Geographic impressions in Facebook political ads: Supplementary Information

Adina Gitomer[1,*], Pavel V. Oleinikov[2], Laura M. Baum[3], Erika Franklin Fowler[4], and Saray Shai[5]

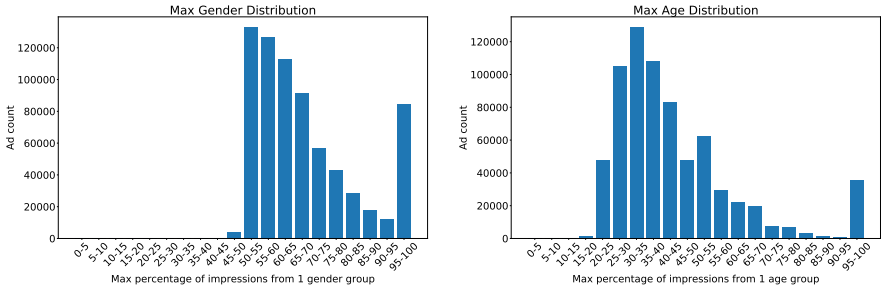
[1]Network Science Institute, Northeastern University, Boston, MA, USA
[2]Quantitative Analysis Center, Wesleyan University, Middletown, CT, USA
[3]Wesleyan Media Project, Wesleyan University, Middletown, CT, USA
[4]Department of Government, Wesleyan University, Middletown, CT, USA
[5]Department of Mathematics and Computer Science, Wesleyan University, Middletown, CT, USA
[*]Corresponding author: adinagitomer15@gmail.com

1 Demographic vs. regional impressions

In our analysis, we mention how the regional distribution of ads stands out from the distribution of ads according to the other demographic data available – namely, age and gender. We provide further justification for this here.

To remind, in addition to the region vector, we have information regarding the percentage of impressions that come from each age range (13-24, 25-44, 45-64, 65+) and gender ("male", "female", and "unknown"). Figure S1 shows how the ads are distributed according to the maximum percentage from each of these demographic vectors. As was the case for region, the higher the maximum percentage, the more concentrated is the ad on a specific gender or age. One can see that ads do not display a clear pattern in terms of concentration by age or gender, but rather exhibit a wide range of maximum age/gender percentages. Recall, for comparison, Fig. 4(a) in the main text, which shows the same histogram, but for regional impressions. This figure plainly demonstrates that ads are typically either equally-impressed upon by many regions or concentrated within one.

Beyond the fact that gender and age clustering is not nearly as well-defined as regional clustering, there is also evidence that funding entities treat their

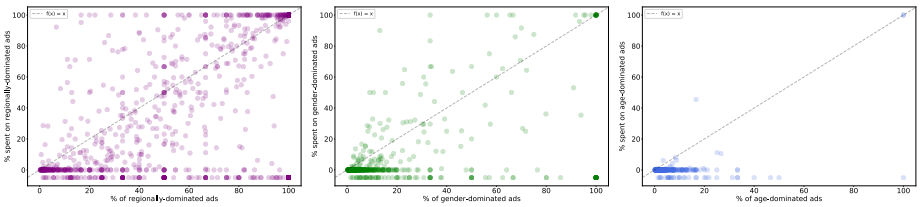


(a) number of ads per max percentage bin, gender

(b) number of ads per max percentage bin, age

Figure S1: Each ad is mapped to the "maximum percentage" from its gender vector (a), indicating how gendered its audience is, and age vector (b), indicating how age-specific its audience is.

regionally-dominated ads differently. When we plot each funding entity's percentage of regionally-dominated ads versus the percentage of their overall spending that went into their regionally-dominated ads (Fig. S2(a)), it is revealed that 28% of entities spend a disproportionate amount on their regionally-dominated ads (i.e., 28% of the data points fall above the $x = y$ line). This can be compared with 15% for gender-dominated ads (Fig. S2(b)) and .6% for age-dominated ads (Fig. S2(c)). To be clear, we define gender-dominated ads and age-dominated



(a) the % of regionally-dominated ads versus the % spent on regionally-dominated ads per funding entity

(b) the % of gender-dominated ads versus the % spent on gender-dominated ads per funding entity

(c) the % of age-dominated ads versus the % spent on age-dominated ads per funding entity

Figure S2: Each funding entity's percentage of regionally-dominated ads versus the percent spent on regionally-dominated ads (a), gender-dominated ads versus the percent spent on gender-dominated ads (b), and age-dominated ads versus the percent spent on age-dominated ads (c). A marker that falls above the $x = y$ dashed line indicates a funding entity that spent a disproportionate amount on their set of X-dominated ads.

ads in the same way we define regionally-dominated ads; for instance, an ad that receives ≥99% of its impressions from a singular age range is considered age-dominated. Therefore, aside from displaying a particularly high degree of clustering, regionally-dominated ads seem to be of particular importance to the funding entities themselves, and thus worthy of our focus.

2 Regional dispersion/concentration

To verify that our definition of regionally-dominated ads is robust, we also considered the dispersion of impressions across regions by computing the Gini coefficient of the region vector of each ad. In Fig. S3, we show the distribution of Gini coefficients (b), alongside the distribution of maximum percentages ((a); same as Fig. 4(a) in the main text). We see similar distributions, wherein most ads are either regionally concentrated (Gini close to 1 and maximum percentage close to 100) or not (maximum percentage close to 0 and Gini close to 0.5). For example, the following region vector has a Gini coefficient of 0.45 and a maximum percentage of 7:

```
[(0.07008, 'Texas'), (0.063446, 'New York'), (0.046777, 'Florida'), (0.043545, '
    Massachusetts'), (0.040313, 'Iowa'), (0.038782, 'Michigan'), (0.038612, '
    California'), (0.038442, 'Pennsylvania'), (0.037932, 'North Carolina'),
    (0.037591, 'Virginia'), (0.03572, 'Ohio'), (0.03521, 'Washington'), (0.034189,
    'Colorado'), (0.031298, 'Illinois'), (0.025685, 'South Carolina'), (0.025344, '
    Minnesota'), (0.021943, 'New Hampshire'), (0.021772, 'Georgia'), (0.021772, '
    Nevada'), (0.021092, 'Tennessee'), (0.020071, 'Indiana'), (0.019901, 'Wisconsin
    '), (0.01718, 'Missouri'), (0.016329, 'Oregon'), (0.015989, 'New Jersey'),
    (0.015819, 'Maryland'), (0.015479, 'Utah'), (0.014288, 'Arizona'), (0.012927, '
    Connecticut'), (0.012077, 'Maine'), (0.011396, 'Oklahoma'), (0.008675, '
    Arkansas'), (0.008165, 'Vermont'), (0.007995, 'Washington, District of Columbia
    '), (0.007995, 'Alabama'), (0.007824, 'Kentucky'), (0.006634, 'Louisiana'),
    (0.006464, 'Nebraska'), (0.006123, 'Kansas'), (0.004933, 'Rhode Island'),
    (0.004423, 'Idaho'), (0.004423, 'New Mexico'), (0.004082, 'Alaska'), (0.003912,
     'Montana'), (0.003742, 'Hawaii'), (0.003742, 'West Virginia'), (0.002722, '
    Mississippi'), (0.002041, 'Delaware'), (0.001871, 'South Dakota'), (0.001701, '
    North Dakota'), (0.001531, 'Wyoming')]
```

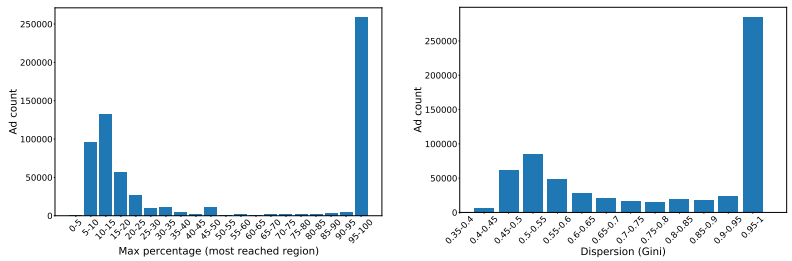
while the following vector has a Gini coefficient of 0.98 and a maximum percentage of 99:

```
[(0.999833, 'New Hampshire'), (0.000167, 'Massachusetts')].
```

Note that regions with zero impressions are not shown in the vector, but were included in the calculation of the Gini coefficient for consistency.

In Fig. S4, we show the relationship between the Gini coefficient and the maximum percentage of each ad (a), as well as the Gini coefficient versus the number of regions impressed upon for each ad (b). We can see that that a maximum region percentage of $\geq 99\%$ is firmly correlated with a Gini coefficient close to 1, indicating that these two measures are strongly related.

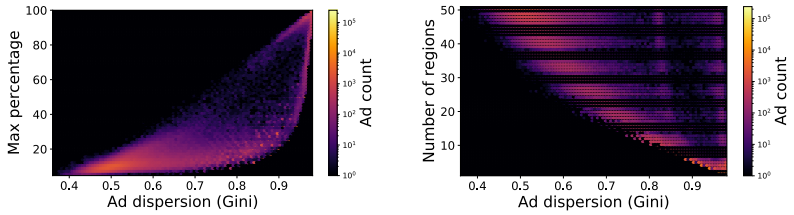
Finally, in Fig. S5(a), we show the number of funding entities with a given fraction of "regionally-dominated ads," defined as those with a maximum percentage $\geq 99\%$. In comparison, we show the number of funding entities with a given fraction of ads with a Gini coefficient ≥ 0.95. We see that both distributions are qualitatively the same, and partitioning funding entities based on the Gini coefficient would yield a similar partition to the one we obtain in the main text.



(a) distribution of ads' maximum percentage

(b) distribution of ads' Gini coefficient

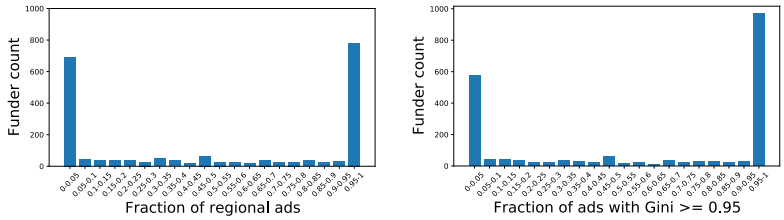
Figure S3: Distribution of ad dispersion. Each ad is mapped to a maximum percentage (indicating the percentage of impressions coming from the most impressed upon region), as well as to a Gini coefficient measuring the dispersion of impressions across regions. (a) number of ads versus maximum percentage; (b) number of ads versus Gini coefficient.



(a) maximum percentage versus Gini coefficient

(b) number of impressed regions versus Gini coefficient

Figure S4: Relationship between Gini coefficient and maximum percentage: (a) heat map of the maximum percentage versus the Gini coefficient per ad; (b) heat map of the number of regions hit versus the Gini coefficient per ad.



(a) distribution of fraction of regional ads per funder

(b) distribution of fraction of ads with Gini coefficient >= 0.95 per funder

Figure S5: Distribution of each funding entity's fraction of regional ads. (a) regional ads defined as those with a maximum percentage ≥ 99, just as in the main text; (b) regional ads defined as those with a Gini coefficient ≥ 0.95.