to whom the Facebook algorithm chooses to show the ad. Namely, the region vector does not necessarily refer to an entity's preliminary intentions, or desire to target a specific state—it simply reveals an ad's eventual reach. Ultimately, the region vector is relatively precise and sufficiently refined, making it a field of interest.

In conjunction with the core dataset of Facebook ads and their attributes, we also utilize a meta-data file with supplementary entity classification. This file is a product of a partnership with the Center for Responsive Politics—an organization that monitors and analyzes campaign donations—and the work of the Wesleyan Media Project team.⁴ The meta-data introduces information such as the election type that an entity is associated with (e.g., presidential, senatorial, etc.), and whether the entity represents an interest group or an official campaign. Beyond these fields and among others, the file encodes an entity's "disclosure type," or the extent to which it discloses its donor list: fully, partially, or not at all. Considering our underlying goal of funding transparency, this field is especially noteworthy. Unfortunately, a major feature of the meta-data file at large is that it is mostly blank; even though 70% of spending and 57% of ads were classified, only 22% of entities were identified [see Table 6.2 in Fowler et al. (2020b)]. After merging the core data with the meta-data, we are left with 631,816 unique ads funded by 2251 unique funding entities, which is the set our analysis will focus on.

⁴ Home: Center for responsive politics. <https://www.opensecrets.org>.

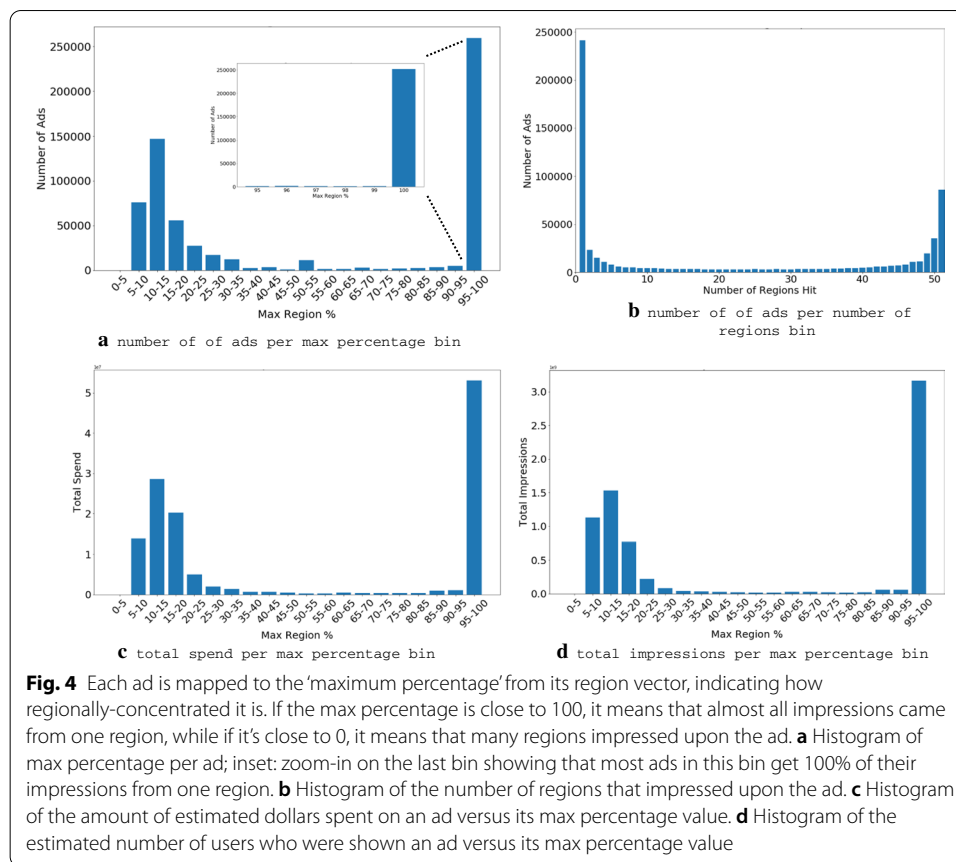


Networks emerge as a powerful tool and central method throughout our analysis because they emphasize connectivity, and can draw out subsurface relationships between entities. We construct bipartite networks in particular, where funding entities always represent one set of the nodes. The significance of such networks depends on what is chosen as the second set of nodes, but in general, our bipartite networks are useful for their ability to expose entity clustering around external items, such as ad content or strategy. We use network analyses to validate potential partitions, both structurally through network statistics, and visually through curated visualizations. In much the same vein, we employ networks to compare behavior across found partition groups.

Analysis

In our first attempt to characterize funding behavior behind political ads on Facebook, we examine the activity levels of funding entities. In Fig. 3, we show the distribution of each entity's total number of ads (left), estimated total spend (middle), and estimated total impressions, or number of users who were shown the ad (right). Recall that the total spend and impressions are obtained by summing the estimated *lower bound* for each ad (see Fig. 2 for more information). We observe that all three distributions have a heavy tail, meaning funding entities typically exhibit behavior at all scales, as opposed to groups or "types" of behavior. For instance, the complementary cumulative distribution function (CCDF) shown in the left panel indicates that entities sponsor every possible number of ads between 1 and 1,000,000, without clear boundaries as to what constitutes funding "a lot" of ads versus "a few." Therefore, grouping entities according to the number of ads they fund is untenable.

Instead, we discover a way to break up the universe of *advertisements* based on how concentrated they are in a singular geographic region. Recall that each ad is associated with a vector that reports the percentage of impressions that came from each region (i.e., state). If the maximum percentage in the vector is 100, this means that all of the impressions came from one region, or that every user who was shown the ad is located in the same state. On the other hand, if the maximum percentage is low, this means that users from many distinct regions were shown the ad. This is justified



by the fact that the sum of the percentages in the vector always equals 100; hence, if the largest percentage is small, the vector must comprise many other small values in order to reach 100.

In Fig. 4a we show the extent to which ads are regionally-concentrated. Specifically, we map each ad to the maximum percentage in its region vector, and plot the distribution of the obtained values. We notice an interesting pattern, wherein ads are either very strongly regionally-concentrated or very weakly regionally-concentrated, with little activity in the middle. The tall bar on the right (95–100%) indicates that 259,463 ads—41%—have a very high maximum region percentage, meaning they are chiefly viewed in a singular region. In contrast, the 3 tallest bars on the left (5–10%, 10–15%, 15–20%) collectively reveal that 278,776 ads—44%—have a very small maximum region percentage, meaning they are not significantly concentrated in any one region. Putting these two facts together, we get that only 15% of ads manifest something in between regional concentration and regional dispersion.

The same pattern arises when we plot the total spend on ads per maximum percentage bin (Fig. 4c), as well as the total impressions received by ads per maximum percentage bin (Fig. 4d). For example, the last bar in Fig. 4c indicates that over 50 million dollars were spent on ads with a maximum region percentage between 95 and 100. Similarly, in

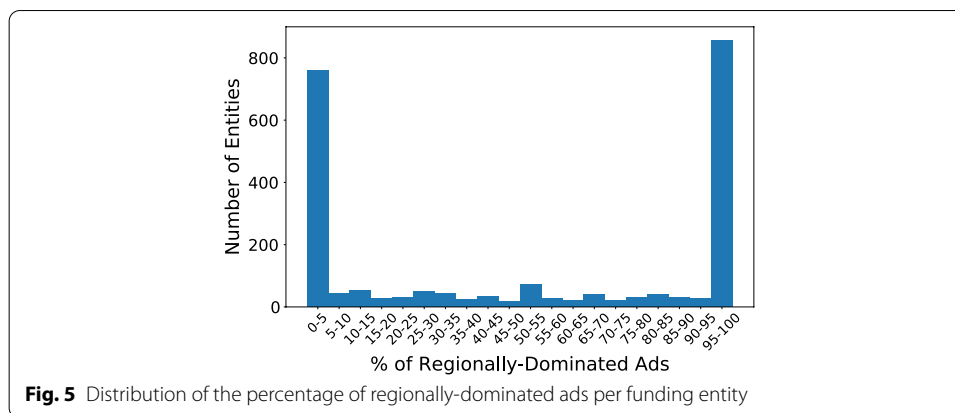


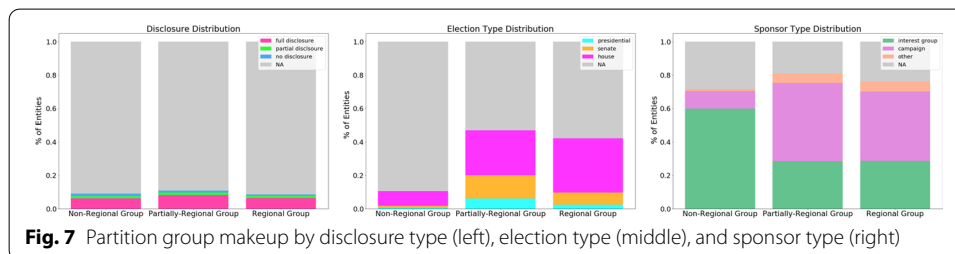
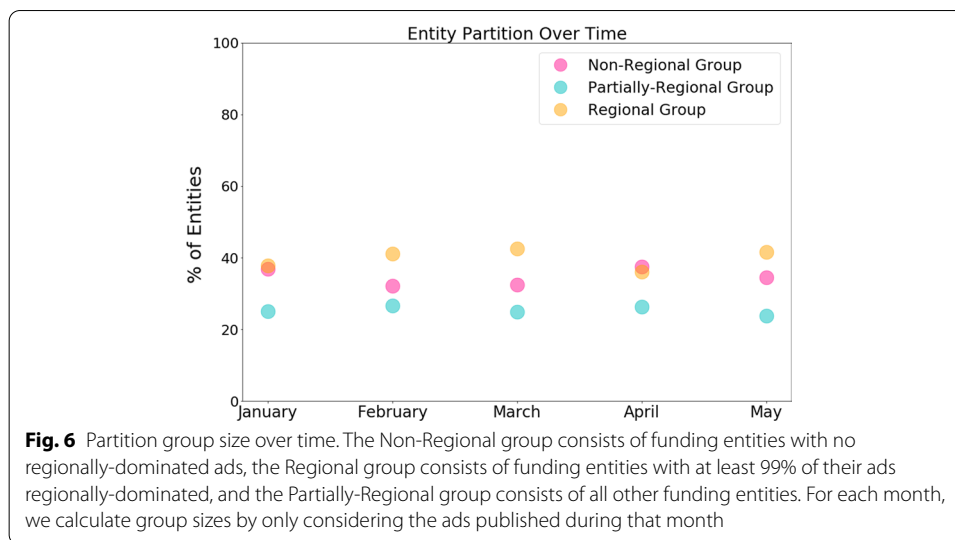
Fig. 4b we observe that the number of unique regions from which ads gain impressions is typically either large (i.e., close to 50) or small (i.e., close to 1), with very few ads in between (e.g., only 3052 ads get impressions from 20 distinct regions).⁵

Drawing on the results above, we define an ad to be **regionally-dominated** if it receives $\geq 99\%$ of its impressions from a singular region. Intuitively then, a regionally-dominated ad is an ad that is only seen in one state (or D.C.). Figure 4 reveals that 35% of the ads in our dataset are regionally-dominated, with at least \$53,058,600 invested in them, and at least 3,164,925,000 users having viewed them. We observe very similar behavior when we use the Gini coefficient of the region vector instead of the maximum percentage defined above (see Section 2 in the Additional file 1).

Next, we consider funding entities' relationships to regionally-dominated ads. In Fig. 5, we show the distribution of percentages of regionally-dominated ads per funding entity. We observe that 676 entities (30%) have 0% (none) of their ads regionally-dominated, 762 entities (36%) have at least 99% of their ads regionally-dominated, and the remaining 813 entities (34%) fall somewhere in between. Thus, a partition emerges, splitting the entities into three respective groups: the "Non-Regional group," the "Partially-Regional group," and the "Regional group."

The marked evenness of this partition is somewhat surprising, especially when noting that gender-dominated ads and age-dominated ads do not yield similar results. To clarify, every ad in our dataset is also associated with an age range (13–24, 25–44, 45–64, 65+) and gender ("male," "female," "unknown") vector, which functions similarly to the region vector. When trying to partition funding entities based on their percentage of age-dominated ads (that is, at least 99% of the impressions came from one age group) and their percentage of gender-dominated ads (that is, at least 99% of impressions came from one gender group), we find that 87% of entities have 0 of the former, and 78% have 0 of the latter. See section 1 of the Additional file 1 for further elaboration on the distinction between regional, age-based, and gendered impressions. What's more, the region-based partition is remarkably stable over time, as demonstrated in Fig. 6.

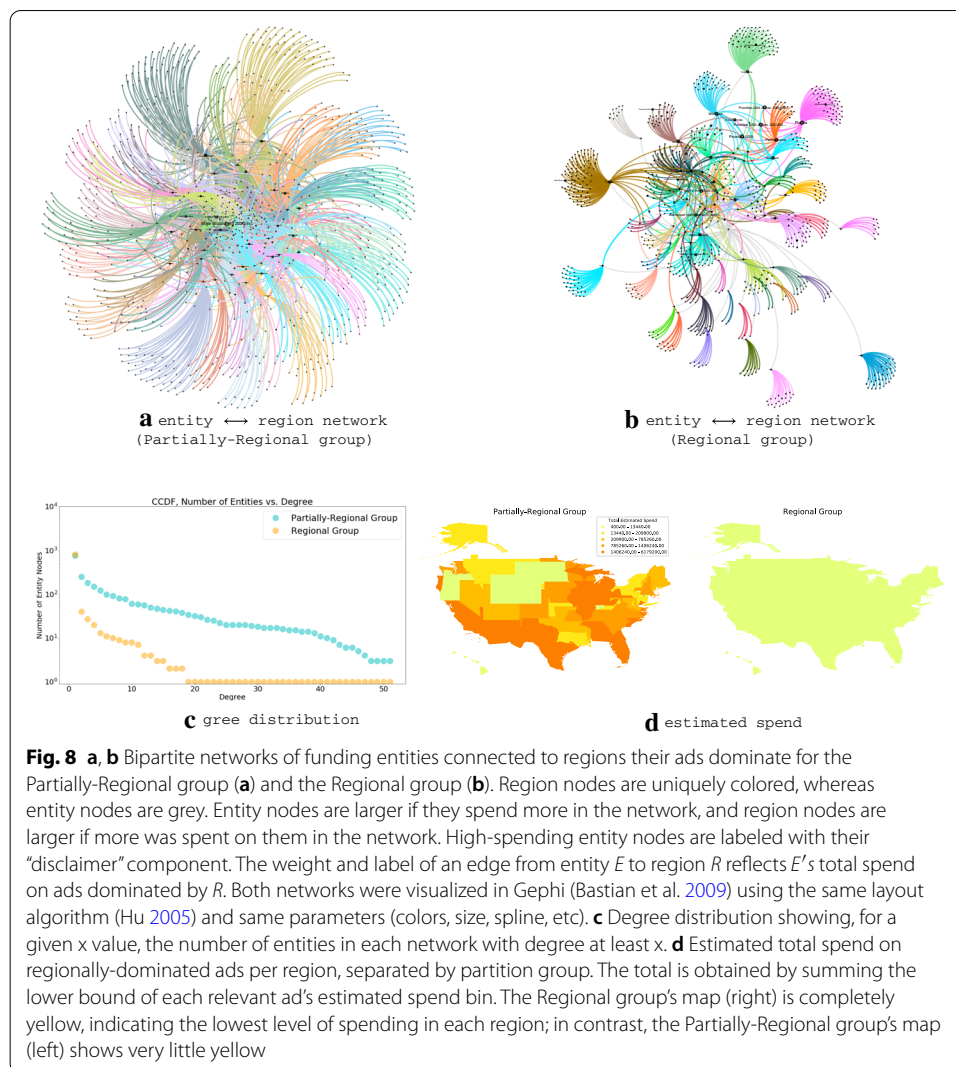
⁵ Keep in mind that a percentage in the region vector can be extremely small, so there may be regions counted in this measure that do not necessarily supply serious engagement.



Having found a potentially successful and stable partition, we examine the makeup of each partition group based on entity meta-data. We find that, although the vast majority of entities do not have a disclosure label, each group comprises entities of all disclosure types in near-even percentages (Fig. 7, left). Likewise, the partition does not prove to form around election lines, for entities of all election types (President, U.S. Senate, U.S. House) appear in every group (Fig. 7, middle). Nevertheless, a strong distinction seems to lie within the distribution of entity “sponsor type,” a variable developed by the Wesleyan Media Project that mainly distinguishes between political “interest groups” and “campaigns” (i.e., candidates).⁶ The right panel of Fig. 7 conveys that the Non-Regional group is almost entirely made up of interest groups, whereas the Partially-Regional and Regional groups are clearly more campaign-heavy.

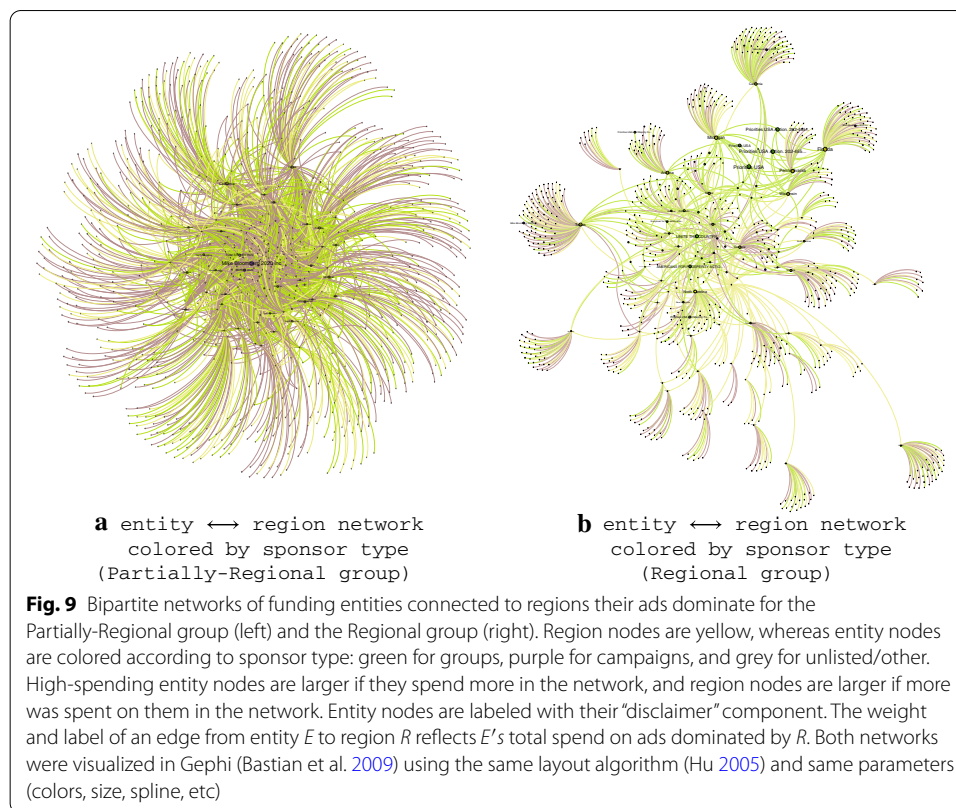
Given the relative non-specificity of the Partially-Regional group, as well as its similarity to the Regional group in the distributions discussed above, we look at how entities’ relationships to their regionally-dominated ads differ between this group and the Regional group—beyond just in proportional share. Figure 8 visualizes bipartite networks of entities connected to regions their ads dominate for the Partially-Regional group (left), and the Regional group (right). A funding entity node is connected to a region node if the entity sponsors at least one regionally-dominated ad that gets the majority of its

⁶ The sponsor type labels beyond “interest group” and “campaign” only exist in very small numbers; they include: leadership PAC, party, coordinated, super PAC, candidate for another office, government official, and government agency.



impressions from that region. Spend in the networks is expressed through node size and edge weight. For an entity node, spend refers to how much that entity spent on regionally-dominated ads; for a region node, spend refers to how much was spent on ads dominated by that region. Additionally, node color—unique for region nodes and grey for entity nodes—emphasizes the number of entities advertising in each region.

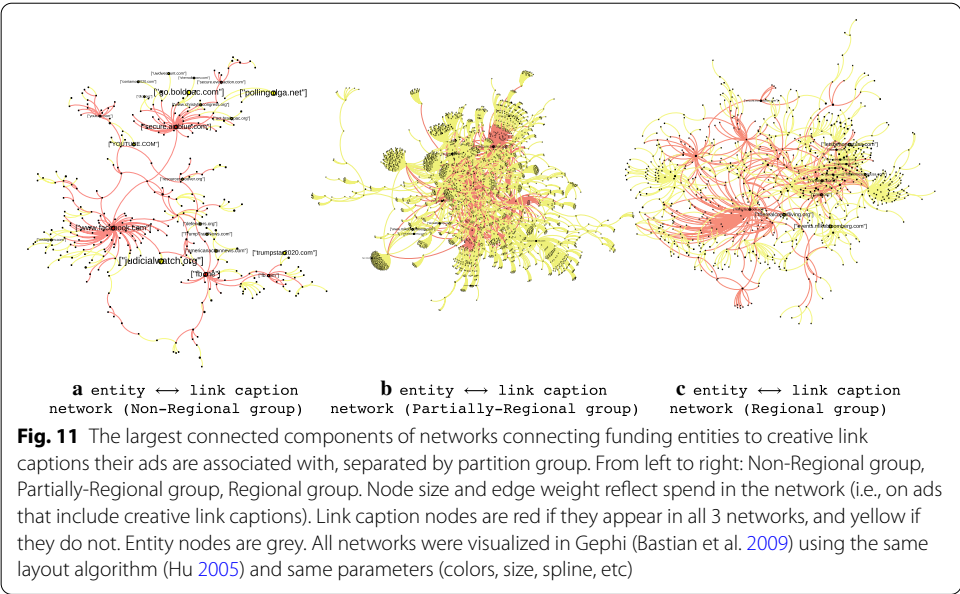
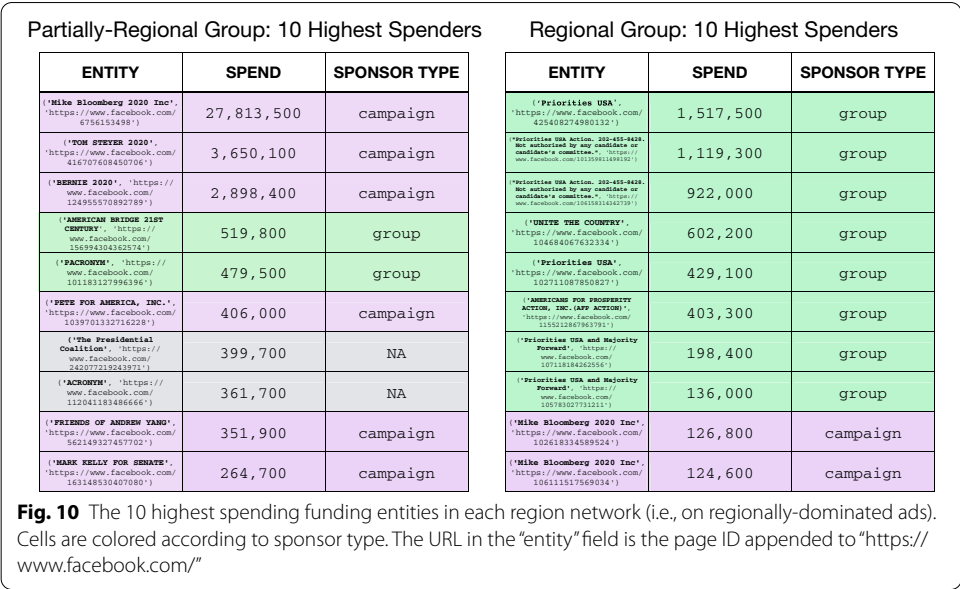
Both networks were visualized in Gephi (Bastian et al. 2009) using the same layout algorithm (Hu 2005) and same parameters (colors, size, spline, etc.), yet they appear strikingly dissimilar. Moreover, the two networks present distinct structural properties. In particular, the degree distribution shown in Fig. 8c reveals that $\sim 90\%$ of entity nodes in the Regional group's network have degree equal to 1, compared with only $\sim 70\%$ in the Partially-Regional group's network. To compound this, we also find a major disparity in modularity scores in the projections of these networks onto their entity nodes (that is, two entity nodes are connected in the projection if they are connected to the same region node in the bipartite network). Network modularity measures the extent to which a network is organized into tightly connected modules, also referred to



as communities (Newman 2006). Using the Clauset–Newman–Moore greedy modularity maximization algorithm (Clauset et al. 2004) to extract communities, we obtain 6 communities with a modularity score of .17 in the Partially-Regional group’s network (Fig. 8a), and 13 communities with a modularity score of .63 in the Regional group’s network (Fig. 8b). In other words, the Regional group network has a significantly larger modularity score, or more strongly defined community structure, which is exactly what one would expect from just looking at the visualizations. Overall then, entities in the Regional group are more likely than those in the Partially-Regional group to have all of their regionally-dominated ads concentrated in one particular region, as opposed to many. At the same time, despite individually investing in a wider set of regions, the entities in the Partially-Regional group outspend the entities in the Regional group in almost every single region, as can be seen in the maps in Fig. 8d.

Finally, when we color entities in the network according to their “sponsor type”—green for interest group, purple for campaign, grey for unlisted/other—we see that the high spending entities are campaigns in the Partially-Regional group’s network (Fig. 9, left), and interest groups in the Regional group’s network (Fig. 9, right). This contrast is reinforced by the tables in Fig. 10, which list the 10 highest spending funding entities in each of the networks, along with their sponsor type.

After differentiating between the Partially-Regional group and the Regional group, we move on to compare the qualitative behavior of all three partition groups. For each group, we construct a bipartite network, where entities are connected to “creative link captions” that their ads are associated with. The “creative link caption” of an ad is the



text accompanying its embedded URL, if one is present; with some exception, “creative link captions” are usually Internet domain names. We choose these as our second set of nodes because they get at entities’ content, whilst remaining relatively general and shared. The largest connected components of these networks are displayed in Fig. 11: node size and edge weight are correlated with spend, link caption nodes are red if they appear in all three networks⁷—yellow if they do not, and entity nodes are grey.

⁷ The shared link captions are: abcnews.go.com, actionnetwork.org, apnews.com, azmirror.com, berniesanders.com, bit.ly, businessinsider.com, cbsnews.com, cnn.com, denverpost.com, facebook.com, fb.com, fb.me, foxnews.com, huffpost.com, latimes.com, mobilize.us, nationalreview.com, nbcnews.com, newsweek.com, npr.org, nytimes.com, politico.com, realclearpolitics.com, secure.actblue.com, secure.everyaction.com, thehill.com, usatoday.com, washingtonpost.com, www.facebook.com, youtube.com

Non-Regional Group			Partially-Regional Group			Regional Group		
CREATIVE LINK CAPTION	DEGREE	SPEND	CREATIVE LINK CAPTION	DEGREE	SPEND	CREATIVE LINK CAPTION	DEGREE	SPEND
"www.facebook.com"	60	82,300	"secure.actblue.com"	153	19,196,800	"www.facebook.com"	75	100,900
"secure.actblue.com"	35	65,800	"www.facebook.com"	109	165,900	"secure.winred.com"	42	36,700
"fb.me"	16	91,900	"secure.winred.com"	70	2,719,800	"fb.me"	33	43,100
"youtube.com"	4	7,100	"fb.me"	55	170,700	"youtube.com"	27	1,700
"fb.com"	9	19,000	"youtube.com"	32	5,100	"secure.actblue.com"	17	3,100

Fig. 12 The 5 creative link caption nodes with the highest degree in each network, along with their degree and spend

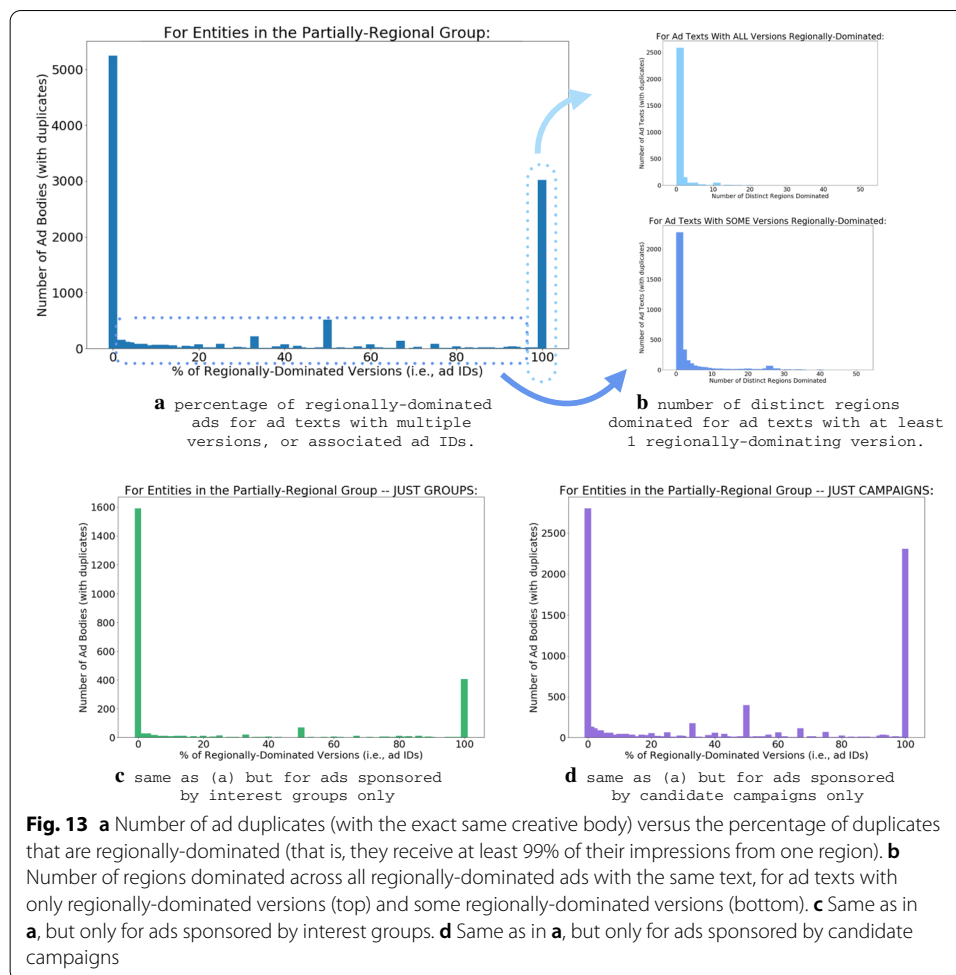
Again, all three networks look remarkably different, despite identical visualization processes. In the Non-Regional group's network, entity nodes tend to cluster around link caption nodes, while the opposite is true in the Partially-Regional group's network. Meanwhile, entities in the Regional group's network exhibit a balanced mix of both clustering habits. This characterization is partially substantiated by the ratio of entity nodes to link caption nodes in each network—55:45 in the Non-Regional group, 22:78 in the Partially-Regional group, and 39:61 in the Regional group. Moreover, a similar relationship between partition groups emerges in the sponsor type distributions of each network's 10 highest spending entities. That is, the Non-Regional group's highest spenders are predominantly interest groups (80%), the Partially-Regional group's highest spenders are predominantly campaigns (90%), and the Regional group's highest spenders are evenly split between the two types. In addition, the net worth of each network⁸, or the total spend over all of the link caption nodes, varies dramatically across groups. We have \$1,061,100 in the Non-Regional group, \$99,432,000 in the Partially-Regional group, and \$4,160,500 in the Regional group—yet another measure that places the Regional group in the middle.

To begin inspecting content-based distinctions, we identify the prominent link captions in each network. The table in Fig. 12 lists the 5 creative link captions with the highest degree from each network, as well as their degree and spend. The variant rankings of "secure.actblue.com" and "secure.winred.com"—the donation sites for the Democratic and Republican parties, respectively—suggest disparate partisan makeup among the three partition groups, and hence disparate content output.⁹

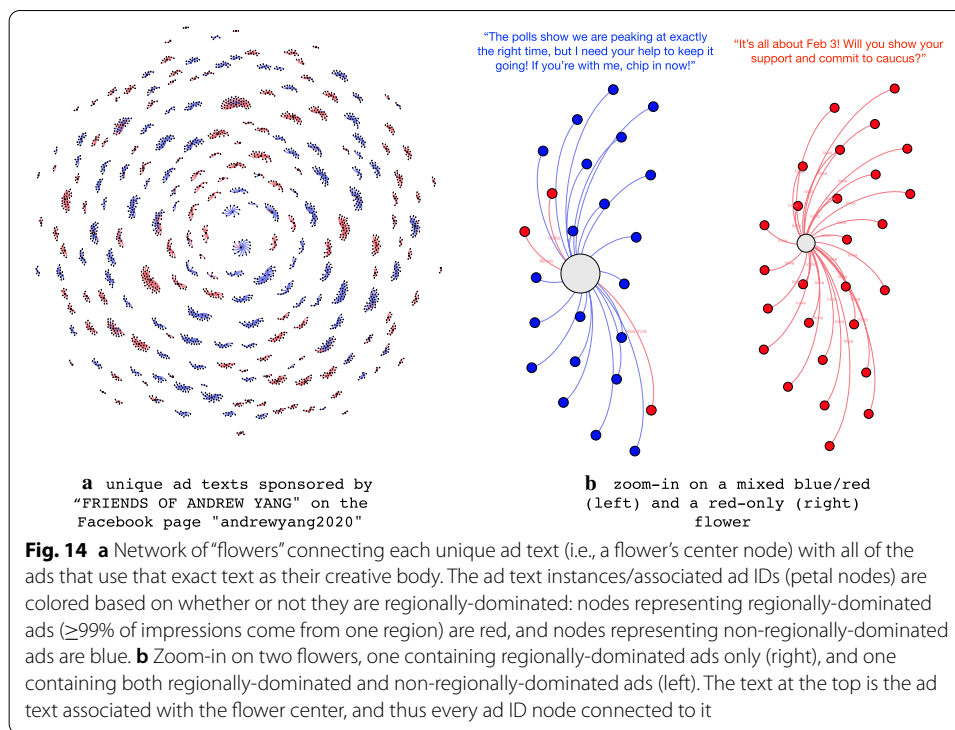
Our concluding set of results continue to investigate the way content interacts with regional strategy, but with a slightly different angle. Our main goal in this segment is

⁸ The idea of the "net worth" of a "network" should probably go viral.

⁹ "secure.winred.com" is present among entities in the Non-Regional group, but it does not appear in the largest connected component. Instead, it makes up its own isolated component, with 8 entity neighbors, and a total spend of \$43,000.



to consider the extent to which an ad's content is correlated with its regional behavior, or whether or not it is regionally-dominated. Included in our data is a string representing each ad's "creative body," or its central *text*. Often, funding entities reproduce multiple instances of the same ad text—that is, they publish multiple unique ad IDs with the exact same creative body string, but potentially other differences (e.g., with respect to spend, audience, attached image, etc.). To explore how such text equivalency relates to regionally-dominated ads versus non-regionally-dominated ads, we hone in on entities in the Partially-Regional group, for they sponsor ads of both sorts. More specifically, we look at their distribution of regionally-dominated ads and non-regionally-dominated ads per recycled ad text. As Fig. 13a shows, it is most common for all versions of the same ad text to be either regionally-dominated (100% on the x-axis) or not (0% on the x-axis). Generally, however, a repeat ad text in the Partially-Regional group has a 45% chance of having no regionally-dominated instances, a 26% chance of having entirely regionally-dominated instances, and a 29% chance of having mixed instances. These numbers change quite a bit when we consider interest groups and campaigns independently (see Fig. 13c, d). We find that groups have a 65% chance of having 0 regionally-dominated versions for a given text, whereas campaigns only have a 37% chance. Beyond that, 16% of interest groups' recycled texts—compared



with 34% of campaigns' recycled texts—are associated with both regionally-dominated and non-regionally-dominated versions. Further, upon examining the number of distinct regions dominated by a specific text across all funding entities, we see that one text is typically reserved for one region alone (Fig. 13b). At the same time, this is 21% *more likely* if the ad text has exclusively regionally-dominated versions, rather than a mix (Fig. 13b, top vs. bottom).

A case study of the entity "(FRIENDS OF ANDREW YANG, 'https://www.facebook.com/562149327457702') and their recycled ad texts confirms the above phenomena. Note that the entity in question represents a presidential campaign, and has 47% of their ads regionally-dominated. Figure 14a visualizes a network where unique ad IDs are connected to the unique ad text they contain. Blue nodes represent non-regionally-dominated ads, and red nodes represent regionally-dominated ads; ad text nodes are grey, and sized according to the number of distinct regions they dominate. For ease of explanation, we will refer to an ad text and its neighbors as a "flower." In the given network, there are 70 blue flowers (that is, 70 unique ad texts with only non-regionally-dominated versions), 48 mixed flowers (that is, 48 unique ad texts with both regionally-dominated and non-regionally-dominated versions), and 98 red flowers (that is, 98 unique ad texts with only regionally-dominated versions). The red flowers shown only ever dominate in one region, while a handful of mixed flowers dominate in multiple regions. To be clear, this is non-trivial—there is nothing preventing a red flower from being dominated by multiple regions. Figure 14b exemplifies this trend by offering a close-up of two salient flowers—one mixed and one red-only—along with their associated text. The red flower is concentrated in Iowa, and is about participating in the Iowa caucus; the mixed flower dominates three distinct regions, and is meant for fundraising purposes.

The fundraising-oriented text of the mixed flower displayed in Fig. 14b leads us to question how fundraising figures into regional behavior more broadly. At a first pass, we designate ads to be about fundraising if they have “secure.actblue.com” or “secure.winred.com” as their creative link caption. Surely, we miss some fundraising ads through this measure, but it is also unlikely to turn up false positives. We discover that out of every such fundraising ad that the Partially-Regional group publishes, 41% would be part of a blue flower, 56% would be part of a mixed flower, 2% would be part of a red flower, and 1% would not be part of a flower at all because their ad text is one of a kind. In other words, a fundraising ad (by our definition) is most likely to share ad text with both regionally-dominated and non-regionally dominated versions, and least likely to share ad text with exclusively regionally-dominated versions.

Discussion

Given the tendency of funding entities in our data to demonstrate a range of behaviors, rather than discrete groups of behaviors, our findings related to the regional distribution of ads and the entity partition they prompt are noteworthy. Yet, they are also noteworthy in and of themselves. The emergence of “regionally-dominated” ads, for example, is simultaneously salient in the data, and understandable on a practical level. That is, some ads are only seen in one region, intentionally or otherwise. While we can’t directly comment on the intentions of entities, as noted earlier, it is reasonable to assume that most regionally-dominated ads are indeed meant for an audience in the region they dominate. This is especially reasonable knowing that, upon purchasing an ad, entities *can* request for it to circulate in specific regions.¹⁰ In any case, the partition over all entities based on their relationships to regionally-dominated ads is extremely useful. For one, it provides an intuitive framework for thinking about the universe of entities through their interest in particular regions, and the scope of their audiences. This works to accentuate strategic and positional differences between sets of entities. Additionally, the partition allows us to split up the 2251 entities into three clean and meaningful groups, thus making the data “smaller” and more manageable to navigate and analyze.

It certainly would have been compelling to have been able to draw this partition along disclosure type lines, and identify a way to predict non-disclosure funding entities. Instead, our results suggest equal amounts of each disclosure type in every partition group, making every group of interest as far as transparency is concerned. At the same time, the strongest result from the disclosure type distributions is perhaps the dearth of disclosure labels in the meta-data; ultimately, little can be drawn from our disclosure information, since it only exists in trace amounts. Election type likewise does not strongly correlate with partition groups, although one might expect that some races are regionally-concentrated (or at least more regionally-concentrated than others). The sponsor type distributions hence provide our only partial distinction, by showing a significant concentration of interest group entities in the Non-Regional group, and not in the others. On the upside, interest groups comprise precisely those actors capable of evasively hiding their donor lists, so it is helpful to know where to find them.

¹⁰ Facebook for business, ad targeting. <https://www.facebook.com/business/ads/ad-targeting>.

While the difference between the Non-Regional group and the Regional group is quite obvious (no regionally-dominated ads vs. all regionally-dominated ads), the difference between the Regional group and the Partially-Regional group is less clear. However, when we construct networks for each of these groups, connecting funding entities to regions their ads regionally dominate, we encounter divergent structures. Through degree distributions and modularity assessments, we find that entities in the Partially-Regional group are significantly more likely than entities in the Regional group to concentrate their ads in more than one region. This implies that entities in the Regional group most frequently produce regionally-dominated ads because they have a genuine interest in some region, while entities in the Partially-Regional group use regionally-dominated ads as a strategic method. Consequently, we can confidently distinguish between the Regional group and the Partially-Regional group. In addition, we can conclude that regionally-dominated ads may serve a number of purposes.

Having validated our discovered entity partition extensively, we proceed to explore the qualitative differences between its three groups. In assembling networks separated by partition group, where funding entities are connected to their creative link captions, we observe multiple patterns and points of contrast, visually and beyond. Importantly, by taking the net worth of each network, we confirm that our biggest players/highest spenders lie within the Partially-Regional group. We also see that, by many measures, the Regional group falls in between the Non-Regional group and the Partially-Regional group. On the contrary, one might *expect* to see the Partially-Regional group play that role, since entities in this group sample strategies from the other two. The reasons driving this phenomenon deserve further scrutiny. Encouragingly, our investigation of each network's most popular creative link captions reveals varying partisan leanings between partition groups. This indicates content-based distinctions between the groups, and marks the beginning of the search for more (Conover et al. 2011). Moreover, it reinforces the integrity of the partition, along with our choice to study it.

Finally, we hone in on entities in the Partially-Regional group in order to explore how the presence of regionally-dominated ads interacts with messaging. Although there is still uncertainty around the precise relationship between an ad's regional concentration and its content, our results offer important insight. First, in line with previous findings, we see that regionally-dominated ads are used in various ways: broadly speaking, some are organized around a specific text, and others are not. We find that this is dependent, to an extent, on whether the funding entity behind a given text represents a group or a campaign; for example, campaigns are 15% more likely than groups to publish an ad text with a mixture of regionally-dominated and non-regionally-dominated versions. Still, among themselves, both groups and campaigns publish more texts with mixed regional behavior than texts with regionally-dominating behavior only. The scenarios wherein specific texts exclusively exhibit regionally-dominating behavior suggest that certain content is simply meant for a regional audience. To go even further, the fact that only one distinct region is typically dominated in such instances implies that certain content is not just meant for a regional audience, but for a *particular* regional audience. On the flipside, when regional behavior is not organized around text—that is, when some text has both regionally-dominated and non-regionally-dominated versions—we are poised to conclude that regional concentration is being used for strategic ends. This is

substantiated by the increased likelihood for such texts to reach multiple regions, as well as the discovery that ads linking to fundraising sites (i.e., ads that inherently center strategy) in the Partially-Regional group are most commonly associated with such texts, and least commonly associated with those with regionally-dominating behavior only.

It is crucial to point out that the content analysis here bases content similarity on string equality, meaning one divergent punctuation mark is enough to register two ad texts as dissimilar. Therefore, it is possible that two very similar yet non-equal texts are fed to two distinct regions, which adds nuance to the finding that some regional-concentration is organized around specific text. Put another way, we may say that unique strings of text are reserved for unique regions, but we may not go as far as to say that unique messages or ideas are reserved for unique regions. In fact, empirical observation of select entities' recycled ad texts supports the opposite claim, as does the known existence of alpha/beta testing in advertising (Ridout 2014). Thus, while our results highlight useful trends, as outlined above, subtle differences between textual messaging remains an underdeveloped area of interest.

Conclusions and future work

While much remains to be learned about political advertisements on Facebook, and in particular, how they are funded, this study makes important headway. Most significantly, we discover a hard-to-come-by partition over all of the funding entities, dividing them into three near-equal groups. Because the partition distinguishes between entities based on their relationships to regionally-dominated ads, or the extent to which their advertisements are concentrated in individual regions, it brings out strategic differences between sets of entities. We additionally lay some groundwork for parsing how these disparate strategies operate in terms of content. Beyond that—and perhaps more crucial to the continuation of this work—the partition breaks up the universe of entities, and offers smaller pieces of the data as units of analysis. In other words, through our partition, the task of investigating political ad funding on Facebook becomes much more tractable.

Alas, political advertising and campaigning are ever-evolving practices, and their Facebook form is certainly no exception. Therefore, we cannot say definitively that this partition and its related findings will stand the test of time. Comfort may be found in the fact that the partition group sizes are quite stable month-to-month, but it is still unsure what would be the case given a different set of candidates and a different election cycle. Nevertheless, it is also unlikely that our results will become *completely* obsolete—for one, the presence of regionally-dominated ads is strong, and as we have seen, multi-purpose. Overall, then, this work successfully takes a step toward demystifying the mechanics of online political advertising.

Moving forward, we would like to utilize the results presented here in order to perform funding entity classification at the individual level. That is, we hope to provide pathways for filling in the empty fields in the meta-data file—especially when it comes to entity disclosure type. We imagine associating each entity with an attribute vector drawn from the meta-data file, with support for unlisted entries. Once that is complete, we can make various networks, separated by partition group, and cross-compare attribute similarity with detected communities and/or clustering in the networks. Such a procedure

would ultimately provide label suggestions for unlabelled entities, and thus hasten the process of individual entity classification. Developing a partially automated, expedited tool for entity classification would be a major, long-lasting step toward ensuring funding transparency in online political advertising.

In addition, our study lays groundwork for looking into spatial questions surrounding digital advertising, and in particular, what makes it successful. Namely, the regional ad clustering we have shown lends itself quite nicely to correlation with polling data, and measurement of how the two move together, if at all. Through such an approach, we may also be able to gain insights into how campaigns react to polling information, or, should there be a puzzling misalignment between polls and ad placement, the extent to which they respond to non-public data.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1007/s41109-020-00350-7>.

Additional file 1. Supplementary Information.

Acknowledgements

We are grateful to the Center for Responsive Politics for partnering with us on the entity classifications. We thank Michael Franz, Travis Ridout, and all of the members of Delta Lab for their feedback along the way. Finally, we thank the anonymous reviewers for their comments, which helped us improve the paper.

Authors' contributions

AG, PO, LB, EFF and SS designed the study. PO acquired the data. AG and SS analyzed the data. AG, PO, LB, EFF and SS interpreted the results and wrote the paper. All authors revised the work, and read and approved the final manuscript.

Funding

The authors recognize support from the James L. Knight Foundation (# GR-2019-60044), Democracy Fund (R-201912-05114), and Wesleyan University.

Availability of data and materials

The datasets used in this article, as well as the code to analyze it and create figures will be made available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Network Science Institute, Northeastern University, Boston, MA, USA. ² Quantitative Analysis Center, Wesleyan University, Middletown, CT, USA. ³ Wesleyan Media Project, Wesleyan University, Middletown, CT, USA. ⁴ Department of Government, Wesleyan University, Middletown, CT, USA. ⁵ Department of Mathematics and Computer Science, Wesleyan University, Middletown, CT, USA.

Received: 23 September 2020 Accepted: 26 December 2020

Published online: 22 February 2021

References

- Ali M, Sapiezynski P, Korolova A, Mislove A, Rieke A (2019) Ad delivery algorithms: the hidden arbiters of political messaging. arXiv preprint [arXiv:1912.04255](https://arxiv.org/abs/1912.04255)
- Ansolahehere S, Iyengar S (1995) Going negative: how political advertisements shrink and polarize the electorate. Free Press, New York
- Bastian M, Heymann S, Jacomy M et al (2009) Gephi: an open source software for exploring and manipulating networks. ICWSM 8(2009):361–362
- Brooks DJ, Murov M (2012) Assessing accountability in a post-citizens united era the effects of attack ad sponsorship by unknown independent groups. Am Polit Res 40(3):383–418
- Clauset A, Newman ME, Moore C (2004) Finding community structure in very large networks. Phys Rev E 70(6):066111
- Clement J (2020) Number of monthly active facebook users worldwide as of 2nd quarter 2020. <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide>, August 2020. Accessed 24 Aug 2020
- Conover MD, Ratkiewicz J, Francisco MR, Gonçalves B, Menczer F, Flammini A (2011) Political polarization on twitter. ICWSM 133(26):89–96
- Dommett K, Power S (2019) The political economy of facebook advertising: election spending, regulation and targeting online. Polit Q 90(2):257–265

- Dowling CM, Wichowsky A (2013) Does it matter who's behind the curtain? Anonymity in political advertising and the effects of campaign finance disclosure. *Am Polit Res* 41(6):965–996
- Dowling CM, Wichowsky A (2015) Attacks without consequence? Candidates, parties, groups, and the changing face of negative advertising. *Am J Polit Sci* 59(1):19–36
- Edelson L, Sakhuja S, Dey R, McCoy D (2019) An analysis of united states online political advertising transparency. CoRR, [arXiv:abs/1902.04385](https://arxiv.org/abs/1902.04385)
- Fowler EF, Franz M, Ridout T (2016) *Political advertising in the United States*. Routledge, New York
- Fowler EF, Franz MM, Ridout TN (2020a) Online political advertising in the United States. In: Persily N, Tucker JA (eds) *Social media and democracy: the state of the field, prospects for reform, SSRC anxieties of democracy*. Cambridge University Press, Cambridge
- Fowler EF, Franz MM, Ridout TN (2020b) Online political advertising in the United States. The state of the field, prospects for reform, social media and democracy, p 111
- Franz MM, Fowler EF, Ridout TN (2015) Loose cannons or loyal foot soldiers? Toward a more complex theory of interest group advertising strategies. *Am J Polit Sci* 60:738–751
- Franz MM, Freedman P, Goldstein K, Ridout TN (2008) *Campaign advertising and American democracy*. Temple University Press, Philadelphia
- Geer JG (2006) *In defense of negativity: attack ads in presidential campaigns*. University of Chicago Press, Chicago
- Hu Y (2005) Efficient, high-quality force-directed graph drawing. *Math J* 10(1):37–71
- Kaid LL (2002) Political advertising and information seeking: comparing exposure via traditional and internet channels. *J Advertis* 31(1):27–35
- Kim Y, Gonzenbach W, Vargo C, Kim Y (2016) First and second levels of intermedia agenda setting: Political advertising, newspapers, and twitter during the 2012 U.S. presidential election. *Int J Commun* 10:4550–4569
- Leathern R (2018) Shining a light on ads with political content. <https://about.fb.com/news/2018/05/ads-with-political-content/>, May 2018. Accessed 24 Aug 2020
- Mattes K, Redlawsk DP (2015) *The positive case for negative campaigning*. University of Chicago Press, Chicago
- Morris M, Ogan C (1996) The internet as mass medium. *J Comput Mediat Commun*. <https://doi.org/10.1111/j.1083-6101.1996.tb00174.x>
- Newman ME (2006) Modularity and community structure in networks. *Proc Natl Acad Sci* 103(23):8577–8582
- Oleinikov P (2020) Long-term monitoring helps identify gaps in facebook reporting. <https://deltalab.research.wesleyan.edu/2020/08/03/facebook-reports/>, August 2020. Accessed 24 Aug 2020
- Ridout TN (2014) The market research, testing and targeting behind American political advertising. In: Lees-Marshment J, Conley B, Cosgrove K (eds) *Political marketing in the U.S.* Routledge, New York
- Ridout TN, Franz MM, Fowler EF (2015) Sponsorship, disclosure, and donors: limiting the impact of outside group ads. *Polit Res Q* 68(1):154–166
- Varol O (2018) Deception strategies and threats for online discussions. *First Monday*. <https://doi.org/10.5210/fm.v22i5.7883>
- Victor JN, Reinhardt GY (2018) Competing for the platform: how organized interests affect party positioning in the United States. *Party Polit* 24(3):265–277
- Weber C, Dunaway J, Johnson T (2012) It's all in the name: source cue ambiguity and the persuasive appeal of campaign ads. *Polit Behav* 34(3):561–584

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
