Socializing Alone: How Online Homophily Has Undermined Social Cohesion in the US

Ruben Enikolopov[1], Maria Petrova[2], Gianluca Russo[3], and David Yanagizawa-Drott*[4]

[1]ICREA-UPF & IPEG & BSE
[2]ICREA-UPF & IPEG & BSE
[3]UAB & BSE
[4]University of Zurich

February 15, 2024

Abstract

Online social networks have changed how people interact across large distances. We examine the long-run effect of a key feature of these networks - online homophily - on interpersonal interactions in local communities. Using Facebook data, we measure online homophily across counties in the United States. To identify effects, we exploit a conflict between Facebook and Google in the early expansion phase of Facebook, which induced persistent variation in online homophily across counties. Using various measures of both online and offline behavior, we find evidence that local social interactions have been affected. In particular, homophilic connections made people use Facebook more often but socialize less offline, as measured through bar, restaurant, and live sports events visits. Individuals have also become less connected across income strata. Political opinions within counties became more diverse, with a lowered probability that two voters in a county support the same political party. Overall, our results indicate that when a natural demand for connecting with socially similar people is met by the supply of a 'death-of-distance' technology, it comes at the cost of short-distance social cohesion.

Keywords: Social Media, Networks, Homophily, Social Capital

*Authors are in alphabetical order. Enikopolov is an ICREA Professor at Universitat Pompeu Fabra, IPEG, and Barcelona School of Economics. Petrova is an ICREA Professor at Universitat Pompeu Fabra, IPEG, and Barcelona School of Economics. Russo is Postdoctoral Fellow at University Autonoma Barcelona and Barcelona School of Economics. Yanagizawa-Drott is a Professor at the University of Zurich. We are grateful to Matt Gentzkow, Lisa George, Ro'ee Levy, Joan Monras, Giacomo Ponzetto, Guido Tabellini, Marco Tabellini, Josh Tucker, Katia Zhuravskaya, and seminar participants at Baryreuth, Chicago, Mannheim, Meta, Opportunity Insights, Princeton, Stanford, Trinity College Dublin, University College London, UPF, UWA, NYU CSMaP 2021 Annual Conference, "Social Media and Democracy" Munich Lectures, CEPR Media and Technology conference in Rome, and Behavioral Political Economy conference at LSE for their valuable feedback. We thank Nikita Belyi for his excellent research assistance.

Internet and social media have been changing people's lives for decades. Early studies suggested that online interactions can bring people closer, cause "the death of distance" (Cairncross, 2002), transform the world into a "global village" (Alstyne and Brynjolfsson, 2005), bridge gaps, and unite communities. However, there is also an argument that online connectivity may lead to fragmented interactions and divide, creating filter bubbles, or "echo chambers" (Sunstein, 2001, 2007). What happens in the real world likely depends on how the Internet and social media change the patterns of social interactions. Network literature uncovered that social interactions are normally characterized by homophily, i.e. the tendency of like-minded people to form connections with each other, taking the degree of homophily as a given characteristic of network formation (Bakshy et al., 2015; Conover et al., 2021; Tarbush and Teytelboym, 2012). At the same time, the causal effects of homophily remain unclear. In this paper, we aim to fill this gap and estimate the causal impact of homophily of interpersonal connections in social media on online and offline interactions, local social capital, and the distribution of political preferences.

More specifically, we study the impact of the network structure of county-to-county Facebook links on various types of interpersonal interactions and political preferences across the U.S. counties. For identification, we use a natural experiment: a conflict between Google and Facebook that made Gmail users less connected to each other if they joined the network between November 2010 and April 2012. By comparing connections with like-minded and different-minded counties, we create a source of quasi-exogenous variation for the network homophily of county-to-county connections. We document that this variable positively predicts average online homophily, measured as an average similarity between counties, weighted by their Facebook links between each other. Then, we look at whether conflict-driven online homophily affects the intensity of use of Facebook, usage of other social media, offline interpersonal contacts, as measured by visits to bars, restaurants, live sports events, and amusement parks, and friendship links across income strata, i.e. "economic social capital" (Chetty et al., 2022b). Finally, we study the implications of these changes on the dispersion of local political preferences.

Our ultimate goal is to study the impact of online network structure. However, this task is not trivial, because friendship links are highly endogenous to the preferences of users who self-select into networks characterized by high degrees of homophily. Homophily is one of the most salient features of social networks that spontaneously form both offline and online, as people are more likely to form connections with those who resemble them in terms of race, socioeconomic status, political preferences, and other attributes (see (Jackson, 2008) for an overview). We begin by creating a measure of online homophily, by weighting social distance between two counties by their Facebook friendship links. For identification, we then exploit a conflict between Facebook and

Google in November 2010 over data sharing. Before the conflict, new users of the network were offered to import their email contact data to Facebook and to easily connect to their contacts who were already on Facebook. After the conflict, new Facebook users lost the ability to use the API to automatically import their Gmail contacts on Facebook. At the same time, new users, who were not using Gmail to sign up for Facebook, were not unaffected by the Facebook-Google conflict. Hence, after November 2010, new Facebook users were less likely to connect to each other if both of them had a Gmail account. Facebook only provided the "Find friends through email" feature for its new users till April 2012, thus we expect our natural experiment to work this way only till April 2012.

To create a source of quasi-exogenous variation in the homophily of Facebook network connections, we create a measure of Gmail complementarity between different counties, by looking at the relative popularity of Gmail and other email clients across different counties in different moments in time. We document a persistent decline in bilateral county-to-county Facebook connections when Gmail complementarity is computed after November 2010, taking county-fixed effects into account, but not before November 2010. This helps us to document that Google-Facebook conflict had a long-term effect on the patterns of bilateral friendship links. We then compare Gmail complementarity before and after the conflict for counties with high and low social distance from each particular county. As a result, we find that some counties got connected to like-minded counties (i.e. low social distance counties), while others got connected to different-minded counties as a result of a change induced by Google policy change. More specifically, we create a variable called Gmail Homophily Shock (GH shock) that subtracts the average Gmail complementarity with like-minded counties from the average Gmail complementarity with different-minded counties between November 2010 and April 2012. We control for the corresponding differences in Gmail complementarity in the pre-conflict period.

With this variable at hand, we proceed to study the impact of online homophily on social interactions. First, we test and confirm that this policy change had a persistent effect on the degree of online Facebook homophily in US counties, here online homophily is an average social distance to other US counties weighted by their Facebook connections with a given county. Then, we document that Gmail complementarity *before* November 2010 did not have a significant impact on online homophily in 2016, but GH shock *between* November 2010 and April 2012 had a strong positive effect on subsequent online homophily. This effect becomes much smaller quantitatively if Gmail complementarity is computed *after* April 2012, when "Find friend by email" stopped working, but long-term consequences are, nevertheless, likely to remain.

Second, we study how GH shock affects various online and offline interactions. Based on Com-

Score (2016), we document that people spend more time on Facebook if they live in a county with a higher homophily shock, i.e. in a county that was pushed to have more like-minded connections for exogenous reasons. Furthermore, we find that higher Facebook homophily implies less visits to other social media, such as Twitter, Reddit, Instagram, etc. Similar to the results for the first stage, we also document an inverted U-pattern for the effect of Gmail Homophily shock: it is not significant before November 2010, it is positive (negative) and significant for Facebook (other social media) visits between November 2010 and April 2012, and it is again small, though showing some degree of persistence, after 2012. We obtain similar estimates for the time (in minutes) spent on Facebook and other social media. Moreover, we find that both GH shock (in a reduced form) and online homophily (in an IV specification) increase the total number of minutes spent on social media.

One important question is whether our GH variable is a shock to homophily or a shock to the total number of friends. By construction, we always control for the county's own preferences for email clients in the pre-2010 time period, and we look at the difference between high- and low- social distance counties. However, it could be the case that our GH shock also positively predicts the total number of connections. Ultimately, it is an empirical question. We do not find a significant effect of GH shock on the county's own total connections in either direction, while the impact of GH shock on the measure of online homophily is positive and highly significant. We conclude that the impact of our shock on online homophily is a first-order effect, while there is no evidence that the impact of this shock on the total number of connections is quantitatively important.

Third, we use SafeGraph data from 2019 to study the impact of online Facebook homophily on offline interactions. We classify different kinds of establishment visits by type. We find that higher GH shock led to a persistent decline in visits to places with a high degree of social interactions, such as bars, restaurants, live sports events, and amusement parks. We did not find significant effects for most other places, and we find more visits to recreational venues (mostly gyms). These findings together are consistent with the following chain of events: in places with higher online Facebook homophily, people spend more time on Facebook, less time on other social media, more time on social media in total, and less time socializing offline with their friends and families.

Fourth, we check if changes in online homophily were associated with actual changes in social capital. We use the data on "economic social capital" from Chetty et al. (2022a),Chetty et al. (2022b), i.e. the probability that people form connections across income strata (e.g. the rich connect to the poor). We document that higher GH shock led to lower economic connectedness.

Fifth, we study the implications of higher (Gmail-induced) homophily online for the distribution of political opinions. Online homophily can affect the distribution of political opinions in two ways.

First, by focusing on communication with like-minded people, and getting constant reinforcement of their pre-existing political preferences, politics could become more extreme, and within-county voting behavior becomes more one-sided. Second, by limiting offline within-county communications, voting preferences within counties could become less similar and more diverse.

We find that in counties, exposed to like-minded countries because of Gmail Homophily Shock, political opinions become more diverse or, in other words, less homogeneous. In other words, we find evidence that political preferences We also find that exposure to the Gmail Homophily Shock decreased the probability of extreme vote margins and increased within-county measures of dispersion of political opinions, such as inter-quartile range and standard deviation of vote shares. Furthermore, we document that GH shock made people less extreme in answering survey questions: CCES respondents were less likely to say that they are "Strong Democrats" or "Strong Republicans."

One sanity check that we do is to show that most of our estimates are weaker in places with a larger share of Facebook connections coming from within their own county. This finding is consistent with the idea that Gmail Homophily shock should have a stronger impact on the network of connections if more of these connections are coming from other counties. Note that the share of own county links did not significantly depend on GH shock, thus it is possible to estimate those specifications. Similarly, we document stronger impact of Gmail Homophily shock in urban areas.

Finally, we document some further null results. We find that GH shock did not benefit either party significantly; the policy change seems to matter for the dispersion of political preferences rather than for the mean. We also document that GH shock did not significantly affect turnout.

Overall, we conclude that the impact of online homophily, estimated with the help of heterogeneous treatment induced by the Gmail-Facebook conflict, was important for the patterns of social media consumption, interpersonal communications, local social capital, cohesion, and political opinions. Thus, our results suggest that technologies capable of transforming the world into a "global village" lead to the unraveling of traditional community bonds.

We contribute to several strands of literature. First, we add to the growing literature on the impact of the internet and social media. Recent literature suggests that exposure to the internet and social media can change economic and political outcomes (Zhuravskaya et al., 2020). Mobile internet and social media positively affect protest participation (Enikolopov et al., 2020; Manacorda and Tesei, 2020), happiness and welfare (Allcott et al., 2020), political polarization, albeit with different results (Barbera, 2020; Boxell et al., 2017; Levy, 2021; Melnikov, 2022; Nyhan et al., 2023), mental health (Braghieri et al., 2022), hate crime and xenophobia(Müller and Schwarz, 2020, 2023; Bursztyn et al., 2024), turnout (Bond et al., 2012), or the trust in government (Guriev et al., 2021). Internet and social media penetration also affected voting outcomes (Fujiwara et al.,

2023; Falck et al., 2014; Campante et al., 2017). There is an ongoing debate on how fact-checking, clicks, and overall regulation of social media could prevent misinformation spread (Barrera et al., 2020; Henry et al., 2022; Guriev et al., 2023). We contribute to this literature by studying the causal impact of homophily in social media, rather than the presence/absence of social media/Internet. Our findings, moreover, help to reconcile some conflicting evidence in this literature.

Second, our work relates closely to the literature on networks and homophily in the networks. Preferences for homophily can increase integration as a result of a random search for friends-of-friends (Bramoullé et al., 2012), while at the same time preference for same-type and biased matching both increase homophily in offline student networks Currarini et al. (2009). The literature documents that online connections are characterized by homophily. People on the Internet mostly interact with like-minded content (Sunstein, 2001, 2007). There is some evidence on homophily online: Bakshy et al. (2015) finds that homophily limits exposure to cross-cutting content. Homophily can affect the diffusion and exposure to like-minded information (Halberstam and Knight, 2016). There is a limited connectivity between right- and left-leaning users (Conover et al., 2021). A the same time, social media allows people to connect to like-minded people when they cannot find them offline (Enikolopov et al., 2021). Langtry (2023) provides a theoretical underpinning of our argument: the more time people spend on out-group connections, the less they provide for the local public good. We contribute to this literature by studying the causal impact of online homophily.

Finally, we contribute to the literature on social capital. "Bowling Alone," a famous Putnam (2000) book, documents the reduction in social capital in the United States in recent years. Social capital seems to be important for governance, democracy, and economic development (Muraskin, 1974; Putnam et al., 1994; Guiso et al., 2004, 2016). Traditional media can reduce social capital and turnout (Gentzkow, 2006; Campante et al., 2022). Social media data could be used to create measures of local social capital (Chetty et al., 2022a,b), while broadband availability might decrease social capital (Geraci et al., 2022) or have positive or no effect (Bauernschuster et al., 2014). Our contribution to this literature is that we study the causal impact of homophily in social media on patterns of offline communications and social capital; our findings also help to reconcile some evidence in earlier studies.

The rest of the paper is organized as follows. Section 1 summarizes the data sources we use. Section 2 discusses our empirical strategy. Section 3 presents results on the usage of Facebook and other online and offline sources of information. Section 4 shows the implications of online homophily on local political consensus. Section 5 concludes.

1 Data

This section describes the sources of the data and construction of the measures used in the analysis. The main unit of analysis is the US county. In few instances the data is available only at the designated market area (DMA) level. We match it to counties using a crosswalk based on population weights.

1.1 Data Sources

Social Connectivity Index. To measure connections between different counties we use information on Facebook users and their friendship networks provided by Facebook Research and described in Bailey et al. (2018).[1] The measures of connectedness are available for 2016 and 2020. The main measure of social connectedness between two counties equals to the number of Facebook connections between users from these two counties, dived by the product of the number of Facebook users in each of the counties (for the 2020 data) or the product of the population of the two counties (for the 2016 data). The measure is scaled to have a fixed maximum value (by dividing the original measure by the maximum and multiplying by 1,000,000,000) and the lowest possible value of 1. Locations are assigned to users based on their information and activity on Facebook, including the stated city on their Facebook profile, and their device connection information.

Email Clients Relative Popularity. To measure relative popularity of different email services across time and space, we use Google's Search Volume Index (SVI) at DMA-level at the quarterly level between 2006 and 2016 for Gmail, Yahoo! Mail and Hotmail.

Demographic and Political County Characteristics. We extract data from the US census (Manson et al., 2021) on demographic and socioeconomic characteristics at the county-level in 2000 and 2010. The data contains the following information: percentage of Whites, Blacks, Hispanics, those with at least college education, average and median income, total population, percentage of rural population, median age, percentage in labor force, and percentage unemployed.

We extract county-level presidential electoral results (1996-2020) from Leip (2021). Precinct-level vote shares for the 2016 presidential election come from Kaplan et al. (2022). To measure the ideology of US counties, we exploit polls from US Tracker Gallup (Gallup, 2017). We collapse individual level self-assessed ideology (ranging from 1 "very liberal" to 5 "very conservative") from 2008, 2009 and 2010 at the county-level.

[1] The data can be downloaded at https://dataforgood.facebook.com/dfg/docs/methodology-social-connectedness-index

Social Media Usage. We collect data from ComScore to gauge social media usage (ComScore, 2016). The data covers the first three months of 2016 and we use it to construct the number of visits in a county to the most relevant websites for our analysis: Facebook and other social media, which include Twitter and Instagram.

Offline Activity. Data on visits to bars and restaurants is from SafeGraph.[2] The data we obtain provides information on visits to several types of commercial establishments for all 2019. We aggregate the data at the county by month level cross-walking data from the census block-level to the county-level.

Economic Connectedness We use data from (Chetty et al., 2022a,b) to measure the degree of economic connectedness in US counties. From this source we borrow the baseline definition of economic connectedness across socioeconomic status (SES). This is constructed as two times the share of high-SES Facebook friends among low-SES individuals, averaged over all low-SES individuals in the county.

1.2 Measure of Social Similarity

To measure how similar are people living in different counties, we look at how close they are in terms of their demographic characteristics and political preferences. In particular, for each county pair we calculate differences in terms of their demographic characteristics (as measured by the percentage of Whites, Blacks, Hispanics, those with at least college education, median income, total population, percentage of rural population, median age, percentage in labor force, and percentage unemployed), their political preferences (as measured by the share of Republican votes in 2004), and their ideology score (as measured by the county average self-assessed ideology from Gallup polls of 2008-2010 that ranges from 1 very liberal through 5 very conservative). We than take the first principal component of these twelve differences and use its inverse as the measure of social similarity between each pair of counties, $Social_Similarity_{ij}$.

1.3 Measure of Online Homophily

To construct a county-level measure of the homophily of online connections, for each county i we take the weighted average of social similarity to all counties j it is connected to, using the share of

[2]SafeGraph is a data company that aggregates anonymized location data from numerous applications in order to provide insights about physical places via the SafeGraph Community. To enhance privacy, SafeGraph excludes census block group information if fewer than five devices visited an establishment in a month from a given census block group.

Facebook connections as weights. Formally, we compute

$$Online_Homophily_i = \sum_{j=1}^{J} \pi_{ij} Social_Similarity_{ij} \tag{1}$$

where π_{ij} is the number of Facebook friendship connections between county i and county j relative to the total number of Facebook friendships of county i. We construct our baseline measure of online homophily using Facebook connections from 2016. Figure 1 maps online homophily for every county in the United States, with green color meaning higher levels of online homophily.

[Figure 1 about here.]

2 Empirical Strategy

In this section, we summarize our empirical strategy. We describe the construction of the key variable that we use as a source of quasi-exogenous variation. We also summarize some descriptive evidence for this variable and show how it relates to online homophily (implied first stage).

2.1 Empirical Challenge and IV Construction

We are interested in studying the effect of online homophily on social media usage, offline behavior, social capital, and local political preferences. Unfortunately, online homophily is likely to be endogenous, as many local characteristics could simultaneously affect both online homophily and our outcome variables. For instance, higher online homophily might reflect self-selection into networks of counties with similar characteristics such as political or ideological preferences, tolerance to other's opinions, and the extent of inter-group contact. These factors might also separately affect our outcomes of interest. Thus, we need an identification strategy that takes care of these possible relationships.

Homophily in social networks is driven by two complementary mechanisms (Feld, 1982; Currarini et al., 2009; Chetty et al., 2022b): differences in exposure (i.e., that people are more likely to meet with more similar individuals) and differences in friending bias (i.e., that they are more likely to form a friendship with more similar individuals after meeting with them). To identify the effect of online homophily we exploit variation in exposure to potential Facebook friends caused by the conflict between Google and Facebook in 2010, which changed the way Facebook suggested friends to new joining users. We show that this variation led to long-term changes in the resulting homophilyy of Facebook connections, indicating that this shock was not fully compensated by endogenous

friendship patterns[3] when users employed alternative methods of searching for friends.

In the next paragraphs, we discuss the conflict between Google and Facebook in 2010 in greater detail and explain how we use it for identification purposes.

Google-Facebook conflict. In the early days of Facebook, new users could use their email contacts to expand their Facebook networks. Figure A.2 shows how a typical entry window looked before and after the 2010 conflict. The window prompted users to type in their emails and passwords so that the program could quickly tell them which of their email contacts were already on Facebook so that they could expand their network from the very beginning. However, in November 2010, Google started invoking reciprocity from Facebook, refusing to share information about Gmail contact of Facebook users without getting information on Facebook users in return (Bodle, 2011). This asymmetry between Google and other email clients lasted until April 2012, when Facebook took down the entry window altogether and switched to algorithmic recommendations of friends. As a result, before November 2010, it was equally easy for people to get connected regardless of their email client, while between November 2010 and April 2012, it was more difficult to do it if both users had Gmail relative to other email clients. The conflict between Google and Facebook was widely covered in the media, see Figure A.3 for an example of the headlines.

To proceed, we note that the relative popularity of different email clients has been changing over time. Back in 2006, the most popular email client was Hotmail. In 2016, Gmail became the most popular one. In between, Yahoo! Mail was the most popular one for some time, with a spike in user interest back in 2010. In Figure 2, we show the evolution of the relative popularity of these three email clients over time, while Figure A.1 presents the geographic distribution of this popularity across the US at different moments in time. We measure the popularity of different email options by employing Google searches for different email clients, thus using Google Search Volume Index (SVI) to proxy for users' interest in various email clients. Interestingly, during the first quarter of the conflict, the relative popularity of all three top email clients was approximately the same.

[Figure 2 about here.]

Gmail Complementarity Shock. Before constructing our instrumental variable, we start by investigating if indeed Facebook connections responded to Gmail-Facebook conflict. We hypothesize that Gmail users had a smaller probability of forming a friendship connection with other Gmail users after November 2010. Hence, we expect that county pairs where Gmail was a more popular

[3] see Ugander et al. (2012) for the study of the friending bias in Facebook.

email client (relative to Yahoo! and Hotmail) in both counties experienced a decrease in the number of Facebook connections as compared with other county pairs with different email client preferences.

Ideally, we would use panel data to inspect Facebook links right around the time of the Google-Facebook incident. Unfortunately, this data before 2016 is not available. Nonetheless, we can test whether 2016 Facebook connections still experienced a decline in connections in the county pairs that had a higher Gmail complementarity right after the 2010 incident relative to Yahoo! and Hotmail. More specifically, we estimate the following equation 2.

$$\pi_{ij} = \alpha + \theta_t gmail_{jit} + \gamma_i + \mu_j + \epsilon_{ij} \tag{2}$$

here π_{ij} is the number of Facebook links between county i and county j in 2016, $gmail_{jit}$ is the Gmail complementarity between counties i and j relative to other email clients in quarter t, γ_i and μ_j are county fixed effects, and ϵ_{ij} is an error term.[4]

Figure 3 plots the estimates of the coefficients θ_t from equation (2) as a function of time t. We find a sharp and significant decline in Facebook connections right after 2010. At the same time, before November 2010, these coefficients were mostly insignificant and had been switching signs around zero. This result is consistent with our hypothesis that the Google-Facebook conflict introduced a discontinuous negative shock for counties with high joint levels of relative Gmail complementarity.

[Figure 3 about here.]

Construction of the Instrumental Variable. We now exploit the Gmail complementarity shock to construct our instrumental variable. We note that when two counties both used Gmail after November 2010, compared to before November 2010, finding friends was more difficult on the margin (higher cost). We also note that Facebook took down the automatic import of contact from any email client after April 2012, effectively ending the window of asymmetry between Gmail and other email clients regarding Facebook and contact importing. Intuitively, large or small Gmail complementarity between November 2010 and April 2012 can mitigate or, on the contrary, exacerbate online homophily between two counties. We argue that if the Gmail complementarity happens to be larger in like-minded counties than in distant-minded counties, it will be harder to connect with similar counties. To formalize this intuition, we construct the Gmail Homophily Shock

[4]We define Gmail complementarity between two counties as the difference between the complementarity in gmail relative to the complementarity with the other email clients

$$gmail_{jit} = gmail_{it} \times gmail_{jt} - .5(yahoo_{it} \times yahoo_{jt} + hotmail_{it} \times hotmail_{jt})$$

(GH_shock_i) in two steps. First, we compute a distance between US counties using the same twelve characteristics we used in the construction of our online homophily measure. Employing this distance, and for each county, we divide US counties between high and low social distance counties.[5] Second, we define our Gmail Homophily Shock as the difference in the average relative Gmail complementarity between high- and low-distance counties between November 2010 and April 2012:

$$GH_shock_i = \sum_{t=1}^{6} \overline{gmail_{it}}^{HI} - \overline{gmail_{it}}^{LO} \tag{3}$$

Here

$$\overline{gmail_{it}}^{d} = \frac{1}{N} \sum_{j=1}^{N} (gmail_{ijt} | SocSim_{ij} = d), \quad d = \{HI, LO\}$$

where d indicates high ($d = HI$) or low ($d = LO$) social distance counties.

This is a key variable in our analysis. Essentially, this variable compares whether county i gets higher complementary to counties with high social distance or low social distance because of the change induced by the Facebook-Google conflict.

[Figure 4 about here.]

For example, Figure 4 shows how we construct the Gmail Homophily Shock for Blount county, AL. The left bar on the graph represents the value of $\sum_{t=1}^{6} \overline{gmail_{it}}^{HI}$ term from equation (3), while the right bar represents $\sum_{t=1}^{6} \overline{gmail_{it}}^{LO}$ term from equation (3). Since the first term in equation (3) is higher than the second term, Blount county is more likely to be connected to counties with low social distance for quasi-random reasons. The construction of the variable for Blount county, AL, is further illustrated in Figure A.4, which plots the Gmail Complementarity between Blount county and all the counties in the US.

Finally, we repeat this exercise for all the counties in the United States. Figure A.5 shows the value of the two different terms in equation (3) in every county in Alabama.[6] As one can see, Gmail-Google conflict introduced heterogenous changes to county-to-county homophily in different counties in the United States: in some counties, the shock, as computed in (3), turn out to be positive, while in others, it is negative. Indeed, the first bar is higher than the second bar for some counties, but for other counties it is the other way around. This translates in a higher complementarity shock with similar counties than with socially distant counties for the first group

[5] In our baseline definition, we split the data in terciles, but check the robustness of our results to other cuts of the data.

[6] The variation of the final variable we construct, the difference between the two bars for each county, is shown in Appendix Figure A.6.

and vice-versa for the second group. We expect that a higher complementarity shock among socially similar counties will increase social distance given the difficulties in making Facebook connections we documented above. Conversely, a higher complementarity shock in socially distant counties should facilitate connecting with like-minded counties. We test this hypothesis in the next subsection.

2.2 Gmail Homophily Shock and Homophily of Online Connections. Implied First Stage

In this sub-section, we check whether Gmail Homophily Shock, the variable we just constructed, is a good predictor of online homophily our independent variable described in subsection 1.3. Formally, we can do so by estimating equation 4.

$$Y_i = \beta_0 + \beta_1 GH_shock_i + \beta_2 X_i + \epsilon_i \tag{4}$$

In this subsection, we define $Y_i \equiv Online_Homophily_i$, our measure of online homophily. GH_shock_i represents the Gmail Homophily Shock and X_i is a host of county-level controls, which includes Gmail Homophily Shock (3), computed in pre-period, i.e. 6 quarters before November 2010; ϵ_i is an error term.

We report our estimates of equation 4 in Table 1 where we gradually add our controls. More precisely, our Baseline Controls include basic demographic and political characteristics: share of whites, share attended college and share unemployed in 2010; turnout and Republican vote share as of 2008. The Demographic Controls include: share Black, share Hispanic, log median income, share in the labor force, share rural, and median age in 2010. Political Controls further include political homogeneity in 2008 to the list of controls.[7] Demographic Trends include differences between 2010 and 2000 of all baseline, demographic, and political controls. In addition, we define the pre-period Gmail complementarity using the last ten quarters before the start of the conflict between Facebook and Google. To compare counties on similar trends over time, we include this control in our set of baseline controls. Finally, to facilitate the interpretation of coefficients we standardize our independent variables. Standard errors are clustered at the state level.

In column 1, where we only control for log population and pre-period Gmail complementarity, the relationship between online homophily and the Gmail Homophily Shock is .525, positive and significant at the 1% level. As we start adding our Baseline Controls (column 2), DMA fixed effects (column 3), Demographic Controls (column 4), and Political Controls (column 5), the coefficient

[7]Political homogeneity is defined as one minus the Herfindal Index calculated using Democrat and Republican vote shares. In this case, assuming the Republican and Democrat vote share sum up to 1, the formula boils down $1 - 2r(1-r)$, where r is the Republican vote share. See more details about this variable and the rationale to use it in section 4.

of interest hovers between an increase of .3 and .4 of a standard deviation. The precision of our estimates is unchanged throughout. In column 6, we add trends in our controls and the point estimate converges to 0.276, significant at 1%. Finally, in column 7 we obtain similar results if we construct our dependent variable, online homophily, using 2020 Facebook links. The point estimate is indeed equal to .236, i.e. it has the same order of magnitude as column 6. We map the residual variation (from the specification in column (4)) in Figure 5. As one can see, there is a high degree of heterogeneity in this variable both within and across American states.

The magnitude of our preferred specification (column 6) implies that one sd increase in the Gmail Homophily Shock increases online homophily by about 28% of a standard deviation. We report the Kleibergen-Paap F-stat and, since its values in most specifications always lie around 20 or more, we do not need to use weak instrument robust methods.

[Table 1 about here.]

[Figure 5 about here.]

Although we cannot distinguish between strong and weak links in Facebook connections, we expect that this variation in the friend suggestion policy affects predominantly the formation of weak links, since strong links are likely to be established through active search for friends regardless of the friend suggestion policy. This is especially relevant for cross-county connections that we study in our paper since within-county connections are much more likely to reflect the structure of offline social networks (Chetty et al., 2022b). However, consistent with the "Strength of Weak Ties" hypothesis (Granovetter, 1973) it has been shown that in online social networks, weak links play an important role in affecting the spread of information (Bakshy et al., 2012) and affecting offline outcomes, such as job mobility (Rajkumar et al., 2022) or housing behavior (Bailey et al., 2018) (Bailey 2018). Thus, significant changes in the structure of weak links can have a substantial effect on the behavior of Facebook users.

Before exploring the effects of our Gmail Homophily Shock, we document that the variation we exploit is balanced on county-level pre-determined characteristics. We carry out our balance test by estimating equation 4, where we put on the LHS one by one our controls. We take them out of the RHS when we use them as outcome variables. Figure 6 plots the estimated coefficient of our balance test. While some coefficients appear unbalanced in our more naive specification, most become indistinguishable from zero when we include the rest of the controls. We still observe some small significant relationship with the share of unemployed (negative) and change in median income (positive), nonetheless consistent with the significance levels we would obtain by random

generation. In any case, we control for socio-demographic, economic, and political variables and their changes throughout the paper.

[Figure 6 about here.]

[Figure 7 about here.]

Another way of seeing our identification graphically is to look separately at the relationship between our outcomes of interest and Gmail Homophily shock, computed during our treatment window, i.e. 6 quarters between November 2010 and April 2012, or Gmail homophily variables constructed during 6 quarters before and 6 quarters after our treatment window. Figure 7a shows these results for the implied first stage as three bar graphs, with coefficients for the pre-period, during the treatment period, and after the treatment period shown side by side. As one can see, the relationship between GH shock and online homophily in pre-period is far from being significant, with a coefficient being several times smaller than the standard error, while, at the same time, our coefficient of interest ("treatment" coefficient) is positive and significant, consistent with the results in Table 1. Nevertheless, in what follows, we always control for pre-period, to take into account possible average differences across counties with different baseline levels of Gmail complementarity, i.e. pre-existing Gmail complementarity before policy change.

In what follows, we will focus on the reduced-form relationship between Gmail Homophily Shock, social capital, and local political cohesion, as well as IV estimation where online homophily is instrumented by GH shock. For the reduced form relationship, we simply estimate equation 4 where the LHS is one of our outcomes of interest. Our emphasis throughout the analysis is on the effect of the shock we reconstruct using the Gmail-Facebook conflict. However, we also provide the result of an instrumental variable approach where we use the Gmail Homophily Shock as the IV and our measure of online homophily as the endogenous variable. More precisely, equation 5 presents the second-stage regression we estimate. $Online_Homophily_i$ is the variable we construct measuring online homophily, X_i is the same set of controls as in the (implied) first stage presented above, and ε_i is an error term.

$$Outcome_{it} = \beta_0 + \beta_1 Online_Homophily_i + \beta_2 X_{it} + \varepsilon_{it} \tag{5}$$

3 Results: Online Homophily and Social Interactions

In this section, we study what happens when a county experiences a higher Gmail Homophily Shock. We document an increase in the usage of Facebook together with a decrease in other social media activities. We also uncover a drop in social capital and interpersonal communications.

3.1 Homophily of Online Connections and Usage of Social Media

We start our analysis by studying the impact of the Gmail Homophily Shock and the homophily of online connections on different measures of online activity. We use different measures computed based on ComScore Internet browsing data. We estimate equation (4) with Internet browsing measures in Panel A of Table 2; Panel B of this Table summarizes the results of the second stage estimation (5). Columns (1)-(4) show the results for the (log) number of Facebook visits, while columns (5)-(8) summarize the results for the (log) visits to other social media, such as Twitter and Instagram.[8]

Our most basic specification always includes baseline controls and DMA fixed effects; we gradually add demographic, political, and trend controls. Importantly, we always control for the total number of visits and its square term to account for differences in total online activities across different counties.

[Table 2 about here.]

[Table 3 about here.]

We show that a one standard deviation increase in the Gmail Homophily Shock increases Facebook visits and reduces visits to other social media. In the most saturated specifications (columns 4 and 8), one standard deviation of a *Gmail Homophily shock* leads to 20.3% increase in the number of Facebook visits and 10.3% decrease in the number of other social media visits, with both coefficients being significant at 5%. Similarly, in Panel B, our results imply that one standard deviation higher *Online homophily* increases visits to Facebook by .660 log points (column 4), significant at 1%, whereas it decreases visits to other social media by .334 log points (column 8), significant at 5%. The Kleibergen-Paap F-stat is consistently between 20 and 50, ruling out weak instrument concerns.

In Table A.3 in the Online Appendix, we repeat this exercise using the (log) number of minutes instead of the number of visits. Consistent with Table 2, Gmail Homophily shock increases the number of minutes spent on Facebook and reduces the number of minutes spent on other social media. The magnitudes in the most saturated specifications (columns 4 and 8) imply that one standard deviation increase in GH shock leads to 25.5% increase in the number of minutes spent on Facebook and to 22.4% decrease in the number of minutes spent on other social media.

[8] In the main text, we stick to log(x) specifications for the ease of interpretation, besides, they rarely become zeroes. However, for all log specifications, we report similar results with inverse hyperbolic sine (IHS) transformation of the dependent variables in the Online Appendix, and they are qualitatively similar.

Finally, Table 3 summarizes the results for total visits and total time spent on social media activity, which includes Facebook and other social media as a single variable. The results imply that one standard deviation increase in Gmail Homophily shock increases the total number of social media visits by 18.5% and the total number of minutes spent on any social media by 22.4%.

Another way of seeing these results is to look at the coefficients for Gmail Homophily shock, computed during, before, and after 6 quarters of Google-Facebook conflict (Figure 7b). As one can see, most of the effect happens during the treatment period, with the magnitudes for pre- and post-period being numerically smaller. Note that Tables 2-3 and Table A.3 all include Gmail Homophily shock in pre-period as a control, to take possible baseline differences in Gmail usage and their network patterns into account. Finally, we'd like to emphasize that qualitatively, Figure 7b resembles Figure 7a, i.e. the variation in our outcome variable (Facebook visits) seems to come exactly from the variation in the first stage.

Overall, the results in this section suggest that people, who experienced a positive GH shock, and who have more connections to like-minded counties, indeed like their Facebook feed, go to Facebook more often, and they spend more time there, at the expense of time spent on other social media. Moreover, our findings imply that these people like Facebook so much that they spend more total time on social media, and less time on other, presumably offline activities.

3.2 Gmail Homophily Shock and Total Connections

One important question is whether our GH variable is a shock to homophily or a shock to the total number of friends. By construction, we always control for the Gmail complementarity in the pre-period, and look at the difference between high- and low-social distance counties.

However, it could be the case that our GH shock is positively correlated with the total number of connections. Ultimately, it is an empirical question. In Table A.14, we summarize the results of the estimation of equation (4) with the (log) number of total connections as a dependent variable. The initially significant positive effect of GH shock on total connections disappears once we control for demographics, with a gradual reduction in the magnitude. Based on the most saturated specification (column 6), we can rule out the effects of up to 6.3%. At the same time, the impact of GH shock on the measure of online homophily is positive and highly significant. We conclude that the impact of our shock on online homophily is a first-order effect, while there is no evidence that the impact of this shock on the total number of connections is quantitatively important. Thus, we stick to our interpretation of Gmail Homophily shock as a shock to Online Homophily rather than a shock to the Total Number of Connections.

3.3 Homophily of Online Connections and Offline Activity

In this subsection, we test if a higher Gmail Homophily Shock reduces offline interpersonal interactions. As a proxy for interpersonal interactions, we use visits to various points of interest. We leverage data from Safegraph covering all months of 2019 at the count-by-month level. We estimate specifications (4) and (5) using visits to restaurants, bars, live sports events, and other locations, which are often visited jointly with friends and families. Given the structure of the data, here we add monthly fixed effects to our empirical setup. Further, just like for social media usage, we control for the total number of visits to any establishment and its squared term. Table 4 presents our results on bars and restaurants where we progressively include controls in a way that mirrors our implied first-stage table (Table 1).

In all columns of Table 4, the coefficients for Gmail Homophily shock (Online Homophily in the second panel) are negative and significant, which implies fewer visits in places with higher Facebook homophily. The set of controls changes from the minimum one (population) to the most saturated specification with DMA and monthly fixed effects, political controls, demographic controls, and trends, and the resulting coefficient remains remarkably stable, ranging from 0.053 to 0.08 in different columns. The magnitude implies 7% fewer visits to bars and restaurants, significant at 5% level, based on the most saturated specification (column 6). The IV estimates reported in the same table confirm these results. One standard deviation increase in online homophily reduces our proxy of social contact by .261 log points, significant at the 10% level (column 6). While the F-stats presented here are different from the previous exercise, given the different samples, they are still consistent with a strong IV scenario.

Furthermore, to graphically illustrate how our identification works here, we report the reduced form results for Gmail Homophily shock computed during our treatment period (November 2010-April 2012, Figure 7c), and in periods before and after. As one can see, the coefficients for the pre-treatment and treatment periods have different signs, with the coefficient for the treatment period being large, negative, and significant. The coefficient for the post-period is much smaller numerically, with the coefficient being smaller than the standard error. These results are in line with the implied first-stage results (Figure 7a) in that the coefficient of interest is the largest in magnitude and in terms of significance as compared with pre- and post- coefficients.

So far, we documented that Gmail Homophily shock reduced visits to bars and restaurants. To dig further into the patterns of potential offline interactions, we report similar results for the groups of establishments separately, namely bars, restaurants, theaters, live sports events, museums and historical sites, amusement parks, recreational venues, religious, and voluntary organizations (Figure 8). We document the negative and significant effects of Gmail Homophily shock on visits

to bars, restaurants, live sports events, and amusement parks, and a marginally significant negative coefficient for voluntary organizations. All these results are consistent with a reduction in offline social interactions. The only positive coefficient that we document is for recreational venues, but most of them are gyms, i.e. establishments people mostly visit alone.

[Figure 8 about here.]

[Table 4 about here.]

Overall, our results so far are consistent with the hypothesis that Online Homophily changes the patterns of online and offline social interactions: they spend more time on social media online and meet with their friends and families offline less often.

3.4 Homophily Shock and Economic Connectedness

In this subsection, we investigate what are the implications of the Gmail Homophily Shock for social capital. Exploiting data from Chetty et al. (2022a,b), we look at the "economic social capital" at the county level as the degree of economic connectedness between low and high SES individuals. This variable measures the number of Facebook friends from high socioeconomic status among low socioeconomic status individuals. Table 5 summarizes our results, with controls added gradually.

Columns 1 through 6 indicate that the effect of Gmail Homophily Shock on social capital is consistently negative and significant across specifications. In the most saturated specification (column 6), the coefficient is -0.172, significant at the 5% level. The magnitude implies that a standard deviation increase in the Gmail Homophily Shock reduces economic connectedness by approximately 17% of a standard deviation. As in the rest of our results, Table 5 reports the 2SLS estimates using online homophily as the endogenous variable and the Gmail Homophily Shock as the IV. The point estimates are, again, consistent with the reduction in economic social capital across US communities. The magnitude in the IV estimates is .511, meaning a reduction of approximately 50% of a standard deviation for each standard deviation increase in online homophily.[9]

To further illustrate how identification works in this case, we report the coefficients for Gmail Homophily shock computed during, before, and after the treatment period (i.e. November 2010-April 2012, Figure 7d). As one can see, the pre-coefficient is clearly nonsignificant and has a different

[9]One important caveat here is that Chetty et al. (2022a) and Chetty et al. (2022b) partly use Facebook data to construct their variables. They argue that within-county Facebook connections serve as a good proxy for within-county offline connections. We agree with this notion, but it might weaken our interpretations of the results. However, our homophily measure is entirely based on out-of-county connections. Moreover, as we show below, the share of within-county connections does not seem to be significantly correlated with Gmail Homophily shock. Thus, we do not think that the method of construction of the economic connectedness variable would change the interpretation of our results.

sign from the negative and significant coefficient in the treatment period. Overall, the coefficients in Figure 7d are consistent with the notion that identification is coming from the changes in the treatment period and with the rest of the pictures in this figure.

In sum, the results in Table reftab:EC suggest that Gmail Homophily shock reduces economic connectedness, i.e. the probability that people within a county are connected across income strata, i.e. that the rich and the poor form connections with each other. These results further strengthen our interpretation of what's going on, with Gmail Homophily shock reducing local economic social capital.

[Table 5 about here.]

Overall, the results we document in Tables 2, 4 and 5 are consistent with a rise in online homophilic interactions through Facebook, at the expense of lower real-life contact and social capital. The Gmail Homophily Shock leads individuals to spend more time on Facebook and avoid interpersonal communication in bars and restaurants with negative repercussions on the social capital of local communities. In the next section, we ask whether higher homophily online also affects the cohesion of political opinions.

4 Results: Dispersion of Political Opinions

What are the implications of higher online homophily for the distribution of political opinions across American communities? We answer this question by focusing on the effect of the Gmail Homophily Shock on voting behavior and political opinions across US counties.

4.1 Hypotheses

How can online homophily affect political preferences? There are at least two alternative hypotheses here. First, exposure to like-minded communities can lead to polarization of opinions. As an average voter gets into more and more extreme "echo chambers" online (Sunstein, 2001, 2007), users are becoming more extreme, and, as a result, we expect to see the convergence of local preferences to one extreme or another.

On the other hand, people exposed to a more homogeneous network switch away from other forms of political communication, and, as a result, they might have fewer interactions within local communities (Bursztyn et al., 2024; Langtry, 2023) In the end, social learning implies divergence of preferences "within" counties.

These two classes of theories suggest alternative hypotheses. Using our empirical approach, we can separate between the two and find which hypothesis is more consistent with the data.

4.2 Measure of Political Homogeneity

To proceed, we need to construct some measure of divergence/convergence of political preferences within the county. We can use vote margins (the difference between first two candidates) as one such a measure, but we would like to also introduce a measure of political homogeneity that is symmetric around the center. Formally, *political Homogeneity* captures the degree of local political homogeneity (consensus), i.e., the opposite of local political fractionalization, and is defined as

$$PolHomogeneity_i t = 1 - 2r_i t(1 - r_i t) \tag{6}$$

Here $r_i t$ is the vote share of a Republic party in Presidential elections. Figure A.7 shows how our measure is related to vote shares.

In addition to political homogeneity, we also use several other variables to capture the distribution of political preferences, including vote margin, extreme (>40%) vote margins, very extreme (>60%) vote margins, and several within-county measures based on precinct-level data, such as inter-quartile range and standard deviation.

4.3 Online Homophily and Political Homogeneity

We start by investigating the effect of the Gmail Homophily Shock on the distribution of voting outcomes by using 2020 political homogeneity as a left-hand side variable in Table 6. Here, we present the point estimates of our standard set of six specifications. The estimate of our most sparse specification in column 1 points to a small and not significant effect. However, as we include baseline controls (column 2) we uncover a sizeable drop in political homogeneity as a result of Gmail Homophily shock. The magnitude of the coefficient ranges from 0.024 to -0.049 across the remaining 5 specifications, all significant at 1% level. Numerically, one standard deviation increase in the Gmail Homophily Shock lowers political homogeneity by about 24% of a standard deviation. The results based on 2SLS estimation are numerically and qualitatively similar to the reduced form.

[Table 6 about here.]

In addition to the results for 2020 presidential elections, we further provide evidence for the relationship between Gmail Homophily Shock and political homogeneity in every presidential election between 2000 and 2020. We plot the point estimates, using 2008 as the reference year, in Figure 9. The effect of our Gmail Homophily Shock on political homogeneity is indistinguishable from zero in the pre-2010 period. In 2012, the last year of the Gmail-Facebook incident, we still find a null effect, consistent with a lower frequency of political discussions in social media back then. From

2016 onward, we observe a net jump in the point estimates indicating a significant reduction of political homogeneity. The point estimate in 2020 is the same as in column 6 of Table 6 and it points to a reduction of 24% of a standard deviation.

Figure 7e further illustrates our identification by looking into the relationship between Political Homogeneity and Gmail Homophily Shock, computed during the treatment period (November 2010-April 2012), in the pre-period (6 quarters before November 2010) and in the after-period (6 quarters after April 2012). As one can see, the resulting figure strongly highlights that our identification is coming from the variation in the treatment period. The coefficient for the pre-period is indistinguishable from zero and is several times smaller than the standard deviation. The coefficient in after-period is negative, but much smaller numerically. Overall, Figure 7e goes in line the rest of the graphical evidence in this figure.

[Figure 9 about here.]

4.4 Online Homophily and Within-County Distribution of Political Preferences

In the previous subsection, we showed that the shock in online homophily caused by the Gmail-Facebook incident also decreased political homogeneity between US counties. Here, we ask what happens to the distribution of political opinions within US communities. We exploit precinct-level political returns for 2016 (Kaplan et al., 2022). Exploiting the precinct-level data, we characterize different moments of the distribution of voting shares at the county level. In particular, we construct three outcomes: the standard deviation of the Republican vote share, the inter-quartile range of the Republican vote share, and the share of the precinct with extreme margins. We use two definitions of extreme margins: at least 40 points of difference and at least 60 points of difference. Table 7 outlines the results of this analysis. For each outcome, we present the point estimates of two specifications.

The general message of Table 7 is that the Gmail Homophily Shock widens the dispersion of political opinions within US communities. One standard deviation increase in the homophily shock has three effects on the moments of the Republican vote shares at the precinct level: it enlarges the standard deviation by 0.012 (column 2), it increases the inter-quartile range by 0.023 points (column 4) and it increments the share of precincts with very large vote share margins (column 8).

[Table 7 about here.]

Finally, we check if the variation that generates the negative effect on the vote margin indeed comes from our treatment period. Figure 7f summarizes these results. As one can see, the coefficients for pre- and treatment period have different size. Pre-treatment coefficient is positive,

but very small and far from being statistically significant, while the coefficient for the treatment period is negative and significant. All graphs in Figure 7 are, thus, consistent with each other and with our general claim: that the identifying variation comes from our treatment period and not before (even though we control for pre-period in all the specification). There is some evidence of delayed response in post-treatment coefficients, but they are, nevertheless, much smaller than the treatment coefficients in all the specifications from (a) to (f).

We interpret these results as evidence that what drives lower political homogeneity between counties is lower cohesion in political opinions also within counties. This is consistent with the results we documented in section 3. Higher online homophily disrupts the traditional structure of US communities, lowers socialization in real life, and dampens social capital. In the political arena, the "death of distance" online generates higher polarization of political opinions within communities, driving the drop in political homogenization we observe between US counties.

Overall, our results are consistent with the second hypothesis that we outline: exposure to like-minded communities on Facebook positively affect the dispersion of voting and the divergence of political preferences.

4.5 Online Homophily and Intensity of Political Preferences

So far, we have looked at the dispersion or convergence of political preferences but have not studied their intensity so far. At the same time, the intensity of political preferences might be important, eg the literature suggests that the presence of mobile internet and social media can increase political polarization (Allcott et al., 2020; Levy, 2021; Melnikov, 2022). To look at the intensive margin, we use the data from the Cooperative Congressional Election Study (CCES). More specifically, we create a variable $Extreme_id_{it}$ which denotes respondents who defined themselves as either strong Democrats or strong Republicans.[10]

As the survey is available in multiple waves, we estimate the following difference-in-difference equation, where we presume that the effect of Gmail Homophily shock starts kicking in after 2010

$$Extreme_id_{it} = \beta_0 + \beta_1 GH_shock_i + \beta_2 GH_shock_i \times post_t + \beta_3 X_i + \delta_t + \epsilon_{it} \tag{7}$$

Table A.17 presents results on extreme partisanship while Table A.18 shows evidence on extreme ideology. The results are consistent with people becoming more centrist in places with higher Online Homophily. These results speak to the findings in Gentzkow and Shapiro (2011) that interpersonal

[10]The exact wording in the survey question is "Generally speaking, do you think of yourself as a) Strong Democrat; b) Not Very Strong Democrat; c) Lean Democrat; d) Independent; e) Lean Republican; f) Not Very Strong Republican; g) Strong Republican?" We code $Extreme_id_{it}$ equal to one if the respondent defined herself as either Strong Democrat or Strong Republican.

interactions are more segregated than internet connections.

4.6 Heterogeneity

By construction, our main variable of interest, Gmail Homophily Shock, presumes that at least some connections that people have on Facebook are out-of-county connections. Moreover, our results, theoretically, should be stronger if the share of out-of-county connections is higher. In this subsection, we formally test this claim, with a caveat that the share of online connections might be an endogenous variable. We first show that the share of links outside the county itself is not significantly related to our Gmail Homophily Shock. These results are presented in Table 8. As one can see, in the most saturated specifications (columns 3-6), the relationship between Gmail Homophily Shock and the share of outside connections is small and insignificant. In column 6, we can rule out effects of up to 0.47%, with a mean of the dependent variable being 43.9%. Thus, even though in principle the share of outside connection could be affected by Gmail Homophily Shock, that does not happen in practice.

We then proceed by looking at the effect heterogeneity with respect to the share of connections outside the county (Table 9), for the first stage and major outcomes of our analysis (Facebook visits, bar visits, economic connectedness, political homogeneity, extreme vote margin). In all the specifications, the interaction term with the share of outside links has the same sign to the non-interacted coefficient. The interaction term is significant at the 1% level in all specifications except bars, where it is significant at 10% level. These results provide a sanity check: we expect our results to be larger in places with a higher share of outside links, and this is exactly what we find.

[Table 8 about here.]

[Table 9 about here.]

Furthermore, we repeat this exercise by looking at the heterogeneity with respect to the share of the urban population. In urban areas, people tend to interact a lot and know their neighbors, while in urban areas marginal connections are easier to replace with online ones. Thus, we expect our results to be stronger in urban areas. This is exactly what we find, see the results in Table 10. All interaction terms have the same sign of the main effect, and five out of six interaction coefficients are significant at 1% level. Overall, the results in Tables 9 and 10 are consistent with the idea that the results are stronger in places with more out of county connections and/or places where out of county connections could be formed more easily.

[Table 10 about here.]

4.7 Vote shares and turnout

Finally, we summarize a couple of null results in our study. In theory, it could be possible that exposure to like minded counties on Facebook benefited a particular party; e.g. Fujiwara et al. (2023) shows that the penetration of Twitter benefited democratic party. We do not find similar evidence in our paper: Table A.19 documents no significant effects for voting for Republicans/Democrats in the most saturated specifications. With the magnitude of the effect being 0.001, we can rule out the effects of up to 0.57%, with a mean of dependent variable being 66.4%, thus our results are close to be precisely estimated zeroes.

Similarly, we do not find any significant evidence for turnout. Even though Bond et al. (2012) find positive effect of experimental information in Facebook on turnout, we did not find that exposure to more like-minded content alters political participation in presidential elections. However, the results, reported in Table A.20, are pretty noisy and are far from precisely estimated zeroes, thus we cannot say anything definite about turnout.

5 Discussion and Conclusion

In this paper, we ask what happens to communities exposed, through their online network, to more like-minded communities than they would have originally picked. We used an incident between Gmail and Facebook between 2010 and 2012 to construct exogenous variation in the degree of online homophily of connections. The incident inadvertently facilitated (hindered) friending people from communities with similar (distant) characteristics. We use the resulting exogenous variation generated by the Facebook-Gmail incident to characterize the response of users to the higher extent of homophily in their network. We documented that higher online homophily fundamentally affected the cohesion of American communities in several ways. First, online homophily pushed individuals to consume more Facebook, at the expense of other social media. Then, we document a drop in interpersonal contact, proxied by visits to bars and restaurants, together with the reduction of general social capital. At the same time, the impact of online homophily in the political arena mirrored the drop in social cohesion in American communities. We showed that exposure to homophily online led to higher dispersion of political preferences within counties. We highlighted that the divergence in political homogeneity between counties is driven by the increase in the dispersion of opinions within communities.

From a social policy standpoint, our results uncover an important and ignored aspect of higher online homophily. “Death of distance” technologies (Cairncross, 2002) are not only responsible for transforming the world into a global village (Alstyne and Brynjolfsson, 2005), they also fundamen-

tally undermine communities' social and political cohesion. Policymakers wanting to bring people closer online should keep in mind the trade-offs such technologies introduce for the traditional structure of communities.

References

Allcott, H., L. Braghieri, S. Eichmeyer, and M. Gentzkow (2020, March). The welfare effects of social media. *American Economic Review 110*(3), 629–76.

Alstyne, M. V. and E. Brynjolfsson (2005). Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities. *Management Science 51*(6), 851–868.

Bailey, M., R. Cao, T. Kuchler, and J. Stroebel (2018). The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy 126*(6), 2224–2276.

Bailey, M., R. Cao, T. Kuchler, J. Stroebel, and A. Wong (2018). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives 32*(3), 259–80.

Bakshy, E., S. Messing, and L. A. Adamic (2015). Exposure to ideologically diverse news and opinion on facebook. *Science 348*(6239), 1130–1132.

Bakshy, E., I. Rosenn, C. Marlow, and L. Adamic (2012). The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*, pp. 519–528.

Barbera, P. (2020). *Social Media, Echo Chambers, and Political Polarization*, pp. 34–55. SSRC Anxieties of Democracy. Cambridge University Press.

Barrera, O., S. Guriev, E. Henry, and E. Zhuravskaya (2020). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of Public Economics 182*, 104123.

Bauernschuster, S., O. Falck, and L. Woessmann (2014). Surfing alone? the internet and social capital: Evidence from an unforeseeable technological mistake. *Journal of Public Economics 117*, 73–89.

Bodle, R. (2011). Regimes of sharing. *Information, Communication & Society 14*(3), 320–337.

Bond, R. M., C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler (2012). A 61-million-person experiment in social influence and political mobilization. *Nature 489*(7415), 295–298.

Boxell, L., M. Gentzkow, and J. M. Shapiro (2017). Greater internet use is not associated with faster growth in political polarization among us demographic groups. *Proceedings of the National Academy of Sciences 114*(40), 10612–10617.

Braghieri, L., R. Levy, and A. Makarin (2022). Social media and mental health.

Bramoullé, Y., S. Currarini, M. Jackson, P. Pin, and B. Rogers (2012). Homophily and long-run integration in social networks. *Journal of Economic Theory 147*(5), 1754–1786.

Bursztyn, L., G. Egorov, R. Enikolopov, and M. Petrova (2024). Social media and xenophobia: Evidence from russia. Technical report, NBER Working Paper.

Cairncross, F. (2002). The Death of Distance. *RSA Journal 149*(5502), 40–42.

Campante, F., R. Durante, and F. Sobbrio (2017, 12). Politics 2.0: The Multifaceted Effect

of Broadband Internet on Political Participation. *Journal of the European Economic Association 16*(4), 1094–1136.

Campante, F., R. Durante, and A. Tesei (2022). Media and social capital. *Annual Review of Economics 14*(1), 69–91.

Chetty, R., M. O. Jackson, T. Kuchler, J. Stroebel, N. Hendren, R. B. Fluegge, S. Gong, F. Gonzalez, A. Grondin, M. Jacob, et al. (2022a). Social capital i: measurement and associations with economic mobility. *Nature 608*(7921), 108–121.

Chetty, R., M. O. Jackson, T. Kuchler, J. Stroebel, N. Hendren, R. B. Fluegge, S. Gong, F. Gonzalez, A. Grondin, M. Jacob, et al. (2022b). Social capital ii: determinants of economic connectedness. *Nature 608*(7921), 122–134.

ComScore (2016). Comscore, [Metric], 1/01/2016 - 31/03/2016 [Database].

Conover, M., J. Ratkiewicz, M. Francisco, B. Goncalves, F. Menczer, and A. Flammini (2021, Aug.). Political polarization on twitter. *Proceedings of the International AAAI Conference on Web and Social Media 5*(1), 89–96.

Currarini, S., M. O. Jackson, and P. Pin (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica 77*(4), 1003–1045.

Enikolopov, R., A. Makarin, and M. Petrova (2020). Social media and protest participation: Evidence from russia. *Econometrica 88*(4), 1479–1514.

Enikolopov, R., M. Petrova, and A. Szeidl (2021). Homophily offline and online social networks.

Falck, O., R. Gold, and S. Heblich (2014, July). E-lections: Voting behavior and the internet. *American Economic Review 104*(7), 2238–65.

Feld, S. L. (1982). Social structural determinants of similarity among associates. *American sociological review*, 797–801.

Fujiwara, T., K. MÃŒller, and C. Schwarz (2023, 10). The Effect of Social Media on Elections: Evidence from The United States. *Journal of the European Economic Association*, jvad058.

Gallup (2017). US Tracker Political and Economic and Well Being, 2008-2017 [Database].

Gentzkow, M. (2006). Television and voter turnout. *The Quarterly Journal of Economics 121*(3), 931–972.

Gentzkow, M. and J. M. Shapiro (2011, 11). Ideological Segregation Online and Offline *. *The Quarterly Journal of Economics 126*(4), 1799–1839.

Geraci, A., M. Nardotto, T. Reggiani, and F. Sabatini (2022). Broadband internet and social capital. *Journal of Public Economics 206*, 104578.

Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology 78*(6), 1360–1380.

Guiso, L., P. Sapienza, and L. Zingales (2004, June). The role of social capital in financial development. *American Economic Review 94*(3), 526–556.

Guiso, L., L. Zingales, and P. Sapienza (2016). Long-term persistence. *Journal of the European Economic Association 14*(6), 1401–1436.

Guriev, S., E. Henry, T. Marquis, and E. Zhuravskaya (2023). Curtailing false news, amplifying truth. Pse working papers, HAL.

Guriev, S., N. Melnikov, and E. Zhuravskaya (2021). 3g internet and confidence in government. *The Quarterly Journal of Economics 136*(4), 2533–2613.

Halberstam, Y. and B. Knight (2016). Homophily, group size, and the diffusion of political information in social networks: Evidence from twitter. *Journal of Public Economics 143*, 73–88.

Henry, E., E. Zhuravskaya, and S. Guriev (2022, August). Checking and sharing alt-facts. *American Economic Journal: Economic Policy 14*(3), 55–86.

Jackson, M. O. (2008). *Social and economic networks*. Princeton university press Princeton.

Kaplan, E., J. L. Spenkuch, and R. Sullivan (2022). Partisan spatial sorting in the united states: A theoretical and empirical overview. *Journal of Public Economics 211*, 104668.

Langtry, A. (2023). More connection, less community: network formation and local public goods provision.

Leip, D. (2021). *Dave Leip's Atlas of U.S. Presidential Elections*.

Levy, R. (2021, March). Social media, news consumption, and polarization: Evidence from a field experiment. *American Economic Review 111*(3), 831–70.

Manacorda, M. and A. Tesei (2020). Liberation technology: Mobile phones and political mobilization in africa. *Econometrica 88*(2), 533–567.

Manson, S., J. Schroeder, D. Van Riper, T. Kugler, and S. Ruggles (2021). IPUMS National Historical Geographic Information System: Version 16.0 [Database].

Melnikov, N. (2022). Mobile internet and political polarization.

Müller, K. and C. Schwarz (2020, 10). Fanning the Flames of Hate: Social Media and Hate Crime. *Journal of the European Economic Association 19*(4), 2131–2167.

Müller, K. and C. Schwarz (2023, July). From hashtag to hate crime: Twitter and antiminority sentiment. *American Economic Journal: Applied Economics 15*(3), 270–312.

Muraskin, W. (1974). The moral basis of a backward sociologist: Edward banfield, the italians, and the italian-americans. *American Journal of Sociology 79*(6), 1484–1496.

Nyhan, B., J. Settle, E. Thorson, M. Wojcieszak, P. Barberá, A. Y. Chen, H. Allcott, T. Brown, A. Crespo-Tenorio, D. Dimmery, D. Freelon, M. Gentzkow, S. González-Bailón, A. M. Guess, E. Kennedy, Y. M. Kim, D. Lazer, N. Malhotra, D. Moehler, J. Pan, D. R. Thomas, R. Tromble,

C. V. Rivera, A. Wilkins, B. Xiong, C. K. de Jonge, A. Franco, W. Mason, N. J. Stroud, and J. A. Tucker (2023). Like-minded sources on facebook are prevalent but not polarizing. *Nature 620*(7972), 137–144.

Putnam, R. (2000). *Bowling Alone: The Collapse and Revival of American Community*. A Touchstone book. Simon & Schuster.

Putnam, R., R. Leonardi, and R. Nanetti (1994). *Making Democracy Work: Civic Traditions in Modern Italy*. Book collections on Project MUSE. Princeton University Press.

Rajkumar, K., G. Saint-Jacques, I. Bojinov, E. Brynjolfsson, and S. Aral (2022). A causal test of the strength of weak ties. *Science 377*(6612), 1304–1310.

Sunstein, C. R. (2001). *Echo Chambers: Bush V. Gore, Impeachment, and Beyond*. Princeton University Press.

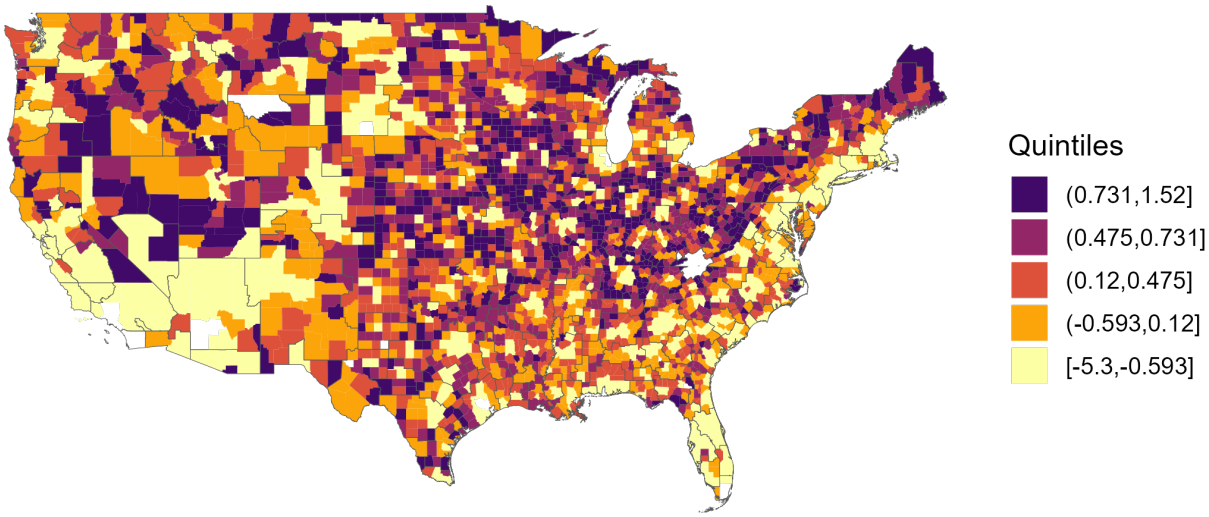
Sunstein, C. R. (2007). *Republic.com 2.0*. Princeton University Press.

Tarbush, B. and A. Teytelboym (2012). Homophily in online social networks. In P. W. Goldberg (Ed.), *Internet and Network Economics*, Berlin, Heidelberg, pp. 512–518. Springer Berlin Heidelberg.

Ugander, J., L. Backstrom, C. Marlow, and J. Kleinberg (2012). Structural diversity in social contagion. *Proceedings of the national academy of sciences 109*(16), 5962–5966.

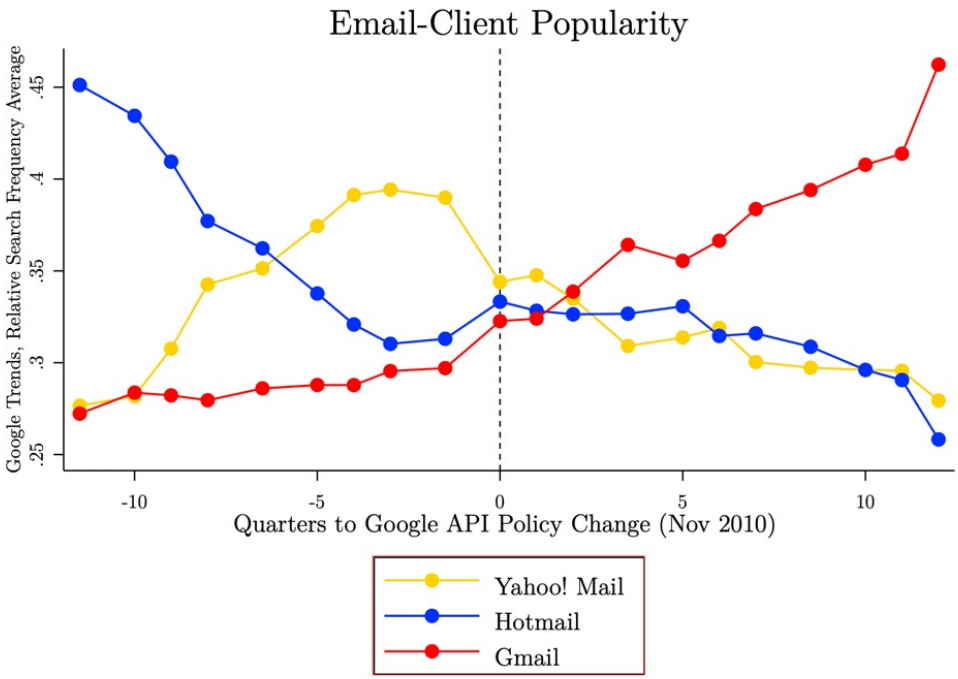
Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). Political effects of the internet and social media. *Annual Review of Economics 12*(1), 415–438.

Figure 1: Online Homophily, 2016



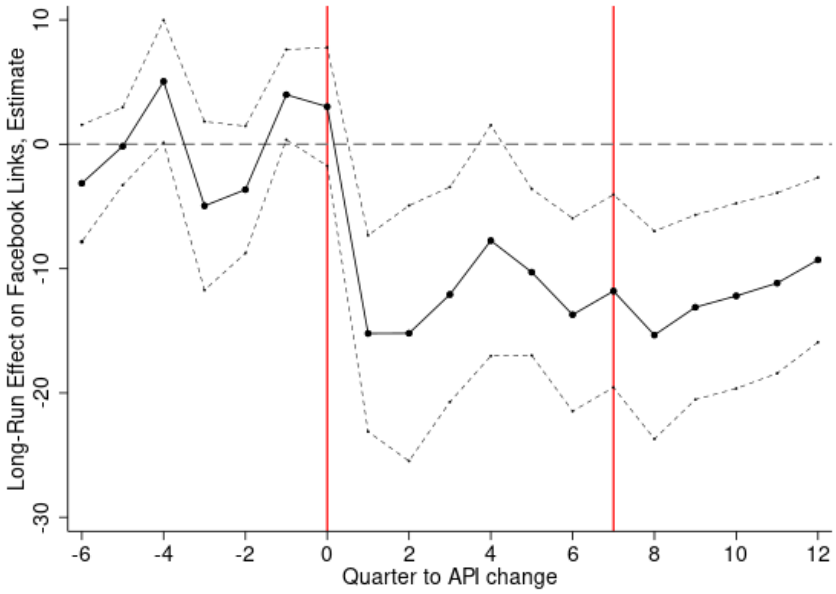
Notes: The map plots the geographic distribution of Online Homophily in 2016 across US counties. We construct Online Homophily in two steps. In the first step, we generate an index of social similarity between US counties. The social similarity index is constructed by taking the inverse of the principal component analysis of twelve socioeconomic differences. In the second step, we use 2016 Facebook links between counties to construct a weighted average of the social similarity each county is exposed to through Facebook. Finally, we standardize the variable.

Figure 2: Relative Popularity of Different Email Clients, 2006-2016



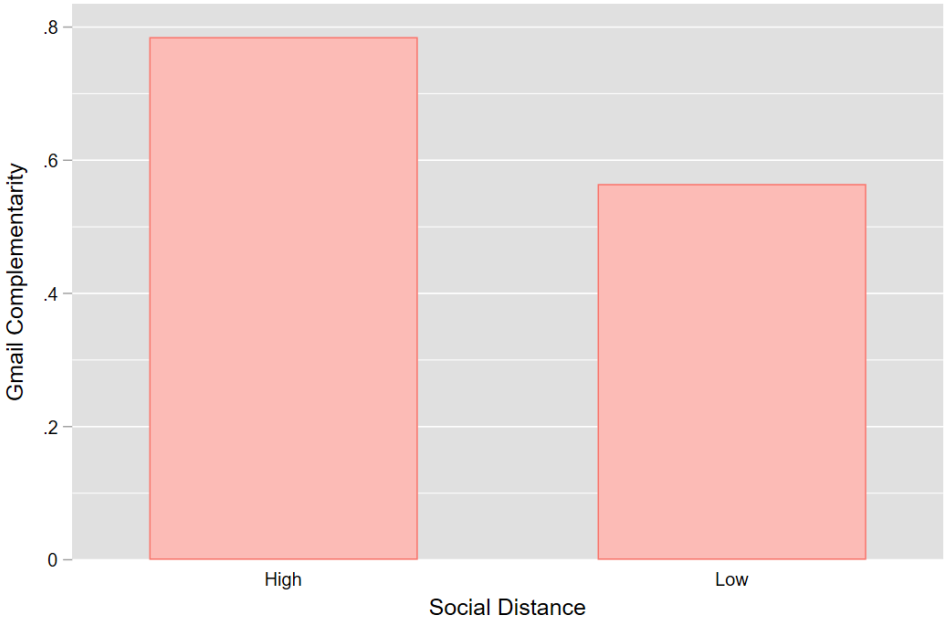
Notes: The figure plots the popularity of the main Email clients in the US, quarterly between 2006 and 2016. The source of the data is Google Trends and popularity is measured as the average search frequency of a given Email client in a DMA. We focus on the three largest Email clients: Yahoo! Mail, Hotmail and Gmail.

Figure 3: Facebook Links by Gmail Complementarity, 2009-2012



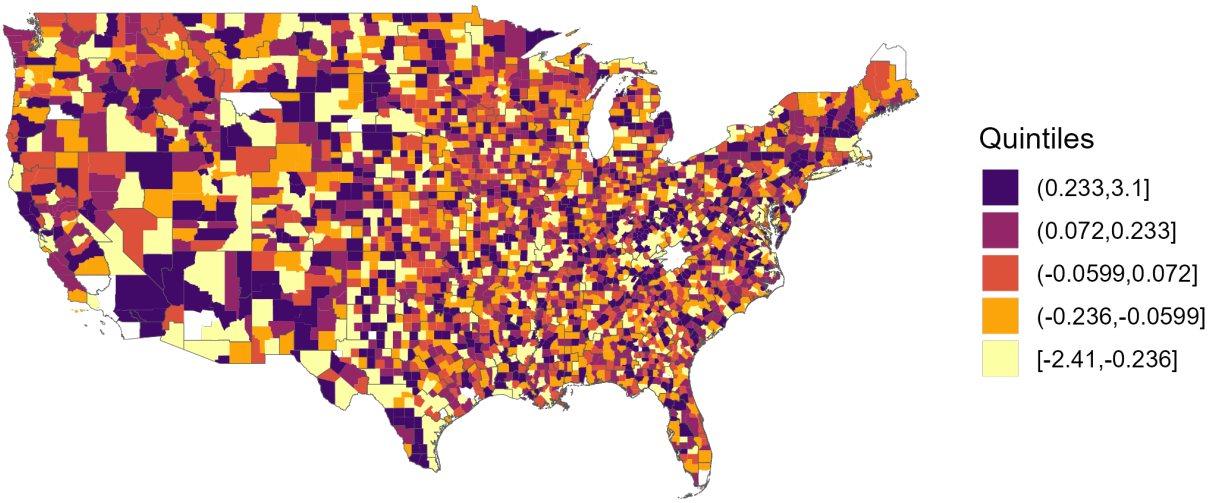
Notes: The graph plots the effect of Gmail complementarity by quarter on 2016 county-pairs Facebook links. We regress and plot the estimates of separate linear models where the outcome variable is the inverse hyperbolic sine of the relative friendship index between county-pairs in 2016. Each estimated coefficient captures the effect of the relative Gmail complementarity in a county-pair six quarters before the API policy change, six quarters during the treatment window, and six quarters after the end of the policy change. The relative Gmail complementarity is computed by taking the complementarity between Email clients across county pairs and computing the difference between Gmail's and the other Email clients' complementarity. Controls include log distance between counties, social distance between counties, and the cumulative Gmail complementarity in the six quarters prior to the API change. The Email client data varies at the DMA-pair level and we cluster standard errors at the DMA-pair level.

Figure 4: Gmail Complementarity by High- and Low-Social Distance in Blount County, Alabama



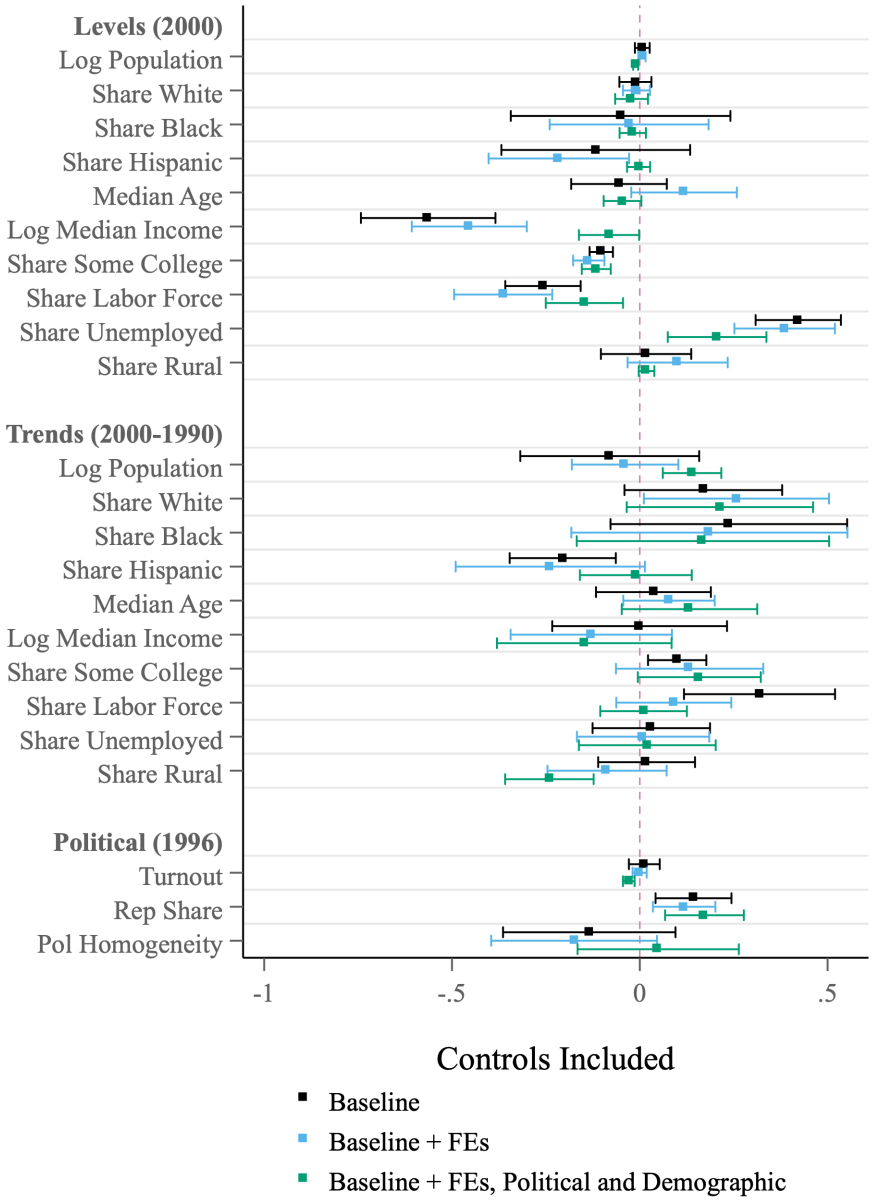
Notes: The figure plots the average Gmail complementarity by social distance for Blount county, Alabama. We denote a county to have high social distance if it belongs to the top tercile of the social distance distribution, whereas we denote the county to have low social distance if it belongs to the bottom tercile of the social distance distribution. We define social distance as the principal component of the difference in absolute value of twelve variables: log population, share White, share Black, share Hispanic, share with at least some college, share unemployed, log median income, share in labor force, share rural, median age in 2010, share of Republican votes in 2004, and their ideology score (county average self-assessed ideology from Gallup polls of 2008-2010 that ranges from 1 very liberal through 5 very conservative). See section 1 for more details.

Figure 5: Gmail Homophily Shock



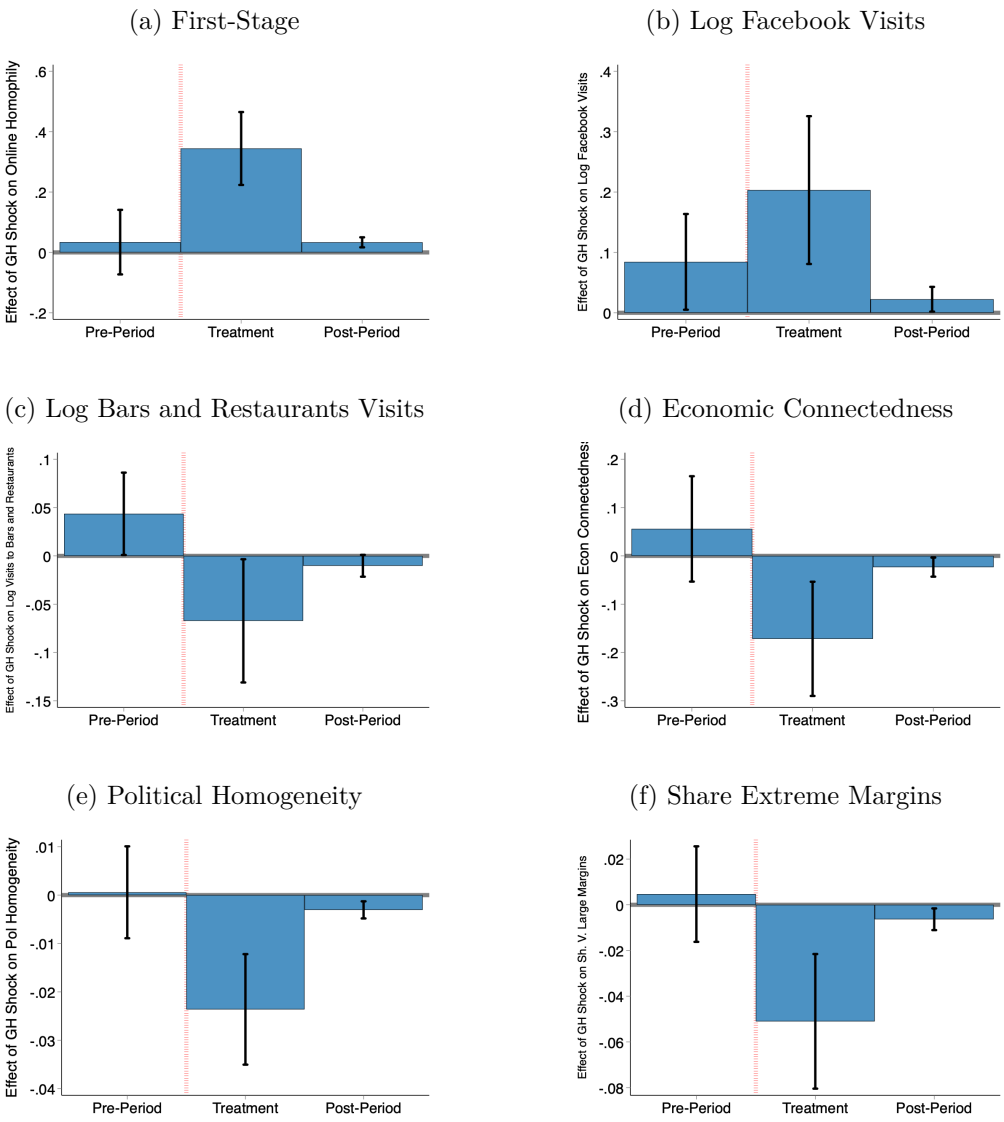
Notes: The map plots the geographic distribution across US counties of the Gmail Homophily Shock residualized on our full set of controls and fixed effects. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. The controls we use to residualize the Gmail Homophily Shock include DMA fixed effects, the pre-period Gmail complementarity using the last six quarters before the API changed, log population, share White, share with at least some college, share unemployed, share Black, share Hispanics, log median income, share in labor force, share rural and median age in 2010; turnout, Republican shares and political homogeneity in 2008; as well as socio-demographic trends defined as the difference for all controls between 2000 and 2010. We cluster standard errors at the state level.

Figure 6: Balance Tests of the Gmail Homophily Shock



Notes: Balance tests of our Gmail Homophily Shock. We plot the estimated coefficients from separate regressions of our baseline model 4 where we test the balancedness of our Gmail Homophily Shock on a host of pre-determined socio-economic and political county-level controls shown on the Y-axis. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. We show the results from three specifications. The first one only employs baseline controls: share White, share with at least some college and share unemployed in 2010; turnout and Republican vote shares as of 2008 and the pre-period Gmail complementarity using the last six quarters before the API changed. In the second specification we add DMA fixed effects. The third specification uses our entire list of controls with the exclusion of trends (as in column 5 of Table 1): share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010; political homogeneity as of 2008. When we perform balance test on log population level in 2000 we omit log population in 2010 from the controls. We cluster standard errors at the state-level.

Figure 7: Gmail Complementarity Has No Effect Outside Treatment Window



Notes: The figure plots the effect of the cumulative Gmail complementarity on our outcome variables in three different windows: the pre-period, the treatment period, and the post-period. We build the cumulative Gmail complementarity in the pre-period using the six quarters before the API change. The cumulative Gmail complementarity in the treatment window is our Gmail Homophily Shock which is constructed using the six quarters after API change. Similarly, we define the post-period window using the six quarters following the end of the Google-Facebook incident. We plot bars and confidence intervals from our most saturated specification which includes the following controls: log population, share White, share with at least some college and share unemployed, share Black, share Hispanic, log median income, share in labor force, share rural and median age as of 2010; turnout, political homogeneity and Republican shares as of 2008; all trends defined as the difference for all socio-demographic controls between 2000 and 2010 as well as DMA fixed effects. We include the pre-period Gmail complementarity as a control when we regress our outcomes on the treatment and post-period Gmail complementarity. We cluster standard errors at the state level. Figure A.8 displays results using a sparser specification including only DMA FEs and baseline controls (same as in column 3 of Table 1).

Figure 8: The Impact of Homophily Shock on Visits by Venues of Interaction

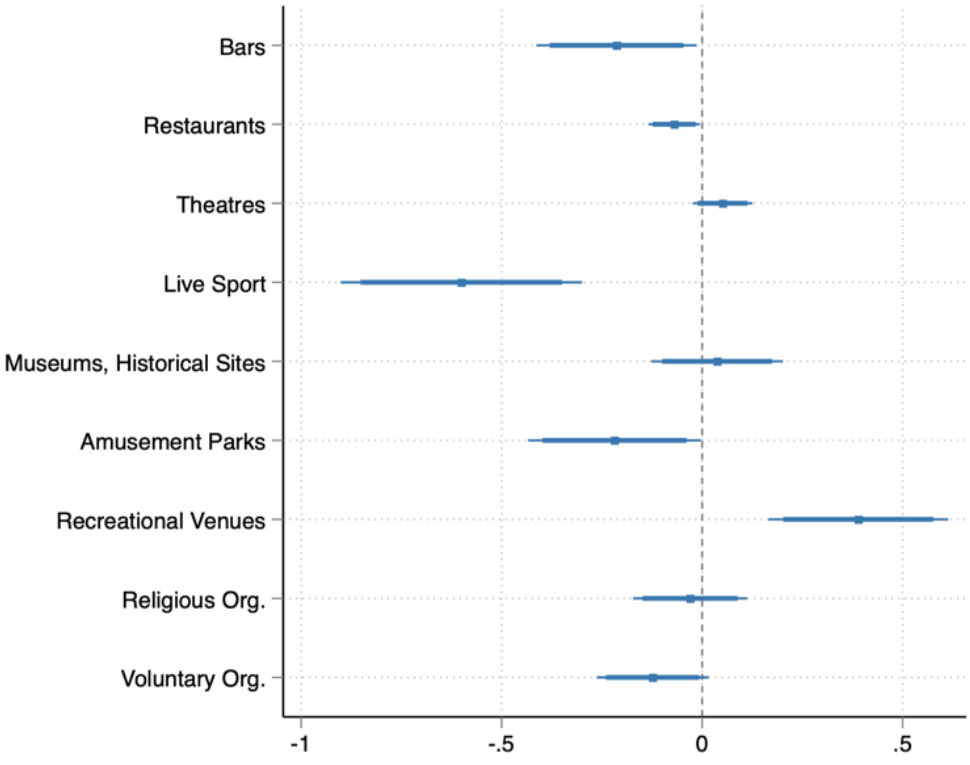
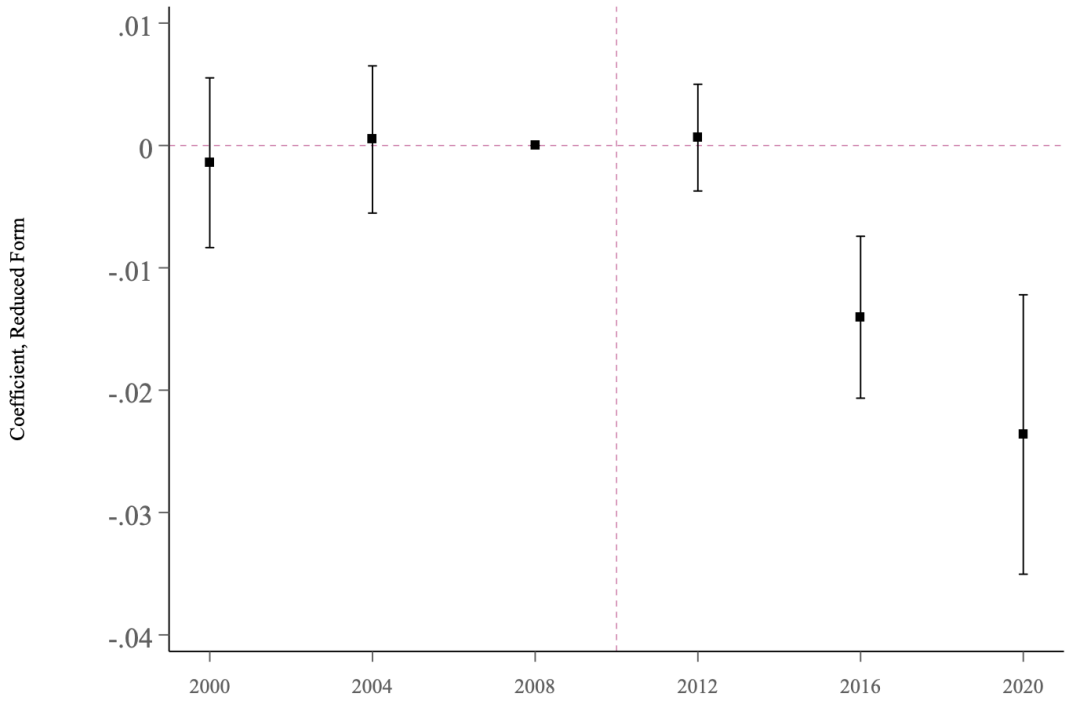


Figure 9: Gmail Homophily Shock and Political Homogeneity



Notes: Event study analysis of the impact of Gmail Homophily Shock on political homogeneity. We plot the estimated coefficients associated with the effect of one standard deviation increase in the Gmail Homophily Shock on political homogeneity every four years between 2000 and 2020. Political homogeneity is computed as one minus the Herfindahl index applied to the Republican and Democrat vote shares, 2008 is our reference year. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Controls include the pre-period Gmail complementarity using the last six quarters before the API changed; log population, share White, share with at least some college and share unemployed, share Black, share Hispanic, log median income, share in labor force, share rural and median age as of 2010; turnout, political homogeneity and Republican shares as of 2008; all trends defined as the difference for all socio-demographic controls between 2000 and 2010 as well as DMA fixed effects. We cluster standard errors at the state level.

Table 1: Long-run effect of Gmail Homophily Shock on Online Homophily

Dep. Variable:	Online Homophily (sd)						
	2016						2020
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Gmail Homophily Shock	0.625***	0.277***	0.419***	0.411***	0.357***	0.344***	0.305***
	(0.148)	(0.084)	(0.063)	(0.069)	(0.058)	(0.060)	(0.055)
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*	*Yes*
KP F-Stat	17.732	10.914	43.657	35.280	37.285	32.856	30.920
Adj R2	0.572	0.719	0.810	0.819	0.827	0.834	0.844
Mean of Dep. Var.	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	3042	3042	3042	3042	3042	3042	3042

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. In columns 1 to 6, we use the share of 2016 Facebook connections as weights, whereas in column 7 we use the 2020 Facebook connections. Please, refer to section 1.3 for further details. The Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Baseline Controls include basic demographic and political county characteristics: share white, share with at least some college and share unemployed in 2010; turnout and Republican vote shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table 2: Homophily Shock and Social Media Visits

Dep. Variable:	Log Facebook Visits				Log Other SM Visits			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GC Homophily Shock	0.262***	0.226***	0.205***	0.203***	-0.078**	-0.081*	-0.099**	-0.103**
	(0.059)	(0.055)	(0.056)	(0.061)	(0.035)	(0.044)	(0.045)	(0.046)
Mean of Dep. Var.	4.853	4.853	4.853	4.853	2.391	2.391	2.391	2.391
Adj R2	0.866	0.866	0.867	0.867	0.841	0.842	0.842	0.842
Observations	2872	2872	2872	2872	2872	2872	2872	2872
Instrumental Variable Estimates								
Online Homophily	0.690***	0.629***	0.658***	0.661***	-0.206**	-0.225*	-0.319**	-0.334**
	(0.127)	(0.145)	(0.170)	(0.182)	(0.102)	(0.129)	(0.154)	(0.162)
F-stat	44.568	35.157	31.467	28.251	44.568	35.157	31.467	28.251
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*Yes*	*Yes*	*Yes*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*Yes*	*Yes*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*Yes*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables measure the log number of visits to Facebook (columns 1-4) and to other social media (columns 5-8). Other social media include Instagram and Twitter. The data on web searches comes from ComScore and refers to the first trimester of 2016. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications, we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. All specifications control for the total number of web visits and its squared term. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table 3: Homophily of Online Connections and Total Time on Social Media

Dep. Variable:	Log Any SM Visits				Log Any SM Minutes			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GC Homophily Shock	0.237***	0.205***	0.184***	0.185***	0.298***	0.277***	0.238***	0.242***
	(0.056)	(0.051)	(0.053)	(0.057)	(0.076)	(0.074)	(0.075)	(0.083)
Mean of Dep. Var.	4.944	4.944	4.944	4.944	7.389	7.389	7.389	7.389
Adj R2	0.879	0.880	0.880	0.880	0.799	0.800	0.800	0.800
Observations	2872	2872	2872	2872	2872	2872	2872	2872
Instrumental Variable Estimates								
Online Homophily	0.625***	0.570***	0.591***	0.602***	0.754***	0.736***	0.731***	0.754***
	(0.118)	(0.133)	(0.156)	(0.165)	(0.183)	(0.212)	(0.247)	(0.276)
F-stat	44.568	35.157	31.467	28.251	45.929	36.322	32.375	28.561
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*Yes*	*Yes*	*Yes*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*Yes*	*Yes*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*Yes*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables measure the log total number of social media visits (columns 1-4) and log total minutes spent on social media (columns 5-8). Social media include Facebook, Instagram and Twitter. The data on web searches comes from ComScore and refers to the first trimester of 2016. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications, we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. All specifications control for the total number of web visits and its squared term. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table 4: Homophily Shock and Bars and Restaurants Visits, 2019

Dep. Variable:	Log Visits to Bars and Restaurants					
	(1)	(2)	(3)	(4)	(5)	(6)
GC Homophily Shock	-0.053**	-0.075*	-0.054	-0.076***	-0.076**	-0.067**
	(0.022)	(0.037)	(0.033)	(0.027)	(0.030)	(0.032)
Mean of Dep. Var.	9.318	9.318	9.318	9.318	9.318	9.318
Adj R2	0.947	0.948	0.953	0.954	0.954	0.954
Observations	36564	36564	36564	36564	36564	36564
Instrumental Variable Estimates						
Online Homophily	-0.122**	-0.323*	-0.159	-0.248**	-0.283**	-0.251*
	(0.061)	(0.194)	(0.096)	(0.097)	(0.124)	(0.129)
F-stat	11.993	10.892	55.892	42.148	38.595	31.939
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable measures the log number of visits to bars and restaurants (NAICS codes 7224 and 7225). The source of the data is Safegraph, it covers all of 2019 and varies at the county-by-month level. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications, we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. All specifications control for the total number of visits and its squared term. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table 5: Homophily Shock and Economic Connectedness

Dep. Variable:	Econ Connectedness					
	(1)	(2)	(3)	(4)	(5)	(6)
Gmail Homophily Shock	-0.516***	-0.312***	-0.308***	-0.197***	-0.184***	-0.172***
	(0.098)	(0.056)	(0.058)	(0.058)	(0.062)	(0.059)
Mean of Dep. Var.	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
Adj R2	0.396	0.705	0.816	0.863	0.863	0.872
Observations	2943	2943	2943	2943	2943	2943
Instrumental Variable Estimates						
Online Homophily	-0.881***	-1.158***	-0.749***	-0.485**	-0.535**	-0.511**
	(0.333)	(0.264)	(0.202)	(0.188)	(0.230)	(0.230)
F-stat	17.199	11.266	48.436	38.553	36.399	30.942
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable measures the economic connectedness across income strata. The source of the data is Chetty et al. (2022a). Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications, we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table 6: Online Homophily and Political Homogeneity, 2020

Dep. Variable:	Political Homogeneity, 2020					
	(1)	(2)	(3)	(4)	(5)	(6)
Gmail Homophily Shock	-0.000	-0.038***	-0.036***	-0.049***	-0.024***	-0.024***
	(0.012)	(0.010)	(0.010)	(0.011)	(0.006)	(0.006)
Mean of Dep. Var.	0.605	0.605	0.605	0.605	0.605	0.605
Adj R2	0.253	0.529	0.620	0.646	0.868	0.876
Observations	3042	3042	3042	3042	3042	3042
Instrumental Variable Estimates						
Online Homophily	-0.000	-0.136**	-0.085***	-0.120***	-0.068***	-0.069***
	(0.020)	(0.059)	(0.022)	(0.021)	(0.019)	(0.017)
F-stat	17.732	10.914	43.657	35.280	37.285	32.856
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable measures political homogeneity in 2020. Political homogeneity is computed as one minus the Herfindahl index applied to the Republican and Democrat vote shares. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications, we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table 7: Online Homophily and Dispersion of Electoral Results

Dep. Variable:	Sd Trump Share		Iqr Trump Share		Share Extreme Margins		Share V. Extreme Margins	
Gmail Homophily Shock	0.014***	0.006	0.027***	0.015*	-0.041	-0.026	-0.064**	-0.051***
	(0.004)	(0.005)	(0.008)	(0.008)	(0.028)	(0.018)	(0.030)	(0.015)
Adj R2	0.646	0.707	0.561	0.638	0.620	0.715	0.612	0.787
Mean of Dep. Var.	0.114	0.114	0.150	0.150	0.626	0.626	0.357	0.357
Observations	2729	2729	2729	2729	2729	2729	2729	2729
	Instrumental Variable Estimates							
Online Homophily	0.035***	0.017	0.069***	0.043*	-0.106	-0.076	-0.165**	-0.148***
	(0.012)	(0.016)	(0.021)	(0.026)	(0.072)	(0.051)	(0.072)	(0.039)
F-stat	37.705	28.428	37.705	28.428	37.705	28.428	37.705	28.428
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*
Political Controls	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*
Demographic Trends	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables are constructed using precinct-level electoral outcomes for 2016 from Kaplan et al. (2022): the standard deviation in the Trump vote share (cols 1-2), the interquartile range for the Trump vote share (cols 3-4), the share of precincts with vote margins of at least 40 points (cols 5-6) and the share of precincts with vote margins of at least 60 points (cols 7-8). Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications, we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in the labor force, share rural, and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table 8: Share Facebook Links Outside County, 2016

Dep. Variable:	Share Out Connections, 2016					
	(1)	(2)	(3)	(4)	(5)	(6)
GC Homophily Shock	-0.022*	0.017*	-0.007	-0.002	-0.004	-0.003
	(0.011)	(0.010)	(0.007)	(0.008)	(0.008)	(0.007)
Mean of Dep. Var.	0.561	0.561	0.561	0.561	0.561	0.561
Adj R2	0.543	0.644	0.786	0.794	0.794	0.801
Observations	3042	3042	3042	3042	3042	3042
Instrumental Variable Estimates						
Online Homophily	-0.036***	0.062	-0.018	-0.005	-0.011	-0.009
	(0.012)	(0.051)	(0.016)	(0.019)	(0.022)	(0.021)
F-stat	17.732	10.914	43.657	35.280	37.285	32.856
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the share of Facebook links outside the county in 2016. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table 9: The Impact of the Gmail Homophily Shock by Share Out Links, 2016

	First-Stage	Log Facebook Visits	Log Bars and Rest. Visits	Econ Connect	Pol Homogeneity	Extreme Margins
	(1)	(2)	(3)	(4)	(5)	(6)
GC Homophily Shock	0.234***	0.084	-0.049	-0.123*	-0.019***	-0.037***
	(0.061)	(0.060)	(0.036)	(0.064)	(0.006)	(0.013)
– × Share Out Links	0.112***	0.121***	-0.022	-0.049***	-0.005***	-0.015***
	(0.012)	(0.025)	(0.014)	(0.014)	(0.001)	(0.004)
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Political Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Trends	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Adj R2	0.902	0.869	0.954	0.876	0.880	0.779
Mean of Dep. Var.	0.000	0.000	-0.000	-0.000	0.000	0.000
Observations	3042	2872	36504	2943	3042	2738

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The Table presents the heterogeneous effect of the Gmail Homophily Shock by the share of Facebook links outside the county. The dependent variables are the six main outcomes of the paper: first-stage in column 1, Log Facebook visits in column 2, Log Bar visits in column 3, Economic Connectedness in column 4, political homogeneity in column 5 and the share of precinct within a county with electoral margins larger than 60 points in 2016 in column 6. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

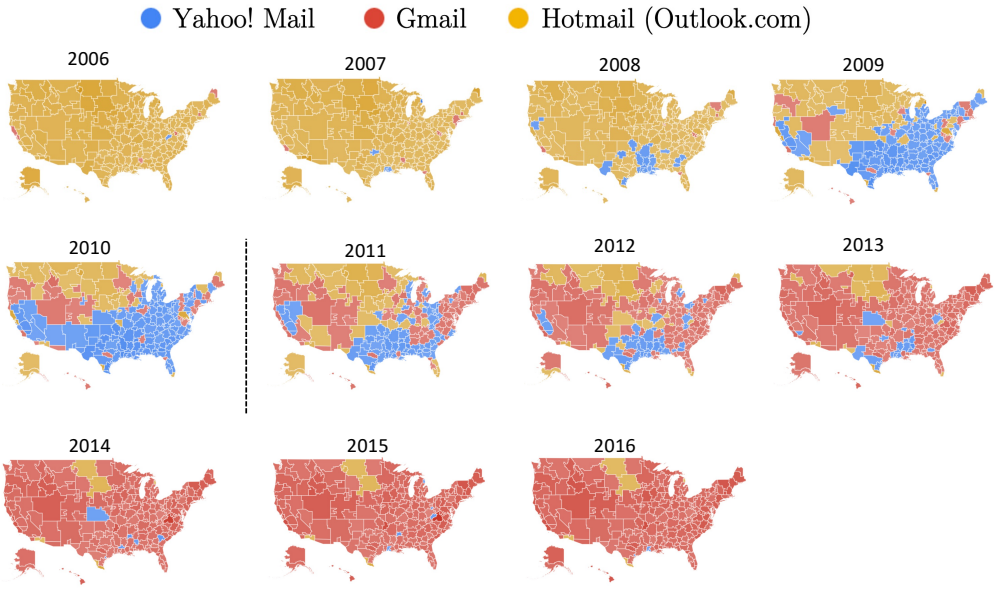
Table 10: The Impact of the Gmail Homophily Shock by Share Urban

	First-Stage	Log Facebook Visits	Log Bars and Rest. Visits	Econ Connect	Pol Homogeneity	Extreme Margins
	(1)	(2)	(3)	(4)	(5)	(6)
GC Homophily Shock	0.193***	0.101	-0.084**	-0.127**	-0.015***	-0.031*
	(0.065)	(0.061)	(0.039)	(0.062)	(0.005)	(0.016)
– × Share Urban	0.157***	0.102***	0.019	-0.041***	-0.009***	-0.020***
	(0.015)	(0.024)	(0.017)	(0.014)	(0.001)	(0.005)
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Political Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Trends	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Adj R2	0.849	0.868	0.953	0.873	0.882	0.780
Mean of Dep. Var.	0.000	0.000	-0.000	-0.000	0.000	0.000
Observations	3042	2872	36504	2943	3042	2738

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The Table presents the heterogeneous effect of the Gmail Homophily Shock by the share of the county population living in Urban areas. The dependent variables are the six main outcomes of the paper: first-stage in column 1, Log Facebook visits in column 2, Log Bar visits in column 3, Economic Connectedness in column 4, political homogeneity in column 5 and the share of precinct within a county with electoral margins larger than 60 points in 2016 in column 6. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

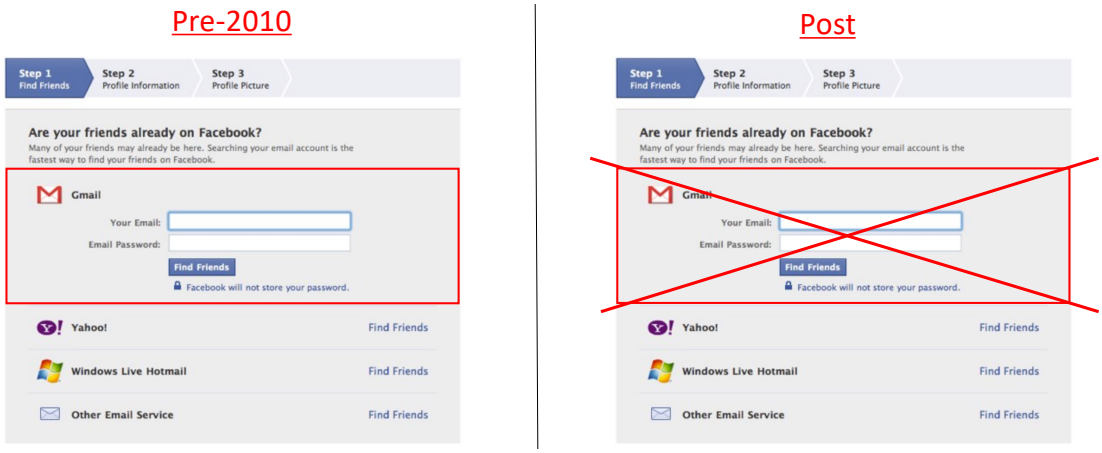
A Online Appendix (Not for publication)

Figure A.1: Geographic Distribution of Relative Popularity Across US DMAs



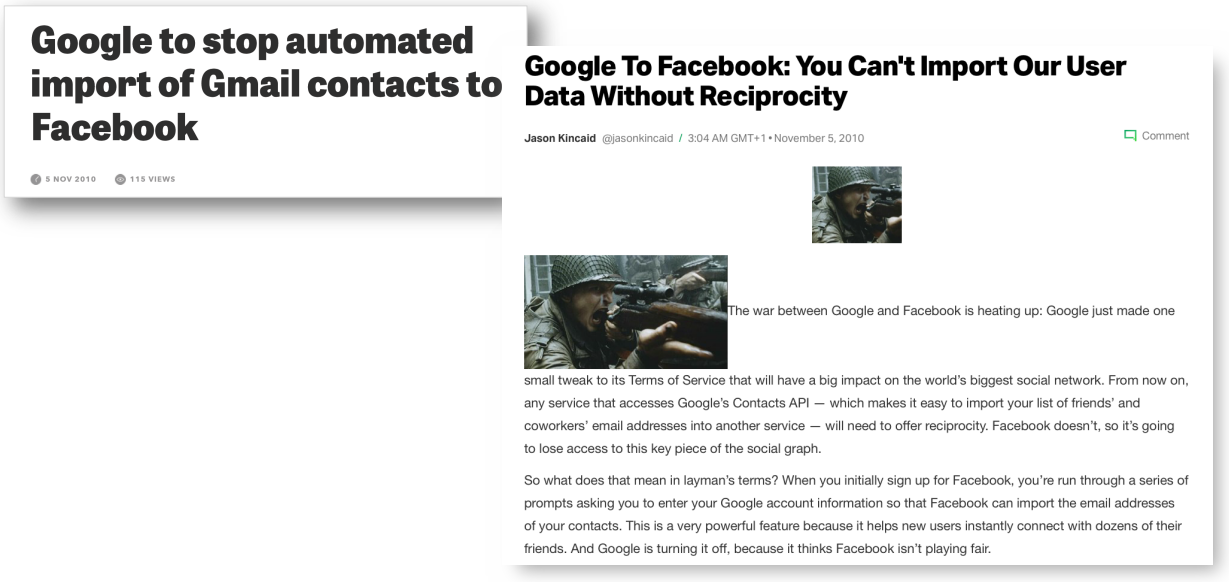
Notes: The map plots the most popular email client between Yahoo!, Gmail and Hotmail at the DMA-level over time. The source of the data is Google Trends and popularity is measured as the average search frequency of a given Email client in a DMA. We focus on the three largest Email clients: Yahoo!, Gmail and Hotmail.

Figure A.2: Exterior Look of Join-Facebook Window in 2009-2012



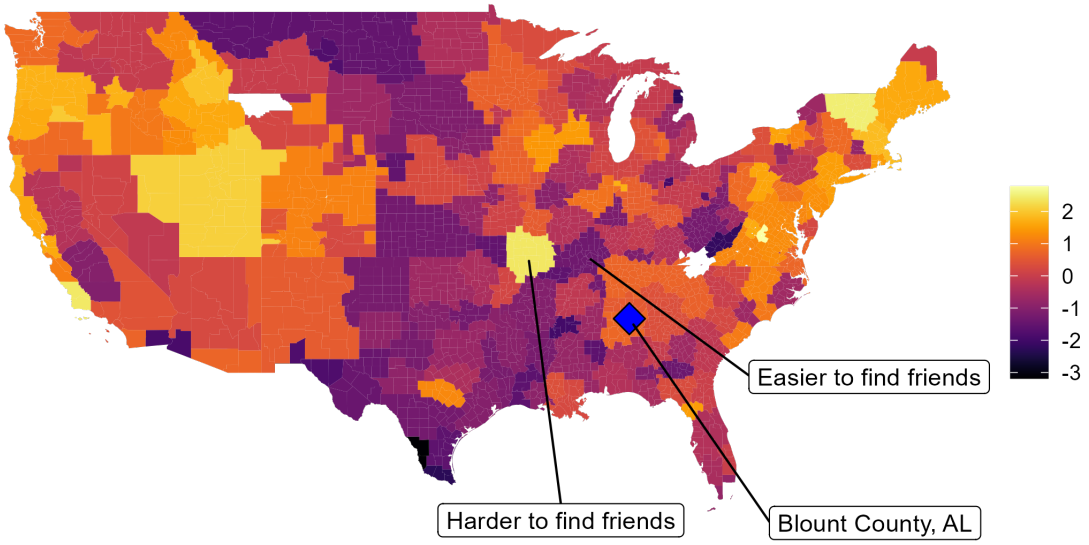
Notes: The figure depicts the typical look of the join-Facebook window that a user would face when joining Facebook before and after November 2010.

Figure A.3: Google-Facebook Conflict Headlines



Notes: The figure shows two headlines contemporary to the incident between Google and Facebook giving details on the source of the conflict. The source of the headlines is the Tech blog TechCrunch.com (https://techcrunch.com/2010/11/04/facebook-google-contacts/?guccounter=1)

Figure A.4: Gmail Complementarity by County for Blount county, AL



Notes: The map plots the geographic distribution of the cumulative Gmail complementarity in the six quarters post-API change between Blount county, AL, and the rest of the counties in the US. The data source of email client usage is Google Trends and comes at the DMA level. Lighter (darker) color indicates higher (lower) complementarity hence higher (lower) difficulty in finding friends on Facebook after the API change.

Figure A.5: Gmail Complementarity by High- and Low-Social Distance by County in Alabama



Notes: The figure plots the average Gmail complementarity by social distance for all the counties in Alabama. We denote a county to have high social distance if it belongs to the top tercile of the social distance distribution, whereas we denote the county to have low distance if it belongs to the bottom tercile of the social distance distribution. We define social distance as the principal component of the absolute difference of twelve variables: log population, share White, share Black, share Hispanic, share with at least some college, share unemployed, log median income, share in labor force, share rural, median age in 2010, share of Republican votes in 2004, and their ideology score (county average self-assessed ideology from Gallup polls of 2008-2010 that ranges from 1 very liberal through 5 very conservative). See section 1 for more details.

Figure A.6: Gmail Homophily Shock by County in Alabama



Notes: The figure plots the Gmail Homophily Shock for each county in Alabama. The Gmail Homophily Shock is calculated as the difference in the average Gmail complementarity among high- and low-social distance counties. The average Gmail complementarity by social distance is plotted in Figure A.5 for all the counties in Alabama.

Figure A.7: Political Homogeneity and Vote Shares

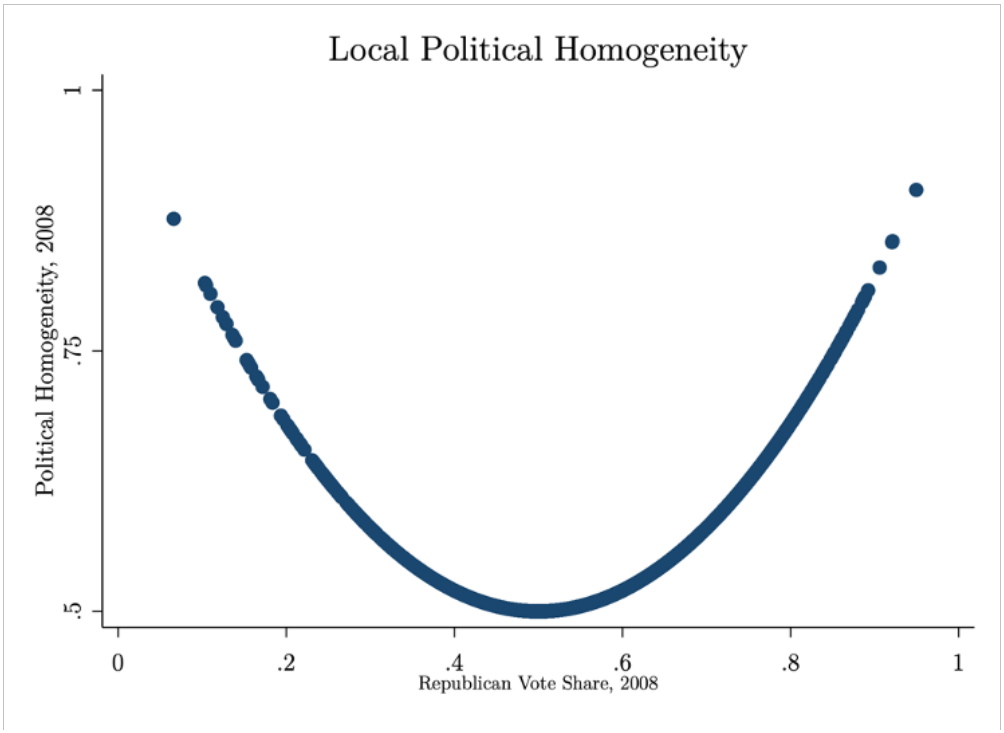
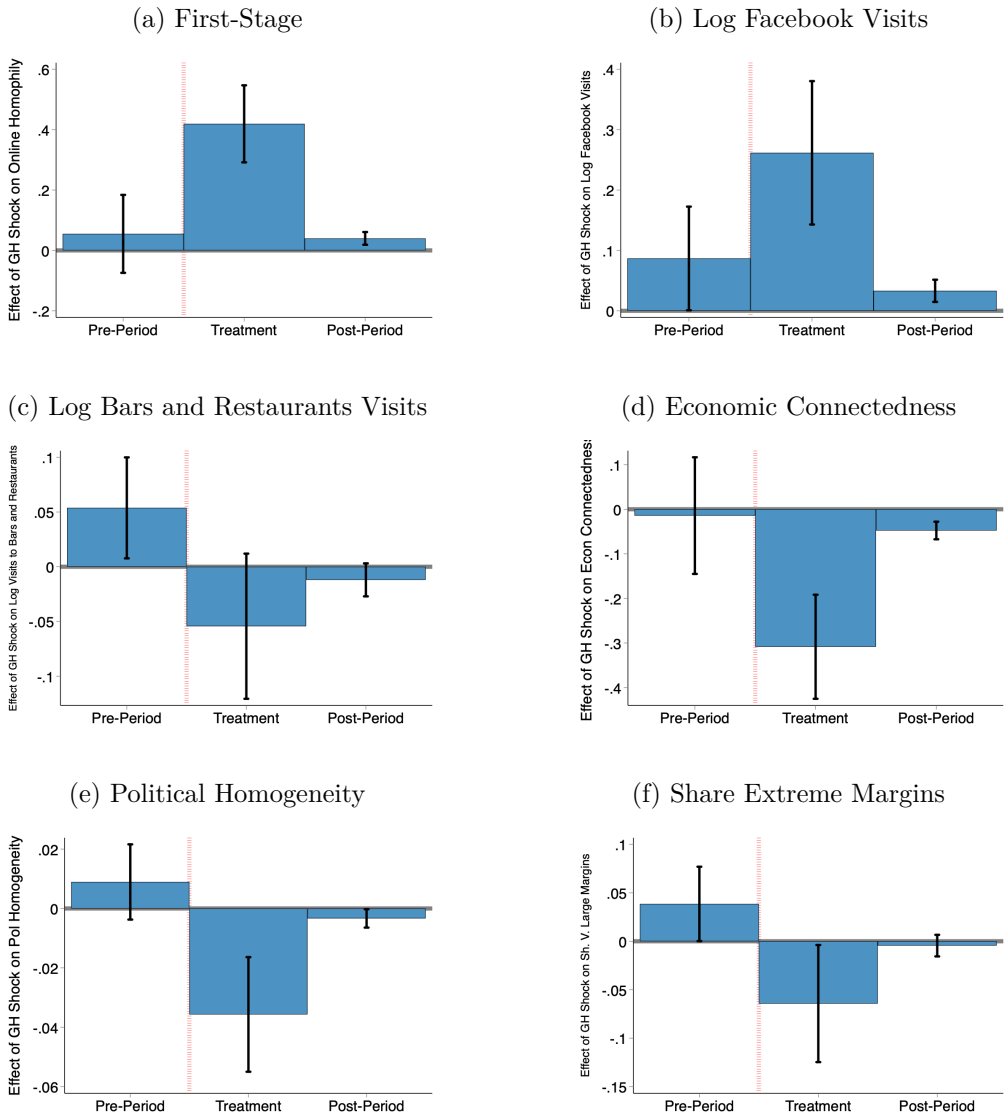


Figure A.8: Gmail Complementarity Has No Effect Outside Treatment Window, Sparse Specification



Notes: The figure plots the effect of the cumulative Gmail complementarity on our outcome variables in three different windows: the pre-period, the treatment period and the post-period. We build the cumulative Gmail complementarity in the pre-period using the six quarters prior to the API change. The cumulative Gmail complementarity in the treatment window is our Gmail Homophily Shock which is constructed using the six quarters after API change. Similarly, we define the post-period window using the six quarters following the end of Google-Facebook incident. We plot bars and confidence intervals from a specification which includes baseline controls and DMA fixed effects. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. We include the pre-period Gmail complementarity as control when we regress our outcomes on the treatment and post-period Gmail complementarity. We cluster standard errors at the state level. Figure 7 displays results using our most saturated specification (same as in column 6 of Table 1)

Table A.1: Homophily Shock and Social Media Visits, OLS

Dep. Variable:	Log Facebook Visits				Log Other SM Visits			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Online Homophily	0.240***	0.262***	0.249***	0.260***	-0.053	-0.041	-0.062	-0.058
	(0.039)	(0.044)	(0.047)	(0.050)	(0.032)	(0.036)	(0.040)	(0.041)
Mean of Dep. Var.	4.853	4.853	4.853	4.853	2.391	2.391	2.391	2.391
Adj R2	0.867	0.867	0.868	0.868	0.841	0.842	0.842	0.842
Observations	2872	2872	2872	2872	2872	2872	2872	2872
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*Yes*	*Yes*	*Yes*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*Yes*	*Yes*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*Yes*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables are log visits to Facebook (columns 1-4) and to other social media (columns 5-8). Other social media include Instagram and Twitter. The data on web searches comes from ComScore and refers to the first trimester of 2016. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. All specifications control for the total number of web visits and its squared term. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Robust standard errors clustered by state in parentheses.

Table A.2: Homophily Shock and Time Spent on Social Media, OLS

Dep. Variable:	Log Facebook Minutes				Log Other SM Minutes			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Online Homophily	0.291***	0.330***	0.302***	0.317***	-0.216***	-0.203***	-0.242***	-0.244***
	(0.058)	(0.067)	(0.072)	(0.071)	(0.058)	(0.061)	(0.063)	(0.063)
Mean of Dep. Var.	7.333	7.333	7.333	7.333	3.046	3.046	3.046	3.046
Adj R2	0.788	0.789	0.789	0.789	0.763	0.764	0.764	0.764
Observations	2872	2872	2872	2872	2872	2872	2872	2872
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*Yes*	*Yes*	*Yes*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*Yes*	*Yes*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*Yes*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables are log minutes spent on Facebook (columns 1-4) and ib other social media (columns 5-8). Other social media include Instagram and Twitter. The data on web searches comes from ComScore and refers to the first trimester of 2016. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. All specifications control for the total number of web visits and its squared term. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Robust standard errors clustered by state in parentheses.

Table A.3: Homophily Shock and Time Spent on Social Media

Dep. Variable:	Log Facebook Minutes				Log Other SM Minutes			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GC Homophily Shock	0.317***	0.293***	0.254***	0.255***	-0.189***	-0.182**	-0.209**	-0.224**
	(0.083)	(0.080)	(0.081)	(0.088)	(0.065)	(0.079)	(0.084)	(0.086)
Mean of Dep. Var.	7.333	7.333	7.333	7.333	3.046	3.046	3.046	3.046
Adj R2	0.787	0.788	0.788	0.788	0.762	0.763	0.763	0.763
Observations	2872	2872	2872	2872	2872	2872	2872	2872
Instrumental Variable Estimates								
Online Homophily	0.804***	0.778***	0.781***	0.793***	-0.478***	-0.483**	-0.643**	-0.699***
	(0.201)	(0.232)	(0.267)	(0.295)	(0.155)	(0.206)	(0.246)	(0.253)
F-stat	45.929	36.322	32.375	28.561	45.929	36.322	32.375	28.561
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*Yes*	*Yes*	*Yes*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*Yes*	*Yes*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*Yes*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables measure the log minutes spent on Facebook (columns 1-4) and on other social media (columns 5-8). Other social media include Instagram and Twitter. The data on web searches comes from ComScore and refers to the first trimester of 2016. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. All specifications control for the total time in minutes and its squared term. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table A.4: Homophily Shock and Restaurant Visits, 2019, OLS

Dep. Variable:	Log Restaurant Visits					
	(1)	(2)	(3)	(4)	(5)	(6)
Online Homophily	0.024	0.023	0.039	-0.002	0.000	-0.005
	(0.019)	(0.018)	(0.030)	(0.030)	(0.033)	(0.034)
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*
Mean of Dep. Var.	9.311	9.311	9.311	9.311	9.311	9.311
Adj R2	0.947	0.948	0.953	0.954	0.954	0.954
Observations	36564	36564	36564	36564	36564	36564

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable measures the log number of visits to restaurants and other eating eating places (NAICS code 7225). The source of the data is Safegraph, it covers all of 2019 and varies at the county by month level. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. All specifications control for the total number of visits and its squared term. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Robust standard errors clustered by state in parentheses.

Table A.5: Homophily Shock and Total Visits, 2019

Dep. Variable:	Log Total Visits					
	(1)	(2)	(3)	(4)	(5)	(6)
Gmail Homophily Shock	-0.005	0.036	-0.006	0.032	0.030	0.048
	(0.045)	(0.045)	(0.044)	(0.044)	(0.045)	(0.041)
Mean of Dep. Var.	11.091	11.091	11.091	11.091	11.091	11.091
Adj R2	0.924	0.936	0.957	0.963	0.963	0.964
Observations	36564	36564	36564	36564	36564	36564
Instrumental Variable Estimates						
Online Homophily	-0.008	0.130	-0.014	0.078	0.085	0.141
	(0.071)	(0.140)	(0.107)	(0.101)	(0.120)	(0.110)
F-stat	17.532	10.861	46.545	37.687	39.842	35.224
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable measures the log number of visits to any establishment. The source of the data is Safegraph, it covers all of 2019 and varies at the county by month level. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table A.6: Homophily Shock and Economic Connectedness, OLS

Dep. Variable:	Econ Connectedness					
	(1)	(2)	(3)	(4)	(5)	(6)
Online Homophily	-0.196***	-0.263***	-0.263***	-0.194***	-0.190***	-0.165***
	(0.069)	(0.046)	(0.032)	(0.027)	(0.027)	(0.026)
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*
Mean of Dep. Var.	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
Adj R2	0.365	0.711	0.822	0.866	0.866	0.874
Observations	2943	2943	2943	2943	2943	2943

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable measures the economic connectedness across income strata. The source of the data is Chetty et al. (2022a). Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Robust standard errors clustered by state in parentheses.

Table A.7: Online Homophily and Political Homogeneity, 2020, OLS

Dep. Variable:	Political Homogeneity, 2020					
	(1)	(2)	(3)	(4)	(5)	(6)
Online Homophily	-0.005	-0.029***	-0.032***	-0.033***	-0.010***	-0.012***
	(0.008)	(0.007)	(0.007)	(0.006)	(0.003)	(0.003)
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*
Mean of Dep. Var.	0.605	0.605	0.605	0.605	0.605	0.605
Adj R2	0.254	0.535	0.634	0.652	0.865	0.875
Observations	3042	3042	3042	3042	3042	3042

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable measures political homogeneity in 2020. Political homogeneity is computed as one minus the Herfindahl index applied to the Republican and Democrat vote shares. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Robust standard errors clustered by state in parentheses.

Table A.8: Online Homophily and Dispersion of Electoral Results, OLS

Dep. Variable:	Sd Trump Share		Iqr Trump Share		Share Extreme Margins		Share V. Extreme Margins	
Online Homophily	0.010***	0.002	0.016**	0.001	-0.071***	-0.042***	-0.088***	-0.042***
	(0.003)	(0.002)	(0.007)	(0.005)	(0.023)	(0.014)	(0.021)	(0.011)
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*
Political Controls	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*
Demographic Trends	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*
Adj R2	0.647	0.707	0.560	0.637	0.629	0.717	0.624	0.788
Mean of Dep. Var.	0.114	0.114	0.150	0.150	0.626	0.626	0.357	0.357
Observations	2729	2729	2729	2729	2729	2729	2729	2729

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables are constructed using precinct-level electoral outcomes for 2016 from Kaplan et al. (2022): the standard deviation in the Trump vote share (cols 1-2), the interquartile range for the Trump vote share (cols 3-4), the share of precincts with vote margins of at least 40 points (cols 5-6) and the share of precincts with vote margins of at least 60 points (cols 7-8). Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Robust standard errors clustered by state in parentheses.

Table A.9: Homophily Shock and Social Media Visits (IHS)

Dep. Variable:	IHS Facebook Visits				IHS Other SM Visits			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GC Homophily Shock	0.281***	0.244***	0.221***	0.218***	-0.057	-0.052	-0.074	-0.076
	(0.064)	(0.060)	(0.062)	(0.067)	(0.040)	(0.051)	(0.052)	(0.055)
Mean of Dep. Var.	5.443	5.443	5.443	5.443	2.803	2.803	2.803	2.803
Adj R2	0.860	0.861	0.861	0.862	0.834	0.836	0.836	0.836
Observations	2872	2872	2872	2872	2872	2872	2872	2872
Instrumental Variable Estimates								
Online Homophily	0.730***	0.668***	0.698***	0.697***	-0.148	-0.142	-0.235	-0.242
	(0.137)	(0.155)	(0.183)	(0.196)	(0.112)	(0.144)	(0.175)	(0.186)
F-stat	45.242	35.665	31.973	28.722	45.242	35.665	31.973	28.722
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*Yes*	*Yes*	*Yes*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*Yes*	*Yes*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*Yes*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables are the inverse hyperbolic sine (IHS) of the visits to Facebook (columns 1-4) and to other social media (columns 5-8). Other social media include Instagram and Twitter. The data on web searches comes from ComScore and refers to the first trimester of 2016. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. All specifications control for the total number of web visits and its squared term. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table A.10: Homophily Shock and Time Spent on Social Media (IHS)

Dep. Variable:	IHS Facebook Minutes				IHS Other SM Minutes			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GC Homophily Shock	0.321***	0.297***	0.255***	0.256**	-0.167**	-0.151*	-0.184**	-0.197**
	(0.088)	(0.087)	(0.088)	(0.096)	(0.064)	(0.083)	(0.087)	(0.090)
Mean of Dep. Var.	7.960	7.960	7.960	7.960	3.478	3.478	3.478	3.478
Adj R2	0.782	0.783	0.783	0.783	0.761	0.763	0.763	0.763
Observations	2872	2872	2872	2872	2872	2872	2872	2872
Instrumental Variable Estimates								
Online Homophily	0.807***	0.783***	0.777***	0.793**	-0.420**	-0.400*	-0.561**	-0.609**
	(0.215)	(0.247)	(0.286)	(0.316)	(0.157)	(0.216)	(0.260)	(0.270)
F-stat	46.340	36.590	32.640	28.763	46.340	36.590	32.640	28.763
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*Yes*	*Yes*	*Yes*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*Yes*	*Yes*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*Yes*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables are the inverse hyperbolic sine (IHS) of minutes spent on Facebook (columns 1-4) and on other social media (columns 5-8). Other social media include Instagram and Twitter. The data on web searches comes from ComScore and refers to the first trimester of 2016. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. All specifications control for the total time in minutes and its squared term. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table A.11: Homophily of Online Connections and Total Time on Social Media, IHS

Dep. Variable:	IHS Any SM Visits				IHS Any SM Minutes			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GC Homophily Shock	0.255***	0.222***	0.199***	0.201***	-11639.949**	-17809.180***	-15089.581**	-14227.653**
	(0.060)	(0.056)	(0.058)	(0.062)	(5441.160)	(6191.858)	(6424.644)	(6544.979)
Mean of Dep. Var.	5.540	5.540	5.540	5.540	2.8e+04	2.8e+04	2.8e+04	2.8e+04
Adj R2	0.875	0.875	0.875	0.876	0.581	0.601	0.605	0.609
Observations	2872	2872	2872	2872	2872	2872	2872	2872
Instrumental Variable Estimates								
Online Homophily	0.663***	0.608***	0.630***	0.642***	-29277.110**	-46964.342***	-46038.050***	-44029.514**
	(0.127)	(0.141)	(0.166)	(0.178)	(12575.841)	(13509.801)	(16418.218)	(16732.049)
F-stat	45.242	35.665	31.973	28.722	46.340	36.590	32.640	28.763
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*Yes*	*Yes*	*Yes*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*Yes*	*Yes*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*Yes*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables measure the IHS total number of social media visits (columns 1-4) and IHS total minutes spent on social media (columns 5-8). Social media include Facebook, Instagram and Twitter. The data on web searches comes from ComScore and refers to the first trimester of 2016. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications, we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. All specifications control for the total number of web visits and its squared term. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table A.12: Homophily Shock and Bars and Restaurants Visits, 2019 (IHS)

Dep. Variable:	IHS Visits to Bars and Restaurants					
	(1)	(2)	(3)	(4)	(5)	(6)
GC Homophily Shock	-0.058**	-0.083**	-0.061*	-0.084***	-0.085**	-0.075**
	(0.024)	(0.040)	(0.036)	(0.030)	(0.033)	(0.035)
Mean of Dep. Var.	10.006	10.006	10.006	10.006	10.006	10.006
Adj R2	0.940	0.941	0.946	0.948	0.948	0.948
Observations	36564	36564	36564	36564	36564	36564
Instrumental Variable Estimates						
Online Homophily	-0.135**	-0.355*	-0.178*	-0.275**	-0.315**	-0.280*
	(0.066)	(0.209)	(0.105)	(0.106)	(0.137)	(0.143)
F-stat	11.996	10.895	55.893	42.157	38.608	31.951
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the inverse hyperbolic sine (IHS) of the number of visits to bars and restaurants (NAICS code 7224 and 7225). The source of the data is Safegraph, it covers all of 2019 and varies at the county by month level. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index computed using the weighted average of socio-economic distance to all counties in each county network, using the 2016 Facebook connections as weights. All specifications control for the total number of visits and its squared term. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table A.13: Homophily Shock and Total Visits, 2019 (IHS)

Dep. Variable:	IHS Total Visits					
	(1)	(2)	(3)	(4)	(5)	(6)
Gmail Homophily Shock	-0.005	0.036	-0.006	0.032	0.030	0.048
	(0.045)	(0.045)	(0.044)	(0.044)	(0.045)	(0.041)
Mean of Dep. Var.	11.784	11.784	11.784	11.784	11.784	11.784
Adj R2	0.924	0.936	0.957	0.963	0.963	0.964
Observations	36564	36564	36564	36564	36564	36564
Instrumental Variable Estimates						
Online Homophily	-0.008	0.130	-0.014	0.078	0.085	0.141
	(0.071)	(0.140)	(0.107)	(0.101)	(0.120)	(0.111)
F-stat	17.532	10.861	46.545	37.687	39.842	35.224
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the inverse hyperbolic sine (IHS) of the number of visits to any establishment. The source of the data is Safegraph, it covers all of 2019 and varies at the county by month level. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index computed using the weighted average of socio-economic distance to all counties in each county network, using the 2016 Facebook connections as weights. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table A.14: Log Total Connections, 2016

Dep. Variable:	Log All Connections, 2016					
	(1)	(2)	(3)	(4)	(5)	(6)
Gmail Homophily Shock	0.127***	0.194***	0.086**	0.007	0.017	0.026
	(0.036)	(0.046)	(0.035)	(0.026)	(0.026)	(0.023)
Mean of Dep. Var.	15.397	15.397	15.397	15.397	15.397	15.397
Adj R2	0.757	0.777	0.875	0.903	0.903	0.911
Observations	3042	3042	3042	3042	3042	3042
Instrumental Variable Estimates						
Online Homophily	0.204**	0.702**	0.204**	0.018	0.049	0.074
	(0.095)	(0.277)	(0.078)	(0.064)	(0.074)	(0.068)
F-stat	17.732	10.914	43.657	35.280	37.285	32.856
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the log number of total Facebook links in 2016. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table A.15: IHS Total Connections, 2016

Dep. Variable:	IHS All Connections, 2016					
	(1)	(2)	(3)	(4)	(5)	(6)
Gmail Homophily Shock	0.127***	0.194***	0.086**	0.007	0.017	0.026
	(0.036)	(0.046)	(0.035)	(0.026)	(0.026)	(0.023)
Mean of Dep. Var.	16.091	16.091	16.091	16.091	16.091	16.091
Adj R2	0.757	0.777	0.875	0.903	0.903	0.911
Observations	3042	3042	3042	3042	3042	3042
Instrumental Variable Estimates						
Online Homophily	0.204**	0.702**	0.204**	0.018	0.049	0.074
	(0.095)	(0.277)	(0.078)	(0.064)	(0.074)	(0.068)
F-stat	17.732	10.914	43.657	35.280	37.285	32.856
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the inverse hyperbolic sine of the number of total Facebook links in 2016. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table A.16: Controlling for Total Facebook Links Does not Affect Main Results

	First-Stage		Log Facebook Visits		Log Bar Visits		Econ Connect		Pol Homogeneity		Extreme Margins	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Gmail Homophily Shock	0.344***	0.340***	0.203***	0.207***	-0.070**	-0.073**	-0.172***	-0.170***	-0.024***	-0.024***	-0.050***	-0.051***
	(0.060)	(0.060)	(0.061)	(0.061)	(0.032)	(0.030)	(0.059)	(0.058)	(0.006)	(0.006)	(0.014)	(0.015)
Log All Links, 2016		0.174**		-0.119*		0.119		-0.161***		0.010***		0.031**
		(0.082)		(0.065)		(0.084)		(0.051)		(0.003)		(0.014)
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Political Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Trends	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
KP F-Stat	32.856	32.259	28.251	28.396	31.951	30.965	30.942	30.793	32.856	32.259	28.935	29.786
Adj R2	0.834	0.836	0.867	0.867	0.953	0.954	0.872	0.874	0.876	0.877	0.777	0.778
Mean of Dep. Var.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	3042	3042	2872	2872	36504	36504	2943	2943	3042	3042	2738	2738

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The Table presents the main results of the paper controlling for log number of total Facebook links each county has. The dependent variables are the six main outcomes of the paper: first-stage in columns 1 and 2, Log Facebook visits in column 3 and 4, Log Bar visits in columns 5 and 6, Economic Connectedness in columns 7 and 8, political homogeneity in columns 9 and 10 and the share of precinct within a county with electoral margins larger than 60 points in 2016 in columns 11 and 12. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Robust standard errors clustered by state in parentheses.

Table A.17: Gmail Complementarity Shock Reduces Extreme Partisan Identity

Dep. Variable:	Extreme Partisanship					
	(1)	(2)	(3)	(4)	(5)	(6)
Post × GC Homophily Shock	-0.007**	-0.007**	-0.006**	-0.006**	-0.007**	-0.007**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*No*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*No*
Demographic Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*No*
Political Controls	*No*	*No*	*Yes*	*Yes*	*Yes*	*No*
Demographic Trends	*No*	*No*	*No*	*Yes*	*Yes*	*No*
Individual Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
County FEs	*No*	*No*	*No*	*No*	*No*	*Yes*
Year FEs	*No*	*No*	*No*	*No*	*No*	*Yes*
Adj R2	0.004	0.004	0.005	0.005	0.029	0.041
Mean of Dep. Var.	0.417	0.417	0.417	0.417	0.417	0.417
Observations	391880	391880	391880	391880	391880	391880

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is an indicator for the respondent self-identifying as either strong democrat or strong republican. Gmail Homophily Shock is standardized and measures the differential Gmail complementarity in the six quarters following the API change. Post is an indicator equal to one for post-2010 observations. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Individual controls include gender, age and age squared, race, indicators for family income brackets, indicators for education brackets, and indicators for marital status. Robust standard errors clustered by state in parentheses.

Table A.18: Gmail Complementarity Shock Reduces Extreme Ideology

Dep. Variable:	Extreme Ideology					
	(1)	(2)	(3)	(4)	(5)	(6)
Post × GC Homophily Shock	-0.006***	-0.006***	-0.006***	-0.006***	-0.006***	-0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*No*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*No*
Demographic Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*No*
Political Controls	*No*	*No*	*Yes*	*Yes*	*Yes*	*No*
Demographic Trends	*No*	*No*	*No*	*Yes*	*Yes*	*No*
Individual Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
County FEs	*No*	*No*	*No*	*No*	*No*	*Yes*
Year FEs	*No*	*No*	*No*	*No*	*No*	*Yes*
Adj R2	0.001	0.002	0.002	0.003	0.011	0.025
Mean of Dep. Var.	0.210	0.210	0.210	0.210	0.210	0.210
Observations	391880	391880	391880	391880	391880	391880

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is an indicator for the respondent self-identifying as either very conservative or very liberal. Gmail Homophily Shock is standardized and measures the differential Gmail complementarity in the six quarters following the API change. Post is an indicator equal to one for post-2010 observations. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Individual controls include gender, age and age squared, race, indicators for family income brackets, indicators for education brackets, and indicators for marital status. Robust standard errors clustered by state in parentheses.

Table A.19: Online Homophily Has No Impact on Republican Vote Share, 2020

Dep. Variable:	Republican Share, 2020					
	(1)	(2)	(3)	(4)	(5)	(6)
GC Homophily Shock	0.111***	0.017***	0.010***	0.004	0.003	0.001
	(0.023)	(0.006)	(0.004)	(0.003)	(0.003)	(0.003)
Mean of Dep. Var.	0.664	0.664	0.664	0.664	0.664	0.664
Adj R2	0.407	0.934	0.960	0.967	0.968	0.971
Observations	3042	3042	3042	3042	3042	3042
Instrumental Variable Estimates						
Online Homophily	0.178***	0.061**	0.024***	0.011	0.008	0.003
	(0.021)	(0.024)	(0.007)	(0.008)	(0.009)	(0.009)
F-stat	17.732	10.914	43.657	35.280	37.285	32.856
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the Republican vote share in the 2020 presidential election. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Individual controls include gender, age and age squared, race, indicators for family income brackets, indicators for education brackets, and indicators for marital status. Robust standard errors clustered by state in parentheses.

Table A.20: Online Homophily Has No Impact on Turnout, 2020

Dep. Variable:	Log Tot Votes, 2020					
	(1)	(2)	(3)	(4)	(5)	(6)
GC Homophily Shock	-0.084***	-0.013	-0.013	0.009	0.012*	0.003
	(0.011)	(0.012)	(0.008)	(0.006)	(0.007)	(0.007)
Mean of Dep. Var.	9.562	9.562	9.562	9.562	9.562	9.562
Adj R2	0.987	0.995	0.997	0.998	0.998	0.998
Observations	3042	3042	3042	3042	3042	3042
Instrumental Variable Estimates						
Online Homophily	-0.134***	-0.045	-0.031	0.023	0.033*	0.008
	(0.032)	(0.039)	(0.020)	(0.016)	(0.018)	(0.021)
F-stat	17.732	10.914	43.657	35.280	37.285	32.856
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the log turnout the 2020 presidential election. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. Individual controls include gender, age and age squared, race, indicators for family income brackets, indicators for education brackets, and indicators for marital status. Robust standard errors clustered by state in parentheses.