Socializing Alone:
How Online Homophily Has Undermined Social Cohesion in the U.S.

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]CUNEF
[4]University of Zurich

January 11, 2026

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 over data sharing of user information in the early expansion phase of Facebook, which induced persistent variation across counties. We find evidence that these homophilic connections on Facebook made people spend more time on the platform but socialize less *offline*, as measured through bar, restaurant, and live sports event visits. The effects are substantial: we estimate that a one-standard-deviation increase in online homophily is associated with approximately a 25 percent reduction in offline socialization. This effect was accompanied by a negative effect on local social capital by making individuals less connected across income strata. Moreover, political opinions within counties became less homogeneous, 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 social cohesion at the local level.

Keywords: Social Media, Networks, Homophily, Social Capital

*The authors are in alphabetical order. Enikolopov 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 a Lecturer at Cunef University. 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 Bayreuth, 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. Ruben Enikolopov and Maria Petrova acknowledge financial support from the Plan Nacional project PID2023-150768NB-I00, and the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S), funded by MCIN/AEI/10.13039/501100011033.

1 Introduction

Homophily – the principle that individuals prefer to associate with similar others – coined by (Lazarsfeld et al., 1954) has been shown to be one of the most robust findings in the social sciences. People are more likely to form connections with those who resemble them in terms of race, socioeconomic status, political preferences, and other attributes (McPherson et al., 2001; Jackson, 2008). In fact, the observation dates back millennia. Plato famously stipulated in *The Phaedrus* that 'similarity begets friendship' and Aristotle noted in *Nicomachean Ethics* that people seem to 'love those who are like themselves'. The empirical regularity is perhaps unsurprising, as friendships and social connections traditionally required face-to-face interactions in shared socioeconomic contexts. Connecting with dissimilar people across distances was prohibitively costly. Social interactions were local, in-person, and offline.

The advent of online social media platforms two decades ago, most notably Facebook, fundamentally shifted the cost and ease of long-distance interactions. In principle, this technology allows users to connect with a diverse and distant set of people living in dissimilar socioeconomic conditions, values, norms, or political attitudes.[1] Yet, empirically, friendship connections on these platforms appear to also be dominated by homophily-based selection, and this tendency can be further amplified by platforms' algorithms (Tarbush and Teytelboym, 2012; Bakshy et al., 2015; Conover et al., 2021).

Meanwhile, this technology-induced shift in connectedness has been accompanied by some concerning social and political trends over the last two decades. First, since Robert Putnam's seminal work that documented the erosion of civic life in the pre-Internet era (Putnam, 2000), time-use surveys show that the US has seen an acceleration of social isolation and a decline in face-to-face socialization across a broad range of measures (Kannan and Veazie, 2023; Anders and Pallais, 2025). For example, in 2004—the year of Facebook's launch in US colleges—Americans spent on average about one hour per day socializing with friends; in 2019 this number was down to 34 minutes per day, a 43% reduction. People feel increasingly lonely and disconnected to their local community, even triggering the US Surgeon General to declare an "epidemic of loneliness and isolation" in 2023.[2] Simultaneously, the political landscape has also shifted. Political identities have become stronger.

[1]Early studies suggested that online interactions could bring people closer, lead to the "death-of-distance", transform the world into a "global village", bridge gaps, and unite communities (Cairncross, 2002; Alstyne and Brynjolfsson, 2005). Alternatively, online connectivity was hypothesized to lead to fragmented interactions and divide people, rather than unite them, creating filter bubbles, or "echo chambers" (Sunstein, 2001, 2007).

[2]US Department of Health and Human Services, Office of the Surgeon General, *Our Epidemic of Loneliness and Isolation: The US Surgeon General's Advisory on the Healing Effects of Social Connection and Community* (2023), available at: `https://www.hhs.gov/sites/default/files/surgeon-general-social-connection-advisory.pdf` (accessed 19 October 2025). The report also notes the potential role due to the rise of social media and that the percentage of US adults who reported using social media increased from 5% in 2005 to roughly 80% in 2019.

In 2004, the share of Americans reporting they "strongly" identified with either the Democratic or Republican party was 33 percent. In 2020, it had risen to 44 percent.[3] These parallel trends raise a fundamental question about how online networks characterized by homophily have reshaped social life and political identification. Has the same mechanism that allowed people to more easily find like-minded communities online simultaneously eroded local bonds? Has it affected political preferences?

In this paper, we pursue the idea that online homophily has fundamentally shifted how people socialize both *online and offline* and we study the implications for local social cohesion in the United States. Since we are interested in capturing potential spill-over effects, we are focusing on homophily in connections at the level of communities rather than individual users. To identify causal effects of online homophily, we face a major empirical challenge: Since online friendship is driven by homophily-based selection, we need some plausibly exogenous shifter in the relative cost of connecting to people living elsewhere under socially similar contexts, versus living elsewhere under dissimilar contexts, that is uncorrelated with other drivers of local social cohesion.[4] We develop a unique identification strategy to overcome this challenge. To our knowledge, this is the first paper to document the long-run impact of online homophily on local social cohesion.

Specifically, we study the impact of the network structure of county-to-county Facebook links on various types of interpersonal interactions and political preferences *within* US counties. Our goal is to document the effects in counties that ended up with Facebook links to more socially similar counties, as opposed to socially dissimilar ones.[5] To overcome the main empirical challenge, we make use of a unique natural experiment by exploiting a change in friendship recommendations that resulted from a data-sharing conflict between Facebook and Google that started in 2010, a period during which many new users were joining the social network. Before the conflict, with the *Find Friends* feature provided by Facebook, new users of the network were offered to import their email contact data to Facebook and easily connect to their contacts who were already on the platform. This was possible as Facebook was effectively given permission to cross-reference

[3]American National Election Studies report these numbers on their homepage. In the 2004 election, the share of Strong Democrats was 16.5% and Strong Republicans 16.5%. In 2020 the shares were 23.2% and 21.0%, respectively; see `https://electionstudies.org/data-tools/anes-guide/anes-guide.html?chart=party_identification_7_pt` (accessed 19 October 2025).

[4]The identification challenge is similar to identifying demand curves. Since there is demand for homophily, we need a supply shock to identify the effects of a greater quantity of homophily consumed, so to speak.

[5]As we describe in the data construction section, the degree of homophily in online connections will be captured by social distance between every pair of counties by looking at the differences in socio-economic and political characteristics of the counties. To construct a measure of online homophily for each county, social distance with all other counties is weighted by their Facebook friendship links. We refer to this variable as Online Homophily.

all Facebook users with all user email addresses.[6] Google, however, did not have reciprocal access to user information from Facebook, which caused a conflict between the companies. In November 2010, Google made a policy change that led to new Facebook users losing the ability to use the API to automatically import their Gmail contacts. At the same time, new Facebook users who were using all *other* email services were not affected by the Facebook-Google conflict and could still easily establish connections with people from their email contact list. The situation continued until April 2012, when Facebook removed the option of finding friends using email contacts entirely and switched to algorithmic recommendations of friends.

As a result of this conflict, between November 2010 and April 2012, new Facebook users were less likely to connect to each other if both of them had Gmail accounts. In several data construction steps, described in detail in our empirical strategy section, we exploit this episode to construct variation in network structure based on the popularity of email clients across space and time, as well as various pre-determined county-level characteristics. In essence, we construct a measure which captures that this change in friendship recommendations, in some counties, made it easier to connect with friends in other socially similar counties — where social distance is based on a broad vector of pre-determined demographics, socioeconomic factors, voting patterns, and ideology — and in other counties, it made it more difficult. We call this variable the *homophily shock*. The main idea is that this shock influenced the hassle cost of online homophily during the critical expansion period of Facebook, which, in turn, under some inertia, may have had a persistent long-term influence on the network structure.

We first document that, as hypothesized, the shock had a persistent long-run effect on the degree of online Facebook homophily in US counties years later, detectable in both 2016 and 2020. We provide numerous robustness checks to establish that this shock can be deemed as a source of credible variation for capturing causal relationships, including placebo tests which exploit data on email client popularity for other time periods.[7] These results not only provide us with a source of quasi-exogenous variation, but also make a substantive contribution by demonstrating that friendship recommendation algorithms have an important long-term effects on the structure

[6]The *Find Friends* feature was a key step during the sign-up process and substantially lowered the hassle cost of connecting to all email contacts who already had Facebook accounts at once through a simple click. It also automatically detected if a user's email contacts were not Facebook users, and allowed the user to invite everyone to sign up in one click. While the old sign-up page can no longer be accessed, the feature was also prominently displayed on the website for existing Facebook users in various ways. An example of this feature, showing that users had to submit their email accounts, thereby allowing Facebook to access your email account, can be found on the Internet Archive from 2009: `https://web.archive.org/web/20091225045036/http://www.facebook.com/find-friends/index.php` (accessed 19 October 2025)

[7]In principle, the shock could also lead to a greater overall number of Facebook friends. Empirically, we do not find a significant effect on the number of connections. We also show that our results are robust even if we control for the number of connections. 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.

of online friendship networks,

We use the shock to study the long-term effect of online homophily. We start by studying how the shock affects various *online* and *offline* interactions. First, we look at the effect on *online* interactions. We use data from ComScore for 2016 to 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 by the shock to have more socially similar, or like-minded, connections. Furthermore, we find that higher Facebook homophily implies fewer visits to other social media platforms, such as X (formerly Twitter), Instagram, or Reddit. The first effect, however, dominates, so that an increase in homophily leads to an increase in the consumption of social media. The implied effects are substantial in magnitude; a one standard deviation increase in online homophily is associated with an increase in visits to Facebook by approximately 66%, and more than 70% increases in time spent on social media in general. This result confirms that more homophilic social networks provide an online experience that is more attractive to users, which results in increased usage of social media.

As social media use shifts toward online networks composed of more socially similar individuals, a key consequence is the potential weakening of traditional, place-based community bonds—an outcome that carries meaningful welfare implications due to the externalities such bonds typically generate. Recent work by Bursztyn et al. (2025) highlights how social media can impose negative externalities on non-users, resulting in reduced overall welfare despite high levels of usage. Heavy social-media users may even coordinate with friends in ways that effectively remove the option to engage with more diverse or higher-quality networks online. Similarly, increased online homophily, while boosting social media engagement, might have unintended negative consequences through its effect on offline socialization and the strength of local communities. For example, if users are preoccupied with online content, they may not be interested in going to a restaurant with their friends. They may then decide to spend time on social media for the simple reason that there is nobody to go to the restaurant with. Over time, this vicious cycle can lead to the erosion of offline socialization.

To test this hypothesis, we examine the effects on offline interactions within counties. Using SafeGraph geographic mobility data for 2019 (capturing long-term outcomes after several years of increased online homophily), we classify establishment visits by type, focusing on places where social interactions are most likely to occur, such as bars and restaurants or live sports events. We find that the homophily shock led to a persistent decline in visits to such places. We do not find significant effects for most other places, and we find a positive effect on visits to recreational venues that are not associated with social interactions (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 in particular and socializing online in general, which results in them spending less time socializing offline with their friends and families. The implied magnitudes are noteworthy. For example, a one-standard-deviation increase in online homophily is associated with approximately a 25% decrease in offline socialization, as captured by visits to bars and restaurants.

Next, we document that these effects were accompanied by important impacts on local social cohesion along several dimensions. First, we look at the effect of online homophily on local social capital. We use the data on "economic social capital" as of 2022 from Chetty et al. (2022a,b), i.e., the probability that people form connections across income strata (e.g., the rich connect to the poor). We document that higher online homophily reduces economic connectedness, leading to a decrease in this measure of local social capital.

Second, we examine the effect in the domain of political cohesion. Online homophily can affect political opinions in two ways. Because users focus on communication with like-minded people online and get constant reinforcement of their pre-existing political preferences, politics could become more extreme, and within-county voting behavior could become more one-sided. Alternatively, by reducing offline within-county communications – social interactions in another realm with a high baseline degree of homophily – political opinions within counties could become less similar and less polarized. We find that the second effect dominates, so that higher online homophily made local political opinions less homogeneous starting in 2016. The probability that two randomly picked county residents vote for the same party goes down. We also find that higher online homophily decreased the probability of extreme vote margins but had little significant impact on within-county measures of dispersion of political opinions, such as inter-quartile range and standard deviation of vote shares.[8] Overall, we see that online homophily reduces local social cohesion by reducing local social capital and political cohesion.

While the paper primarily focuses on the effects related to local social cohesion, we also examine whether online homophily can have implications for the broader debate on polarization in the country as a whole. Does it lead to more extreme viewpoints and stronger partisan divides? In particular, we show that online homophily made people less extreme in answering survey questions; for example, an increase in online homophily made Cooperative Election Study respondents less likely to say that they are "Strong Democrats" or "Strong Republicans". Thus, we find no evidence that online homophily leads to stronger partisanship and greater polarization overall in the country. If anything, it *reduces* such polarization across the country, as measured by the likelihood of identifying strongly with one political party.

[8]It is noteworthy that online homophily did *not* disproportionately benefit any particular political party; the policy change seems to matter for the dispersion of political preferences rather than for their mean level. We also document that there is no significant effect of online homophily on turnout.

As an important robustness test, for all the outcomes affected by online homophily, we show placebo tests based on the popularity of email clients across time. The effects are not significant for the placebo shock based on email client popularity before November 2010, they are consistently significant for the period between November 2010 and April 2012, and they are much smaller for the period after 2012, when "Find Friends" by email was discontinued, but long-term consequences are, nevertheless, likely to remain. This indicates that a shock to online homophily was indeed driven by changes in friendship suggestions that resulted from the temporary conflict between Facebook and Gmail, rather than the differences in the average popularity of different email clients. [9]

Overall, we conclude that the effect of online homophily, estimated with the help of an exogenous shock induced by the change in friendship suggestions due to the Gmail-Facebook conflict, was important for the patterns of social media consumption, interpersonal communications, local social capital, and political opinions. Thus, our results suggest that technologies capable of transforming the world into a "global village" may come at the cost of unraveling traditional community bonds at the local level, to the extent that these "death-of-distance" technologies tend to lead to the creation of online connections dominated by homophily.

Our paper contributes to several strands of literature. First, it adds 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; Bursztyn et al., 2025), mental health (Braghieri et al., 2022), hate crime and xenophobia (Müller and Schwarz, 2020, 2023; Bursztyn et al., 2024), voter turnout (Bond et al., 2012), and 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). The evidence on the impact of social media and the Internet on political polarization is mixed, with some studies finding a positive effect (Allcott et al., 2020; Levy, 2021; Melnikov, 2022), some finding no effect (Boxell et al., 2017; Nyhan et al., 2023), and some a negative effect Asimovic et al. (2021). There is an ongoing debate about 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., 2025). We contribute to this literature by studying the causal impact of homophily in social media, rather than the presence of social media

[9] An additional 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 (which are not used in the construction of our measure of online homophily). This finding is consistent with the idea that the 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 does not significantly depend on the shock, so the results are not simply driven by a non-linear effect. Similarly, we document a stronger impact of the shock in urban areas, which suggests that online homophily undermined social cohesion, particularly in cities, as opposed to rural areas.

and the Internet. Our findings highlight potential heterogeneity in the effects of social media and, thus, help to reconcile some conflicting evidence in this literature.

Second, our work relates closely to the literature on networks and homophily in networks. In offline networks, homophily increases because of the same-type preference and biased matching Currarini et al. (2009), but preferences for homophily can increase integration as a result of a random search for friends-of-friends (Bramoullé et al., 2012). The study of causal effects of network structure in offline networks usually focuses on peer effects (e.g., Sacerdote, 2001; Carrell et al., 2009; Waldinger, 2012). In online networks, connections are characterized by a high degree of homophily that limits exposure to cross-cutting content Bakshy et al. (2015). People on the Internet mostly interact with like-minded content (Sunstein, 2001, 2007). Homophily affects the diffusion and exposure to like-minded information (Halberstam and Knight, 2016) and limits connectivity between right- and left-leaning users (Conover et al., 2021). At 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 providing the first, to the best of our knowledge, evidence on the causal impact of online homophily.

Finally, we contribute to the literature on social capital. *Bowling Alone*, the foundational 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; Olken, 2009; Campante et al., 2022), 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 conflicting evidence in earlier studies.

The rest of the paper is organized as follows. Section 2 summarizes the data sources we use. Section 3 discusses our empirical strategy. Section 4 presents empirical results. Section 5 concludes.

2 Data

This section describes the sources of the data and the construction of the measures used in the analysis. The main unit of analysis is the US county. In a 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.

2.1 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).[10] We use the measures of connectedness for 2016 and 2020. The main measure of social connectedness between two counties equals the number of Facebook connections between users from these two counties, divided 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 the relative popularity of different email services over time and geography, we use Google's Search Volume Index (SVI) by DMA and quarter between 2006 and 2016 for Gmail, Yahoo! Mail, and Hotmail (Outlook.com).

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

We extract county-level electoral results (1996-2020) from Leip (1993). 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.

We leverage survey data from the Cooperative Congressional Election Studies (CCES) to reconstruct variation over time in the intensity of political preferences.[11]

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

[11] The Cooperative Election Studies is the former Cooperative Congressional Election Study (CCES), data available at https://cces.gov.harvard.edu/

Social Media Usage. We use browsing history 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, Instagram, and Reddit.

Offline Activity. Data on visits to different establishments comes from SafeGraph.[12] The data we obtained provides information on visits to several types of commercial establishments for 2019. We aggregate the data at the county-by-month level, cross-walking data from the census block level to the county level.

Social Capital. We use data from Chetty et al. (2022a,b) to measure local social capital. We focus on the degree of economic connectedness in US counties as of 2022, which was shown to be the component of social capital most predictive of inter-generational income mobility. 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.

2.2 Measure of Social Similarity

To measure how similar the people living in different counties are, we look at how close they are in terms of their demographic characteristics and political preferences. In particular, for each pair of counties, 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). The choice of the variables is informed by the literature on the most important dimensions of homophily (e.g., McPherson et al., 2001; Jackson et al., 2023). We then take the first principal component of these twelve differences and use its inverse as the measure of social similarity between each pair of counties, $Social\_Similarity_{ij}$.

[12] 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.

2.3 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 to which it is connected, using the share of 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, excluding connections within the same county. 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 darker colors indicating higher levels of Online Homophily.

[Figure 1 about here.]

3 Empirical Strategy

In this section, we summarize our empirical strategy and describe the construction of the key variable that we use as a source of quasi-exogenous variation.

3.1 Empirical Challenge and the Source of Variation

We are interested in studying the effect of online homophily on social media usage, offline behavior, local social capital, and political preferences. Online homophily is likely to be highly 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 others' opinions, and the extent of inter-group contact. These factors might also separately affect our outcomes of interest. Thus, we need an identification strategy to address these endogeneity concerns.

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 friendships with more similar individuals after meeting with them). In online networks, the most important methods for establishing friendship connections include i) direct search for specific users; ii) responding to friendship recommendations by the platform; and iii) connecting with friends of already existing friends (even if they are not suggested by the platform). To identify the effect of online homophily, we focus on the role of friendship suggestions by Facebook. In particular, we exploit the 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 users as they joined the platform, thus affecting exposure. We show that this variation led to long-term changes in the resulting homophily of Facebook connections, indicating that this shock was not fully compensated by endogenous friendship patterns when users employed alternative methods of searching for friends.[13]

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 email login credentials so that the program could quickly tell them which of their email contacts were already on Facebook, allowing them to expand their network from the outset. However, in November 2010, Google started invoking reciprocity from Facebook, refusing to share information about Gmail contacts 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. After April 2012, there were no explicit differences between the users of different email clients, but Gmail users could still be affected by the dynamic effects of discrimination during the previous period. The conflict between Google and Facebook was widely covered in the media; see Figure A.3 for examples of the headlines.

To proceed, we note that the relative popularity of different email clients has been changing over time. 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 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

[13]See Ugander et al. (2012) for the study of the friending bias on Facebook.

(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 variable that measures the quasi-exogenous shock to online homophily at the county level, we start by investigating if indeed Facebook connections between pairs of counties responded to the 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 during this period, county pairs where Gmail was a more popular 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 how the formation of Facebook links was changing during the time of the Google-Facebook incident. Unfortunately, such data before 2016 is not available. Nonetheless, we can test for long-run effects by examining whether 2016 Facebook connections 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).

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

where 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,[14] γ_i and μ_j are county fixed effects, and ϵ_{ij} is an error term.

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 the coefficients 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 popularity, so we can use it as a source of quasi-exogenous variation in connections between pairs of counties.

[Figure 3 about here.]

[14] We define Gmail complementarity between two counties as the difference between the complementarity in Gmail relative to the average complementarity for the other two main email clients

$$gmail_{jit} = gmail_{it} \times gmail_{jt} - 0.5(yahoo_{it} \times yahoo_{jt} + hotmail_{it} \times hotmail_{jt})$$

Construction of the Homophily Shock Variable. We now exploit the Gmail complementarity shock for the county-pair connections to construct our measure of the shock to Online Homophily at the county level. Intuitively, we exploit the fact that if this shock to pairwise connections happens to be larger for connections with like-minded counties compared to distant-minded counties, it will be harder to connect with similar counties and, thus, will lead to lower homophily. To formalize this intuition, we construct the Homophily Shock in two steps (we refer to this variable as the 'Gmail Homophily Shock', GH_shock_i, since it is built on the Gmail complementarity). First, we compute a social distance between US counties using the methodology described in section 2.2. Employing this distance, for each county we divide the US counties between high and low social distance counties.[15] 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:

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

where

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

d indicates high ($d = HI$) or low ($d = LO$) social distance counties, and t are quarters where $t = 0$ is the last quarter of 2010, when the API first changed.

This is a key variable in our analysis. Essentially, this variable compares whether county i gets more connections to counties with high or low social distance because of the change induced by the Facebook-Google conflict. In subsequent analysis, we always control for the same variable computed during the pre-period using the last six quarters before the start of the conflict between Facebook and Google to ensure that identification comes from the *change* in differences in Gmail complementarity during the Facebook-Google conflict.[16]

For example, let's consider how we construct the Gmail Homophily Shock for Blount County, AL. Figure A.4, plots the Gmail Complementarity between Blount County and all the counties in the US. Figure A.5 plots counties with high or low social distance to Blount County. By combining the two we arrive at Figure A.6. The left bar on the graph represents the value of the $\sum_{t=1}^{6} \overline{gmail_{it}}^{HI}$ term from equation (3), while the right bar represents the $\sum_{t=1}^{6} \overline{gmail_{it}}^{LO}$ term from equation (3). Since the first term in equation (3) is higher than the second term and since

[15] In our baseline definition, we split the data using the top and bottom terciles. Appendix Tables A.20 shows that the results are similar when splitting using the median, the top and bottom quartiles, and when using a continuous measure based on distance to the median value of the social distance distribution.

[16] More specifically, we control for a variable computed using (3) for 6 quarters before the policy change.

higher Gmail complementarity makes it harder to establish connections, Blount County is more likely to be connected to counties with low social distance for quasi-random reasons.

Finally, we repeat this exercise for all the counties in the United States. Figure A.7 shows the value of the two different terms in equation (3) in every county in Alabama.[17] As one can see, Facebook-Google conflict introduced heterogeneous changes to county-to-county homophily in different counties: in some counties, the shock, as computed in equation (3), turns 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 into a higher complementarity shock with similar counties than with socially distant counties for the first group and vice versa for the second group, pushing these counties to have a lower level of homophily.

We map the residual variation of this measure, controlling for Baseline Controls and DMA fixed effects in Figure 4. As one can see, there is a high degree of heterogeneity in this variable both within and across American states.

4 Results

4.1 Gmail Homophily Shock and Online Homophily

In this sub-section, we check whether friendship recommendations have a long-term effect on the structure of the friendship network. In particular, we test whether Gmail Homophily Shock, the variable we just constructed, is a good predictor of online homophily in subsequent years. Formally, we estimate the following equation:

$$Online\_Homophily_i = \beta_0 + \beta_1 GH\_shock_i + \beta_2 X_i + \epsilon_i \tag{4}$$

where $Online\_Homophily_i$ is our measure of online homophily as defined in equation (1); GH_shock_i represents the Gmail Homophily Shock defined in equation (3); X_i is a set of county-level controls, which includes Gmail Homophily Shock defined by equation (3), computed in pre-period, i.e. 6 quarters before November 2010; ϵ_i is an error term. In all specifications, we control for the pre-period Gmail complementarity using the last six quarters before the start of the conflict between Facebook and Google to ensure that the identification comes from the *change* in Gmail complementarity during the Facebook-Google conflict. To facilitate the interpretation of coefficients, we standardize our independent variables. In the benchmark specifications, we cluster standard errors at the state level. We also show that the statistical significance of our results is robust to using

[17] The variation of the final variable we construct, the difference between the two bars for each county, is shown in Appendix Figure A.8.

randomization inference (see Figures A.13 and A.14 in the Online Appendix).

We report our estimates of equation (4) in Table 1, where we gradually add more and more 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 include political homogeneity in 2008.[18] Demographic Trends include differences between 2010 and 2000 for all the Baseline, Demographic, and Political Controls.

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 0.754, positive and significant at the 1% level. The magnitude of the effect decreases to 0.355 after the addition of Baseline Controls, but remains relatively stable with the addition of more controls in the most saturated specifications, converging to 0.308, significant at 1%. Finally, in column 7, we obtain similar results if we construct our dependent variable, Online Homophily, using 2020, rather than 2016 Facebook links. The point estimate is equal to 0.312, which is very similar to the estimate in column 6.

The magnitude of our preferred specification (column 6) implies that a one-standard-deviation increase in the Gmail Homophily Shock increases Online Homophily by about 31% of a standard deviation.

[Table 1 about here.]

[Figure 4 about here.]

For the interpretation of our results, it is important to understand what kind of connections are affected by the change in friendship recommendations. Although we cannot distinguish between strong and weak links in data on Facebook connections, we expect that the variation in the friend suggestion policy predominantly affects the formation of weak links, since strong links are likely to be established through active searches for specific friends, regardless of the friend suggestion policy. This is especially relevant for cross-county connections, which 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

[18] Political homogeneity is defined as one minus the Herfindahl Index calculated using Democratic and Republican vote shares. In this case, assuming the Republican and Democratic vote share sum up to 1, the formula boils down to $1 - 2r(1 - r)$, where r is the Republican vote share. See more details about this variable and the rationale to use it in section 4.5.

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). 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 with respect to the county-level predetermined characteristics. We perform balance tests by estimating specifications similar to equation (4), where instead of the online homophily we use each of the control variables as the dependent variable (taking them out of the list of independent variables when we use them as outcome variables). Figure 5 plots the estimated coefficient of these balance tests. While some coefficients appear unbalanced when we focus on the endogenous variable, most become indistinguishable from zero when we leverage our source of identifying variation. Although the small number of statistically significant coefficients in a battery of tests is consistent with the lack of imbalances, we control for socio-demographic, economic, and political variables and their changes in all specifications.

[Figure 5 about here.]

[Figure 6 about here.]

To provide additional evidence that we identify changes in online homophily that are driven by the conflict between Facebook and Google and not some other underlying differences between counties, we can look separately at the relationship between our outcomes of interest and Gmail Homophily Shock, computed during the 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 6a shows these results 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 Gmail Homophily Shock and online homophily in the pre-period is small in magnitude and far from being statistically significant, while the coefficient of interest is positive and statistically significant, consistent with the results in Table 1. The coefficient for the post-period is much smaller in size but statistically significant, which is consistent with the existence of dynamic effects in network formation.

The results confirm that Gmail complementarity played an important role in the formation of connections in Facebook only during the conflict. Nevertheless, we always control for the 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.

4.2 Online Homophily and Usage of Social Media

We proceed to studying the impact of online homophily on online activity. We use different measures of online activity derived from ComScore's Internet browsing data. The results of the estimation of equation (4) with Internet browsing measures are presented in the top panel of Table 2. 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 platforms, that is Twitter, Instagram, and Reddit.[19]

Our most basic specification always includes baseline controls and DMA fixed effects; we gradually add demographic, political, and trend controls. In all specifications, we control for the total number of visits and its square term to account for differences in total online activity across different counties, but we show in Table A.3 in the Online Appendix that there is no statistically significant effect of homophily on total online activity.

[Table 2 about here.]

[Table 3 about here.]

The results indicate that 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 a 20.3% increase in the number of Facebook visits and a 10.3% decrease in the number of other social media visits, with the former coefficient significant at 1% and the latter at 5%.

In Table A.4 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 a one-standard-deviation increase in GH Shock leads to a 25.5% increase in the number of minutes spent on Facebook and to a 22.4% decrease in the number of minutes spent on other forms of social media.

Finally, Table 3 presents the results for total visits and total time spent on social media activity, which combines Facebook and other social media platforms. The results imply that a 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 platform by 24.2%.

[19] In the main text, we stick to log(x+1) specifications for ease of interpretation. However, for all log specifications, we report similar results with inverse hyperbolic sine (IHS) transformations of the dependent variables in the Online Appendix, and they are qualitatively similar.

We can again look at the coefficients for Gmail Homophily Shock, computed during, before, and after 6 quarters of the Google-Facebook conflict (Figure 6b). Most of the effect happens during the treatment period, with the magnitudes for pre- and post-period being noticeably smaller. Note that Tables 2-3 and Table A.4 all include Gmail Homophily Shock in the pre-period as a control to take possible baseline differences in Gmail usage and any network patterns into account. Finally, we'd like to emphasize that qualitatively, Figure 6b resembles Figure 6a, i.e., the effect on the outcome variable of interest (Facebook visits) seems to mirror the effect on online homophily, thus the identifying variation seems to be coming from the implied first stage.

Overall, the results in this section suggest that people who experienced a positive homophily shock, and who, as a result, have more connections to like-minded counties, enjoy Facebook more, go to Facebook more often, and spend more time there, at the expense of time spent on other social media. The first effect, however, dominates, so that they spend more time on social media platforms in total.

4.3 Online Homophily and Offline Activity

In this subsection, we test whether the homophily shock reduces offline interpersonal interactions. The basic idea is that increased online homophily, while boosting social media engagement, might have unintended negative consequences on offline socialization and the strength of local communities. The intuition is as follows: if users are preoccupied with online content, they may not be interested in going to a restaurant with their friends. Moreover, if they would like to go but their friends are not interested, demand for the restaurant is lower. Users may then decide to spend time on social media for the simple reason that there is nobody to go to the restaurant with. To that end, we examine the effects of online Facebook homophily on offline interactions within the county.

As a proxy for interpersonal interactions, we use visits to commercial establishments where people are likely to socialize, such as bars, restaurants, or live sports events. We leverage the mobile phones tracker data from SafeGraph covering all months of 2019 at the county-by-month level. We estimate specification (4) using visits to different types of establishments as outcome variables. Given the structure of the data, 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, although we do not find a significant effect on total number of visits (see Table A.8 in the Online Appendix).

The results on bars and restaurants are presented in Table 4. In all specifications, the coefficients for Gmail Homophily Shock are negative, which implies that higher Facebook homophily leads to a reduction in offline visits to bars and restaurants. As the set of controls changes, the resulting coefficient remains remarkably stable, ranging from -0.053 to -0.076 in different columns. The magnitude of the effects in the most saturated specification implies that a one standard deviation increase in Gmail Homophily Shock leads to 6.7% fewer visits to bars and restaurants, significant at the 5% level (see column 6).

To illustrate graphically how our identification works here, we report the results for the shock computed during our treatment period, and in the periods before and after (see Figure 6c). 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 results in Figure 6a in that the coefficient of interest is the largest both in magnitude and in terms of significance as compared with pre- and post-treatment coefficients.

So far, we have documented that the homophily shock reduced visits to bars and restaurants. To dig further into the patterns of potential offline interactions, we report similar results for other types of establishments, such as live sports events, amusement parks, religious activities, or volunteer organizations. (Figure 7). We document the negative and significant effects of the homophily shock on visits to bars and restaurants, live sports events, and amusement parks, and a marginally significant negative coefficient for volunteer organizations. All these results are consistent with a reduction in offline social interactions. We do not find any effects on visits to churches and other religious facilities, suggesting that online homophily does not constitute a substitute for this kind of social activity. The only positive coefficients that we document are for visits to facilities which are less intense in terms of face-to-face socialization: fitness centers and golf facilities.[20]

[Table 4 about here.]

[Figure 7 about here.]

Overall, our results so far are consistent with the hypothesis that Online Homophily changes the patterns of online and offline social interactions: increasing online homophily induces people to

[20] Admittedly, it is not entirely obvious how to classify the degree to which we could expect people to talk and interact with others whom they know, or if the purpose is not to socialize but rather achieve some other goal, at each type of venue. For example, take bowling facilities. According to Putnam, bowling used to be a highly social activity. But over time, that eroded, and many people now bowl alone. Similarly, consider golf facilities. They used to be primarily clubs, which required membership, often at a high fee. Players would coordinate with other members by first booking a tee time, and then socialize during play or in the club house. These days, golf facilities enable one to go alone, without any membership, and strike golf balls at the driving range or to play the course without reserving a spot with multiple people in advance. In fact, survey data show a rise of the "solo golfer" phenomenon, which is driven by younger generations (see `https://www.lightspeedhq.com/blog/golf-industry-trends/`, accessed November 13, 2025). Similar to fitness gyms, many view solo golfing as an activity to achieve self-care. This view is most prevalent among younger generations; 51 percent of Gen Z respondents report this, and 39 percent among Millennials. Finally, consider bars and restaurants. Obviously, one can book a table alone at a restaurant. Yet, it seems to us that people overwhelmingly go to bars or restaurants primarily for face-to-face socialization, especially with people they know.

spend more time on social media and meet with their friends and families offline less often. These forces may, in turn, have important unintended impacts on local social cohesion more broadly. Next, we investigate this possibility along a number of dimensions for which the data exists.

4.4 Online Homophily and Local Social Capital

In this subsection, we investigate the effect of online homophily on local social capital, as defined and measured by Chetty et al. (2022a,b). We look at the economic connectedness at the county level, which reflects connections between individuals of low and high socio-economic status. We focus on this measure as it has been shown to be the measure of social capital that is the most predictive of economic mobility.

The results presented in Table 5 indicate that the effect of the 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 1% level. The magnitude implies that a one standard deviation increase in the Gmail Homophily Shock reduces economic connectedness by approximately 17% of a standard deviation. Figure A.10 shows that the outcomes are similar if we use alternative measures of economic connectedness from Chetty et al. (2022a,b).[21]

As in the case of other outcome variables, we report the coefficients for Gmail Homophily Shock computed before, during, and after the treatment period (see Figure 6d). The pre-coefficient is not statistically significant and has a different sign from the negative and significant coefficient in the treatment period. Overall, the results 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 5 show that an increase in online homophily had a negative effect on local social capital by reducing economic connectedness, i.e., the probability that the rich and the poor in a county form connections with each other.

[Table 5 about here.]

4.5 Online Homophily and Political Opinions

4.5.1 Hypotheses

Online homophily can affect political preferences in at least two ways. First, exposure to like-minded communities can lead to polarization of opinions. If an average voter gets into more and more

[21] 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. In contrast, as we show below in our analysis, the share of within-county connections does not seem to be significantly correlated with Gmail Homophily Shock. Thus, we believe that the method of construction of the economic connectedness variable does not change the interpretation of our results.

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. Second, exposure to like-minded communities can crowd out within-county political communications and, thus, weaken the role of contextual factors determining political outcomes(Cantoni and Pons, 2022). As a result, people can vote less similarly to their neighbors, which implies divergence of local political preferences.

These two classes of theories make different empirical predictions, and we can use our data to tell them apart.

4.5.2 Online Homophily and Political Homogeneity

To reflect the extent of divergence/convergence of political preferences within counties, we construct a measure of political homogeneity at the county level. This measure captures the probability that two randomly picked people from a county would vote for the same party, or the degree of local political homogeneity (consensus), i.e., the opposite of local political fractionalization, and is defined as

$$PolHomogeneity_{it} = 1 - 2r_{it}(1 - r_{it}) \tag{5}$$

Here r_{it} is the vote share of the Republican party in county i in Presidential or House elections at time t. Figure A.9 shows how our measure is related to vote shares.

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. The results indicate that once we include baseline controls, there is a sizeable drop in political homogeneity as a result of the Gmail Homophily Shock. The magnitude of the coefficient ranges from -0.024 to -0.049 across the 5 specifications in columns (2)-(5), with all the results being significant at the 1% level. In terms of the magnitude of the effect, the results for the most extensive set of controls in column 6 indicate that a one standard deviation increase in the Gmail Homophily Shock lowers political homogeneity by 24% of a standard deviation.

[Table 6 about here.]

Electoral results are the only outcomes in our analysis for which we have data for different periods, including the periods before the creation of Facebook, which allows for estimating placebo regressions. Figure 8 illustrates 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. The effect of the 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 low degree of polarization (and politicization) of social media at that time. From 2016 onward, we observe a jump in the point estimates, indicating a significant reduction in political homogeneity as a result of an increase in online homophily. 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. We find similar results if we look at the outcomes of congressional elections (see Figure A.11 in the Online Appendix).

Figure 6e further illustrates our identification by looking into the relationship between Political Homogeneity and Gmail Homophily Shock, computed during the treatment period, in the pre- and post-treatment periods. The resulting figure supports the assumption that our identification comes 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 error. The coefficient in the after-period is negative, but much smaller numerically. Overall, Figure 6e is in line with similar tests for other outcomes.

[Figure 8 about here.]

4.5.3 Online Homophily and Political Preferences Within Counties

In the previous subsection, we showed that the shock in online homophily caused by the Gmail-Facebook incident decreased political homogeneity between US counties. In this subsection, we examine the effect of online homophily on the distribution of political opinions within counties. Exploiting the precinct-level data for 2016 from (Kaplan et al., 2022), we characterize different moments of the distribution of voting shares at the county level. In particular, we construct several outcomes for measures of dispersion of political opinions across precincts in a county, including standard deviations, inter-quartile ranges, and overall range. We calculate these measures for both the Republican vote share and the measure of political homogeneity described above. We also examine the prevalence of extreme voting margins, ranging from 30 to 70 percent.

The results presented in Figure 9 indicate that there is no consistent effect of online homophily on the measures of dispersion of political opinions across precincts. However, we do see a consistently negative effect on the likelihood of observing extreme voting margins of 50 percent or more.

[Figure 9 about here.]

We check whether the variation that generates the negative effect on the vote margin indeed comes from our treatment period, focusing on the margin of 70% as an example. Figure 6f summarizes these results. The coefficients for pre- and treatment periods have different sizes. The pre-treatment coefficient is negative, but very small and far from being statistically significant, while the coefficient for the treatment period is negative and significant. All graphs in Figure 6 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 specifications). There is some evidence of persistence in post-treatment coefficients, but they are, nevertheless, much smaller than the treatment coefficients in all the specifications from (a) to (f).

4.5.4 Online Homophily and Intensity of Political Preferences

So far, this analysis has examined the dispersion or convergence of political preferences but not their intensity. However, the effect of online homophily on the intensity of political preferences may have important implications, as it speaks more directly to the effect of social media on political polarization. To look at the intensive margin of political preferences, we use the data from the Cooperative Election Study (CES). More specifically, we create a variable $Extreme\_id_{it}$ which denotes respondents who defined themselves as either strong Democrats or strong Republicans.[22] We create a similar measure for extreme ideology for the respondents who defined themselves as either strongly liberal or strongly conservative.

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

$$Extreme\_id_{it} = \beta_0 + \beta_1 GH\_shock_i + \beta_2 GH\_shock_i \times post_t + \beta_3 X_i + \delta_t + \epsilon_{it} \tag{6}$$

The results in Table 7 show that an increase in online homophily leads to a decrease in the share of people with extreme partisan preferences (the results in Table A.14 show a similar effect for extreme ideological positions). In light of the existing results that show that the presence of mobile internet and social media can increase political polarization (Allcott et al., 2020; Levy, 2021; Melnikov, 2022), it is important to note that we are looking at the effect of online homophily rather than the exposure to social media, and that one of the important effects of increasing online homophily is a decrease in offline interaction at the local level. To the extent that interpersonal interactions are more segregated than internet connections (Gentzkow and Shapiro, 2011), the polarizing effect of offline communications may be even stronger than the polarizing effect of social

[22] 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.

media exposure. [23]

[Table 7 about here.]

4.5.5 Effect of Online Homophily on Vote Shares and Turnout

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 the Democratic party. We do not find similar evidence: Table A.15 documents no significant effects for voting for Republicans once proper controls are incorporated. In the most saturated specification, the magnitude of the effect is 0.001, and we can rule out the effects of up to 0.57%, with a mean of the dependent variable being 66.4%; thus our results are close to being precisely estimated zeroes.

Similarly, we do not find any significant evidence for turnout. However, the results, reported in Table A.16, are noisy and cannot be interpreted as precisely estimated zeroes; thus, we cannot provide definite conclusions about the effect of online homophily on turnout.

Overall, our results on the effect of online homophily on political outcomes are consistent with the second hypothesis that we outline: exposure to like-minded communities on Facebook increases the dispersion of voting and leads to the divergence of political preferences within local communities, leading to a reduction in extreme margins of voting and extreme political preferences.

4.6 Gmail Homophily Shock and Total Connections

One important question is whether the Gmail Homophily Shock variable is a shock to homophily or a shock to the total number of friends. In Table A.11, we present 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 the 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%. In addition, if we include the number of connections as a control variable in the regressions that examine the effect of homophily, the results remain virtually unchanged (see Table A.13 in the Online Appendix). Thus, the results indicate that the Gmail Homophily Shock was a shock to online homophily rather than a shock to the number of connections.

[23] Similarly, the results can be driven by interactions on Facebook crowding out interactions on other, more polarizing social media platforms.

4.7 Gmail Homophily Shock and Long Ties

One potential positive effect of social media is that they help create “long ties”, i.e., the ties that connect individuals who do not share any mutual contact. Such long ties have been associated with important positive outcomes, such as measures of economic opportunity (Jahani et al., 2023). To test how online homophily is affecting long ties, we exploit the baseline measure constructed in Jahani et al. (2023).

The results in Table A.18 indicate that our Gmail Homophily Shock leads to a reduction, not an increase, in the fraction of long ties. Once we take into account our baseline set of controls (columns 2 to 6), the magnitude of the effect is stable across specifications, with our most saturated one showing a reduction of about 16% of a standard deviation, given a one standard deviation increase in our Gmail Homophily Shock. These results rule out the hypothesis that online homophily on Facebook has positive effects through the creation of long ties.

4.8 Heterogeneity of Effects

By construction, our main variable of interest, Gmail Homophily Shock, focuses only on the out-of-county connections that people have on Facebook. 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 start by showing that the share of links outside the county itself is not significantly related to our Gmail Homophily Shock. These results are presented in Table A.17. 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 56.1%. Thus, even though in principle the share of outside connections could be affected by the Gmail Homophily Shock, that does not happen in practice.

We then proceed by looking at the heterogeneity of the effect of online homophily with respect to the share of connections outside the county for all the major outcomes of our interest. In all the specifications, the interaction term with the share of outside links has the same sign as the non-interacted coefficient and is significant at the 1% level in all specifications except for the visits to bars, where it is significant at 10% level (see Table 8). These results are consistent with the intuition that the effect of online homophily is stronger in places with a higher share of outside links.

[Table 8 about here.]

We also look at the heterogeneity of effects with respect to the share of urban population. In rural 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 the results to be stronger in urban areas, and this is exactly what we find in the data (see the results in Table A.19 in the Online Appendix). With the exception of offline interactions, all interaction terms have the same sign of the main effect, and five out of six interaction coefficients are significant at the 1% level.

Overall, the results in Tables 8 and A.19 are consistent with the idea that the effects are stronger in places with more out-of-county connections and places where out-of-county connections could be formed more easily.

5 Discussion and Conclusion

In this paper, we examine what happens to communities whose members are exposed to more like-minded connections through their online networks. We exploit a change in friendship recommendations by Facebook that resulted from a conflict about data sharing between Gmail and Facebook in 2010-2012 to construct exogenous variation in the degree of online homophily between different counties in the US. We document that a temporary change in friendship recommendations had a lasting effect on the structure of the friendship network. The incident inadvertently hindered friending people from some communities, which could be communities with either similar or distinct characteristics. We use the resulting exogenous variation generated by the Facebook-Gmail incident to estimate the causal effect of online homophily at the county level. We documented that online homophily substantially affected the cohesion of American communities in several ways. First, higher online homophily pushed individuals to spend more time on Facebook. This happens partly at the expense of other social media platforms, but increases the overall usage of social media. Second, higher online homophily decreases interpersonal contact, as proxied by visits to bars, restaurants, and other locations where people socialize. Third, it leads to a reduction in local social capital. Overall, the results suggest that homophilic online networks contribute to a decline in offline socialization and local cohesion.

The impact of online homophily in the political arena is more nuanced. In terms of the distribution of votes within localities, it mirrors the drop in social cohesion as it leads to higher dispersion of political preferences within counties; two random people are less likely to agree on which party is preferred. Thus, it weakens the role of place-based factors (Cantoni and Pons, 2022) in determining voting behavior.

At the same time, in terms of its effect on political preferences at the individual level, we find that online homophily leads to a reduction in the prevalence of extreme political positions. Thus, although in general social media usage may increase political polarization (Allcott et al., 2020; Levy, 2021; Melnikov, 2022), the homophily of online network connections seems to attenuate this effect.

References

Allcott, H., L. Braghieri, S. Eichmeyer, and M. Gentzkow (2020). 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.

Anders, J. and A. Pallais (2025). Inefficient social disconnection. *Working Paper*.

Asimovic, N., J. Nagler, R. Bonneau, and J. A. Tucker (2021). Testing the effects of facebook usage in an ethnically polarized setting. *Proceedings of the National Academy of Sciences 118*(25), e2022819118.

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.

Barrera, O., S. Guriev, E. Henry, and E. Zhuravskaya (2020). Facts, Alternative Facts, and Fact Checking in Times of Crisis. *American Economic Review 110*(10), 3256–3275.

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. *American Economic Review 112*(11), 3660–93.

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. *Working Paper*.

Bursztyn, L., B. Handel, R. Jiménez-Durán, and C. Roth (2025). When product markets become collective traps: The case of social media. *American Economic Review 115*(12), 4105–4136.

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

Campante, F., R. Durante, and F. Sobbrio (2017). 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.

Cantoni, E. and V. Pons (2022). Does Context Outweigh Individual Characteristics In Driving Voting Behavior? Evidence From Relocations Within The United States. *American Economic Review 112*(4), 1226–1272.

Carrell, S. E., R. L. Fullerton, and J. E. West (2009). Does your cohort matter? measuring peer effects in college achievement. *Journal of Labor Economics 27*(3), 439–464.

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). 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. *Working paper*.

Falck, O., R. Gold, and S. Heblich (2014). 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. Müller, and C. Schwarz (2023). 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). 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). 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 (2025). Curtailing False News, Amplifying Truth. *Working Paper*.

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). Checking And Sharing Alt-Facts. *American Economic Journal: Economic Policy 14*(3).

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press.

Jackson, M. O., S. M. Nei, E. Snowberg, and L. Yariv (2023). The dynamics of networks and homophily.

Jahani, E., S. P. Fraiberger, M. Bailey, and D. Eckles (2023). Long Ties, Disruptive Life Events, And Economic Prosperity. *Proceedings Of The National Academy Of Sciences 120*(28), e2211062120.

Kannan, V. D. and P. J. Veazie (2023). Us trends in social isolation, social engagement, and companionship - nationally and by age, sex, race/ethnicity, family income, and work hours, 2003–2020. *SSM-population health 21*, 101331.

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. *Working paper*.

Lazarsfeld, P. F., R. K. Merton, et al. (1954). Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society 18*(1), 18–66.

Leip, D. (1993). Dave Leip's Atlas of U.S. Presidential Elections. `https://uselectionatlas.org/`.

Levy, R. (2021). 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].

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology 27*(1), 415–444.

Melnikov, N. (2022). Mobile Internet and Political Polarization. Technical report.

Müller, K. and C. Schwarz (2020). 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). 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.

Olken, B. (2009). Do television and radio destroy social capital? evidence from indonesian villages. *American Economic Journal: Applied Economics 1*(4), 1–33.

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.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics 116*(2), 681–704.

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.

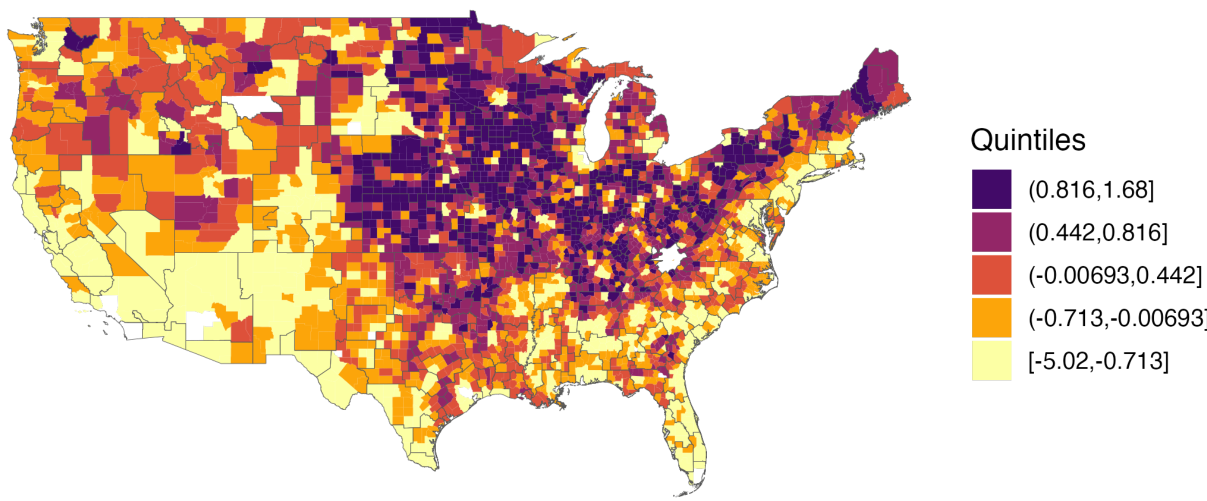
Tarbush, B. and A. Teytelboym (2012). Homophily In Online Social Networks. In P. W. Goldberg (Ed.), *Internet And Network Economics*, Berlin, Heidelberg, pp. 512–518. Springer Berlin Heidelberg.

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.

Waldinger, F. (2012). Peer effects in science: Evidence from the dismissal of scientists in nazi germany. *The Review of Economic Studies 79*(2), 838–861.

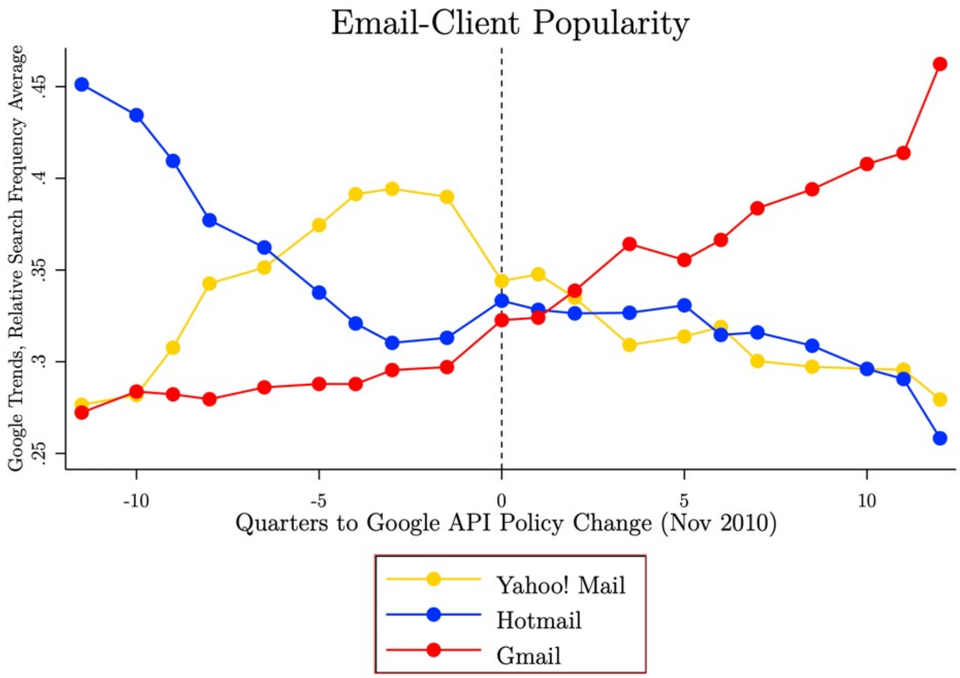
Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). Political Effects of the Internet and Social Media. *Annual Review of Economics 12*(1), 415–438.

Figure 1: Online Homophily, 2016



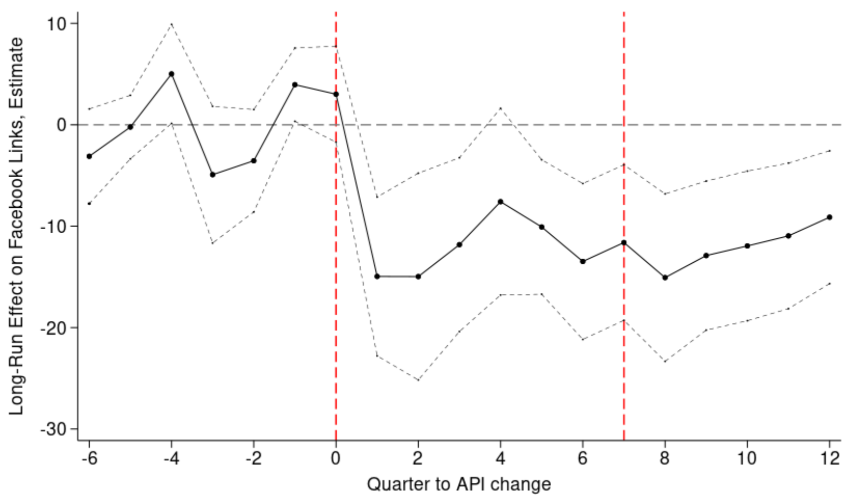
Notes: The map plots the geographic distribution of Online Homophily in 2016 across US counties. We construct Online Homophily in two steps. First, we generate an index of social similarity between US counties by taking the inverse of the principal component analysis of twelve variables: log population, 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 of self-assessed ideology from Gallup polls between 2008 and 2010 ranging from 1 (very liberal) to 5 (very conservative)). Second, we average the social similarity index at the county level, weighting by 2016 Facebook connections. Finally, we standardize the index, see section 2 for more details.

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



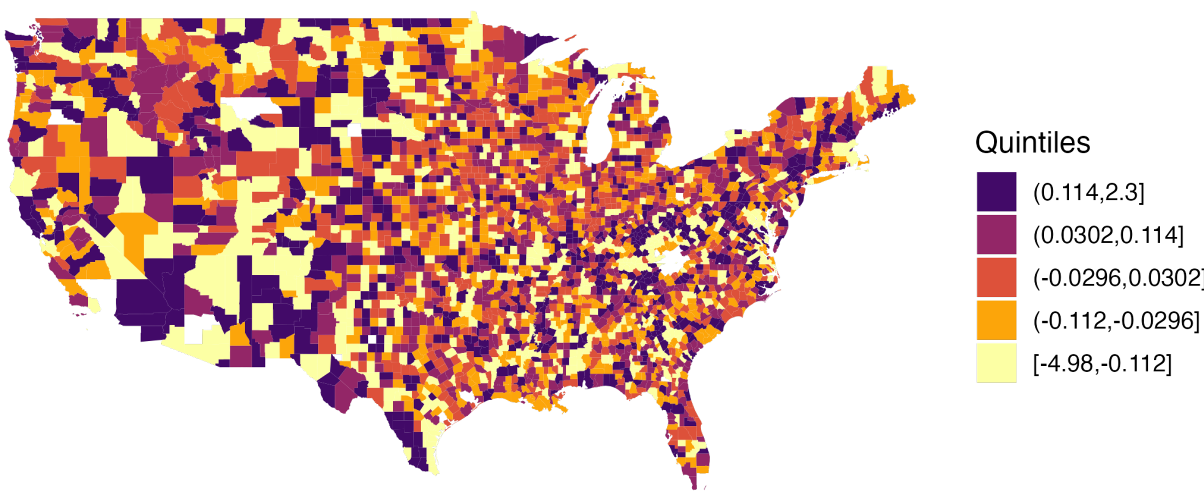
Notes: The figure plots the popularity of the main email clients in the US, quarterly between 2006 and 2016. The source of the data is Google Trends, and popularity is measured as the average search frequency of a given email client in a DMA. We focus on the three largest email clients: Yahoo! Mail,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

Figure 3: Facebook Links by Gmail Complementarity, 2009-2013



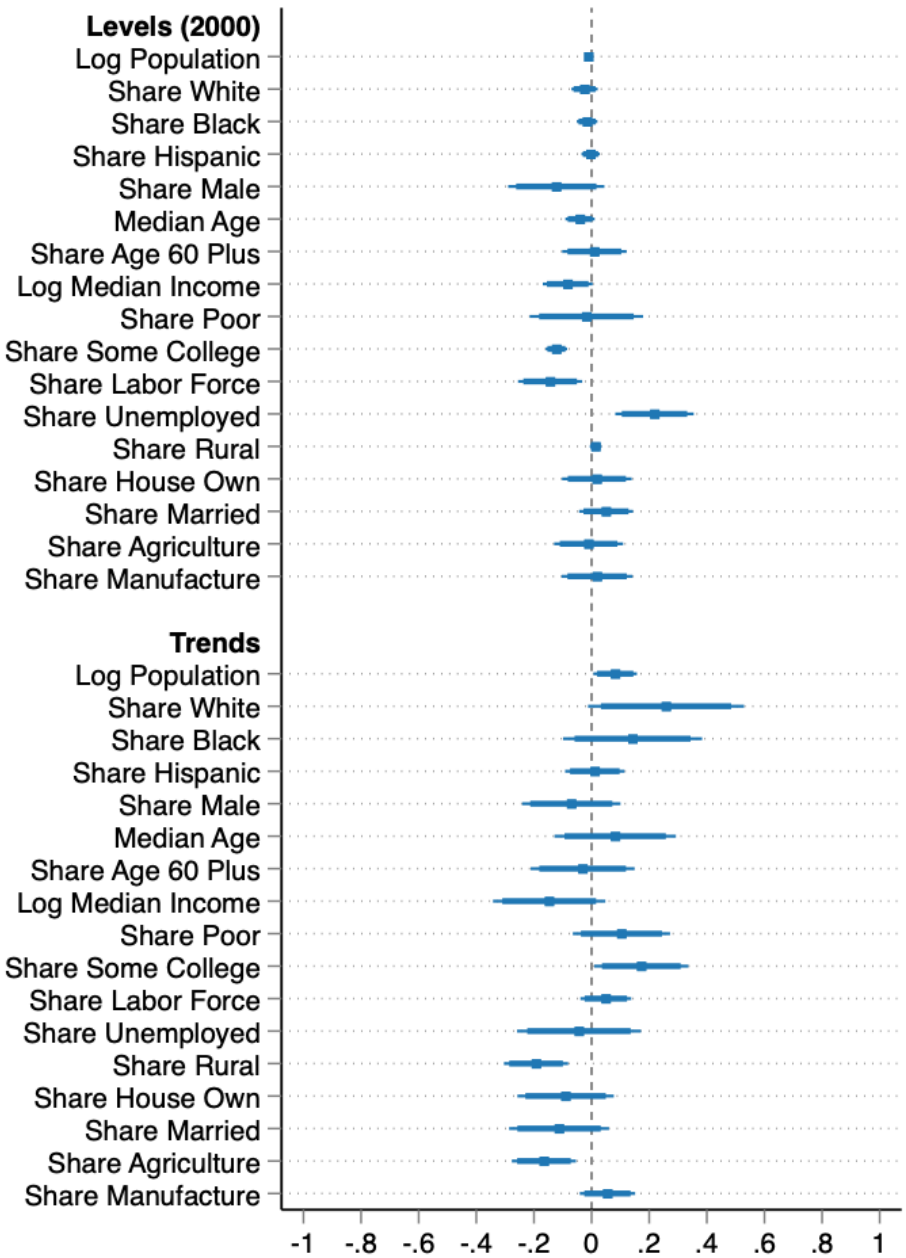
Notes: The graph plots the effect of Gmail complementarity by quarter on 2016 county-pairs Facebook links. We regress and plot the estimates of separate linear models where the outcome variable is the inverse hyperbolic sine of the relative friendship index between county pairs in 2016. Each estimated 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 Homophily Shock



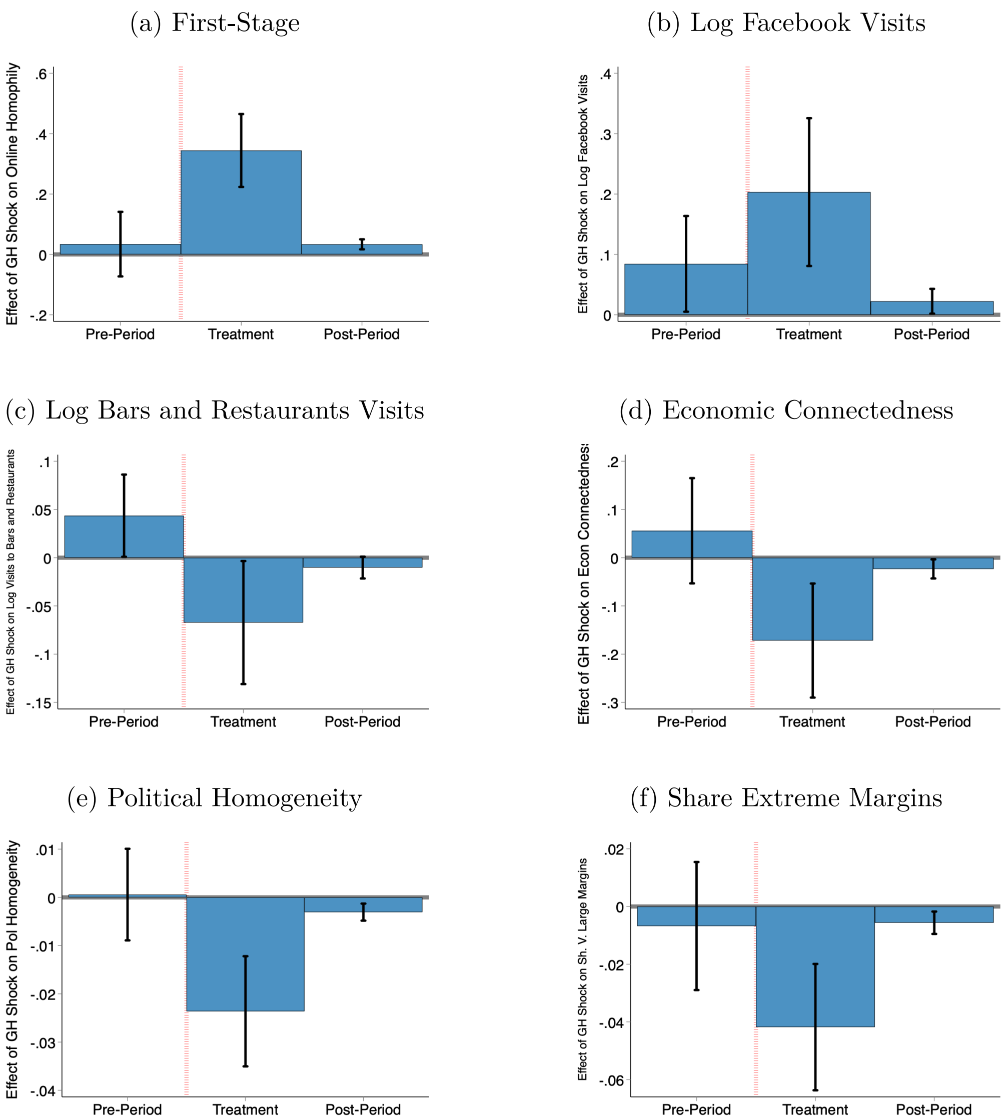
Notes: The map plots the geographic distribution across US counties of the Gmail Homophily Shock residualized on our full set of controls and fixed effects. Gmail Homophily Shock measures the differential Gmail complementarity in the six quarters following the API change. The controls we use to residualize 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; socio-demographic trends defined as the difference for all controls between 2000 and 2010. We cluster standard errors at the state level.

Figure 5: Balance Tests of the Gmail Homophily Shock



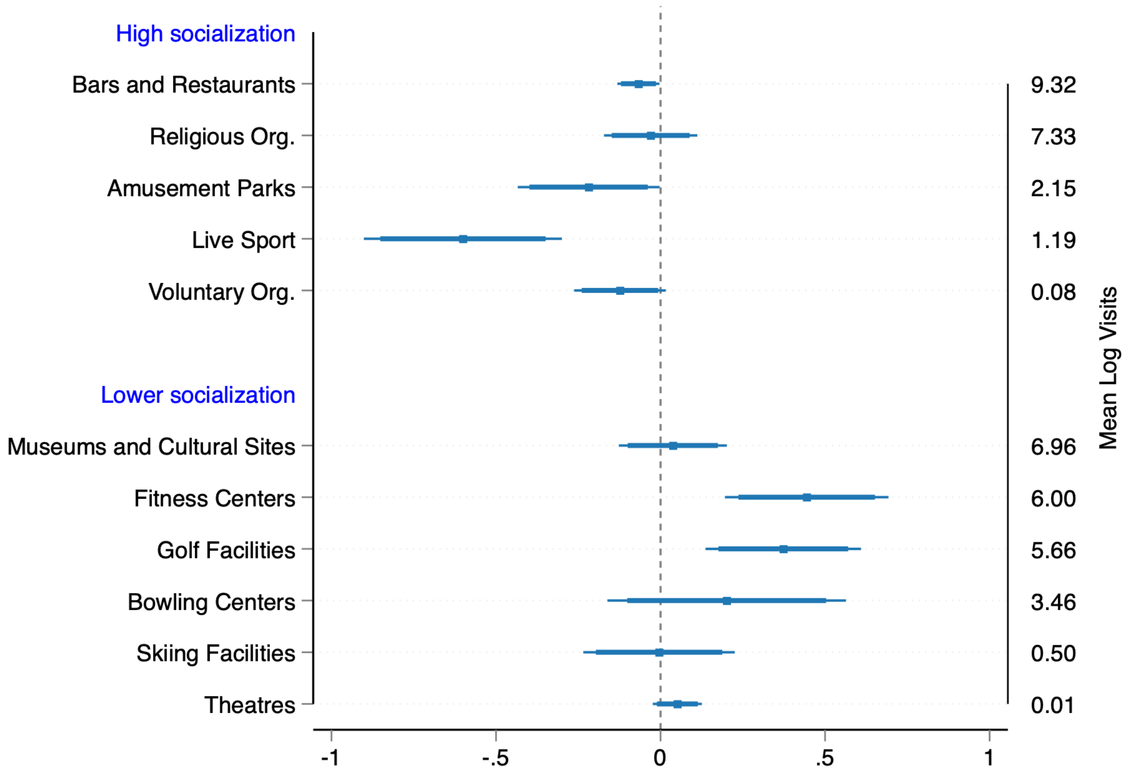
Notes: Balance tests of our Gmail Homophily Shock. We plot the estimated coefficients from separate regressions of our baseline model in equation (4), where we test the balancedness of our Gmail Homophily Shock on a host of predetermined socio-economic and political county-level controls shown on the Y-axis. Gmail Homophily Shock measures the differential Gmail complementarity in the six quarters following the API change. When looking at levels we use a specification that includes DMA fixed effects and all our controls in levels as of 2010: share White, share Black, share Hispanic, log median income, share in labor force, share rural, median age, share with at least some college, share unemployed; political homogeneity, turnout and Republican vote shares as of 2008; We also control for the pre-period Gmail complementarity using the last six quarters before the API changed. When looking at trends between 1990 and 2000, we use our most saturated specification, including trends between 2000 and 2010. We cluster standard errors at the state level.

Figure 6: Gmail Complementarity Has No Effect Outside Treatment Window



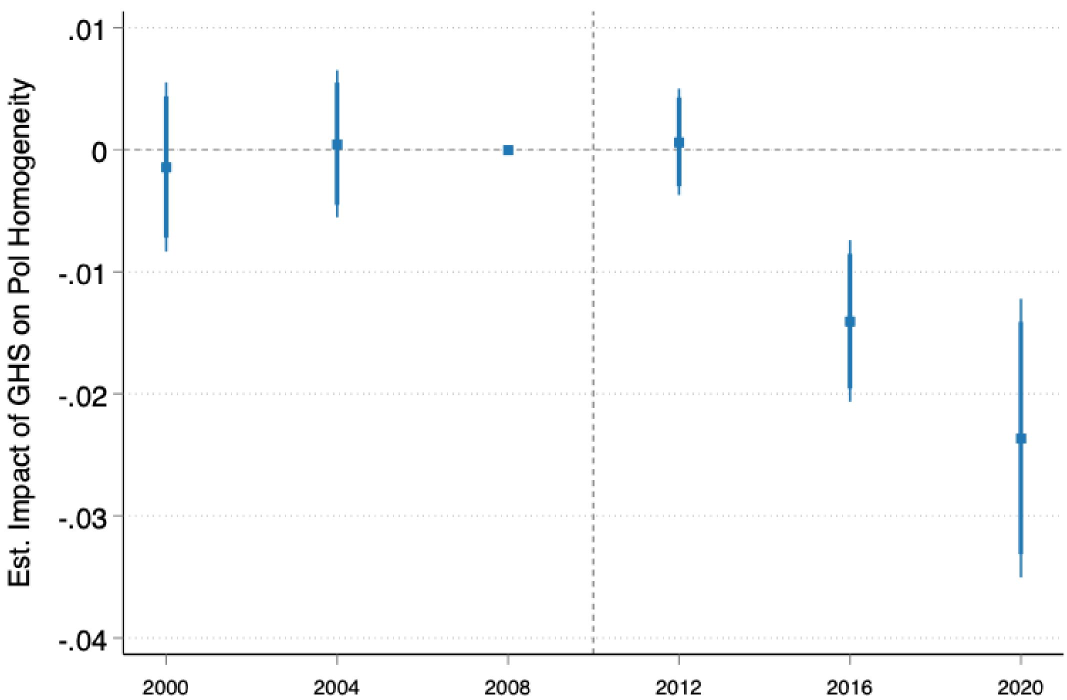
Notes: The figure plots the effect of the cumulative Gmail complementarity on our outcome variables in three different windows: the pre-period, the treatment period, and the post-period. We build the cumulative Gmail complementarity in the pre-period using the six quarters before the API change. The cumulative Gmail complementarity in the treatment window is our Gmail Homophily Shock, which is constructed using the six quarters after the 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.12 displays results using a sparser specification including only DMA FEs and baseline controls (same as in column 3 of Table 1).

Figure 7: The Impact of Homophily Shock on Visits by Venues of Interaction



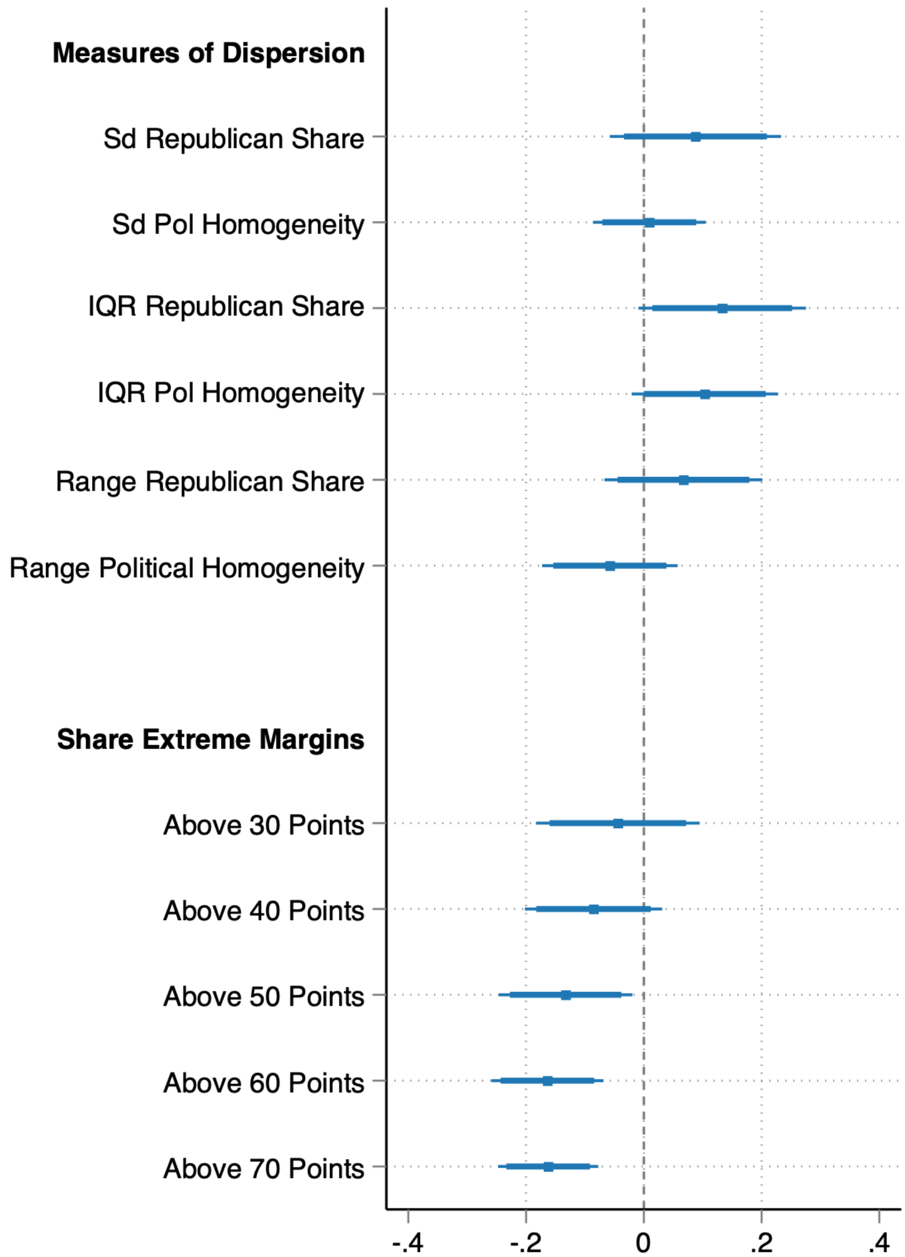
Notes: The figure plots the estimated effect of the Gmail Homophily Shock on the log number of visits for several venues of interaction. The left y-axis shows the venue of interaction, while the right y-axis presents the average number of visits for each venue in logs (Appendix Table A.2 shows means in levels). The source of the data is Safegraph and we reconstruct the number of visits for the following venues of interaction using the relative NAICS codes reported in parenthesis: Bars (NAICS 7224), Restaurants (NAICS 7225), Theatres (NAICS 7111), Live Sport (NAICS 7112), Museums and Historical Sites (labeled as Cultural Sites; NAICS 7121), Amusement Parks (NAICS 7131), Golf Facilities (NAICS 713910), Skiing Facilities (NAICS 713920), Fitness Centers (NAICS 713940), Bowling Centers (NAICS 713950), Other Recreational Centers (NAICS 713990), Religious Organizations (NAICS 813110) and Voluntary Associations (NAICS 8132). We plot bars and confidence intervals from our most saturated specification, as in column 6 of Table 4. We cluster standard errors at the state level.

Figure 8: Gmail Homophily Shock and Political Homogeneity



Notes: The figure plots the event study analysis of the impact of Gmail Homophily Shock on political homogeneity. We plot the estimated coefficients associated with the effect of a 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, using 2008 as our reference year. Gmail Homophily Shock measures 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; trends between 2000 and 2010 for all the socio-demographic controls as well as DMA fixed effects. We cluster standard errors at the state level.

Figure 9: Gmail Homophily Shock and Political Preferences Within County



Notes: The figure plots the estimated effect of our Gmail Homophily Shock on within-county dispersion of political preferences. The dependent variables shown on the y-axis are constructed using precinct-level electoral outcomes for 2016 from Kaplan et al. (2022). Both Gmail Homophily Shock and outcome variables are expressed in standard deviations. We plot point estimates and confidence intervals from our most saturated specification which includes the following controls: 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)
GH Shock	0.754***	0.355***	0.424***	0.411***	0.320***	0.308***	0.312***
	(0.198)	(0.085)	(0.084)	(0.095)	(0.076)	(0.079)	(0.078)
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*
Adj R2	0.291	0.597	0.717	0.736	0.759	0.769	0.783
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: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is our Online Homophily, constructed by standardizing the inverse of the social diversity index. We build the social diversity index by taking the average socio-economic distance to all counties in a county network, weighted by Facebook connections. In columns 1 to 6, we use 2016 Facebook connections as weights, whereas in column 7 we use 2020 Facebook connections. Refer to section 2.3 for further details. The Gmail Homophily Shock (GH Shock) measures 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 attended 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 include 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)
GH 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)
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*
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

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, Twitter, and Reddit. The data on web searches comes from ComScore and refers to the first trimester of 2016. Gmail Homophily Shock (GH Shock) measures 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)
GH 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)
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*
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

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, Twitter, and Reddit. The data on web searches comes from ComScore and refers to the first trimester of 2016. Gmail Homophily Shock (GH Shock) measures 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 Bars and Restaurants Visits					
	(1)	(2)	(3)	(4)	(5)	(6)
GH 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)
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*
Month 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.	9.318	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

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the log number of visits to bars and restaurants (NAICS codes 7224 and 7225). The source of the data is SafeGraph, covering all of 2019 and varying at the county-by-month level. Gmail Homophily Shock (GH Shock) measures 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 offline 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)
GH Shock	-0.516***	-0.312***	-0.308***	-0.197***	-0.184***	-0.172***
	(0.098)	(0.056)	(0.058)	(0.058)	(0.063)	(0.059)
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	0.396	0.705	0.816	0.863	0.863	0.872
Observations	2943	2943	2943	2943	2943	2943

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the economic connectedness across income strata for low socioeconomic status individuals sourced from Chetty et al. (2022a). Gmail Homophily Shock (GH Shock) measures 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)
GH Shock	-0.000 (0.012)	-0.038*** (0.010)	-0.036*** (0.010)	-0.049*** (0.011)	-0.024*** (0.006)	-0.024*** (0.006)
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	0.253	0.529	0.620	0.646	0.868	0.876
Observations	3042	3042	3042	3042	3042	3042

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is political homogeneity in 2020 computed as one minus the Herfindahl index applied to the Republican and Democrat vote shares. Gmail Homophily Shock (GH Shock) measures 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: Gmail Complementarity Shock Reduces Extreme Partisan Identity

Dep. Variable:	Extreme Partisanship					
	(1)	(2)	(3)	(4)	(5)	(6)
GH Shock × Post	-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 a strong democrat or a strong republican. The indicator is derived by transforming the answer to a survey question in the Cooperative Election Study (CES) asking respondents to place themselves on a partisanship scale of seven possible alternatives ranging from "strong democrat" to "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.

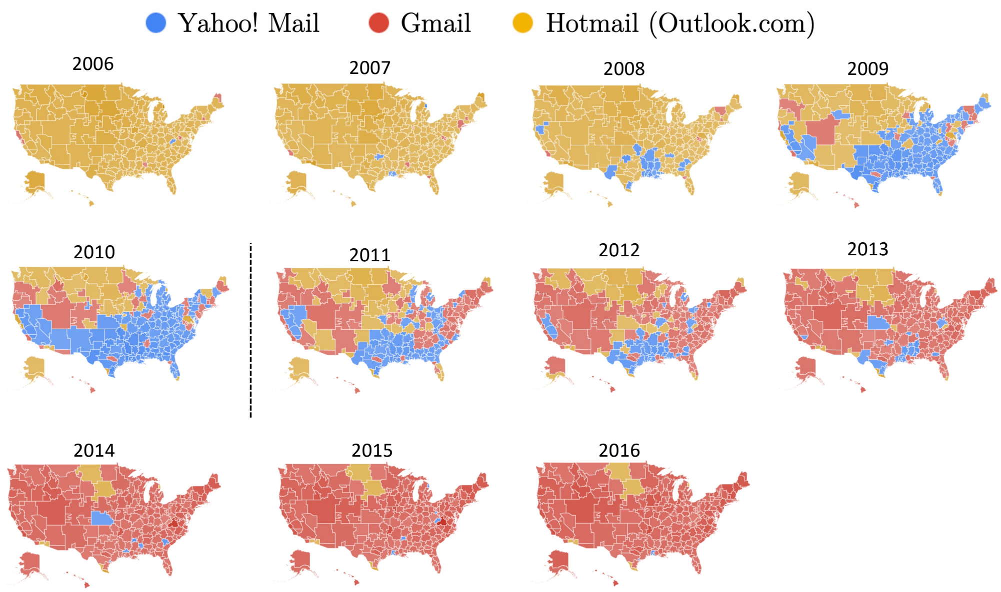
Table 8: The Impact of the Gmail Homophily Shock by Share of Links Outside the County

	Online Homophily	Log Facebook Visits	Log Bars and Rest. Visits	Econ Connect	Political Homogeneity	Extreme Margins
	(1)	(2)	(3)	(4)	(5)	(6)
GH Shock	0.204**	0.084	-0.049	-0.123*	-0.019***	-0.032***
	(0.079)	(0.060)	(0.036)	(0.064)	(0.006)	(0.010)
– × Share Out Connections	0.115***	0.121***	-0.021	-0.049***	-0.005***	-0.011***
	(0.024)	(0.025)	(0.014)	(0.014)	(0.001)	(0.003)
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	0.782	0.869	0.955	0.876	0.880	0.759
Mean of Dep. Var.	0.000	4.853	9.318	0.000	0.605	0.218
Observations	3042	2872	36564	2943	3042	2729

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 connections outside the county. The dependent variables are the six main outcomes of the paper: Online Homophily in 2016 in column 1, Log Facebook visits in column 2, Log Bars and Restaurants visits in column 3, Economic Connectedness in column 4, political homogeneity in column 5 and the share of precincts within a county with electoral margins larger than 70 points in 2016 in column 6. Gmail Homophily Shock (GH Shock) measures 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. Columns 2 and 3 also control for the total number of online and offline visits, respectively. In column 3, we also account for month fixed effects together with DMA fixed effects. Robust standard errors clustered by state in parentheses.

A Online Appendix (Not for publication)

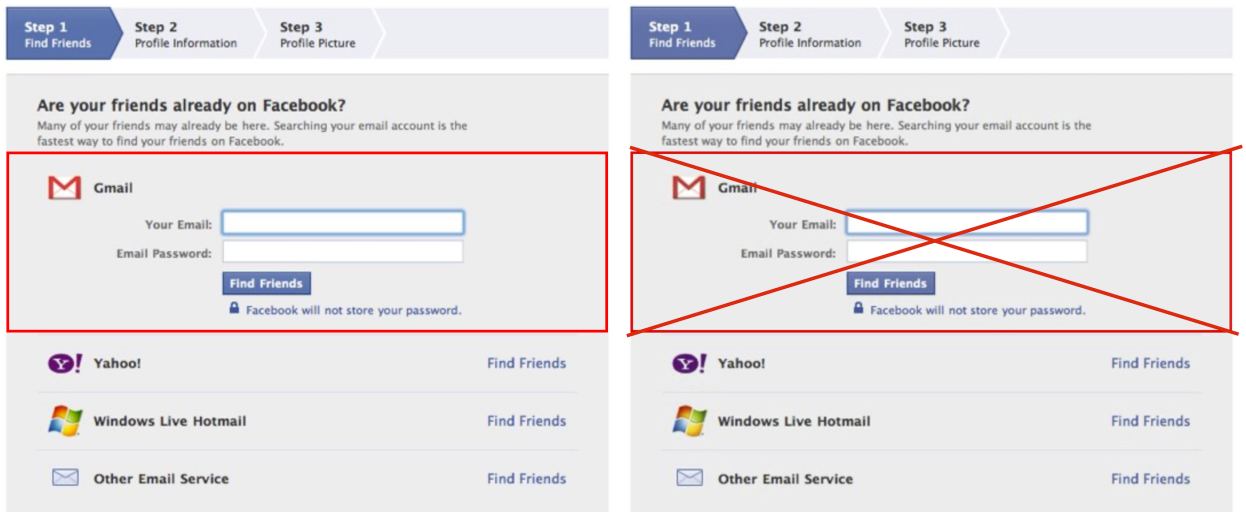
Figure A.1: Geographic Distribution of Email Clients Popularity Across US DMAs



Notes: The map plots the most popular email client between Yahoo!, Gmail and Hotmail (Outlook.com) in each DMA-level between 2006 and 2016. The source of the data is Google Trends where popularity is measured as the average search frequency of a given email client in a DMA.

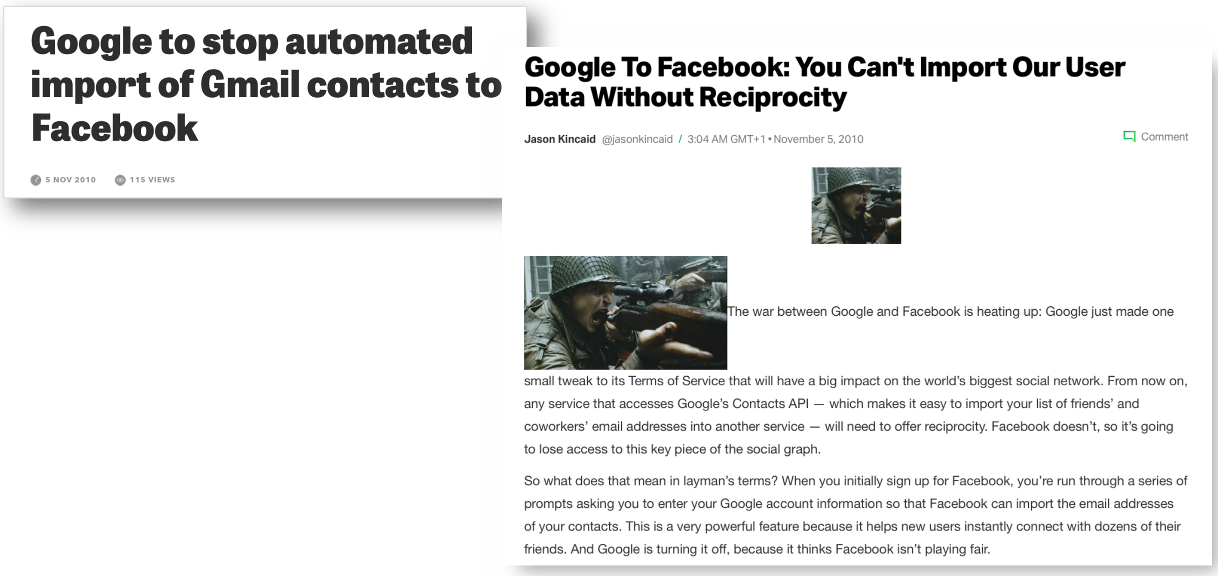
Figure A.2: Exterior Look of Join-Facebook Window, 2009-2013

(a) Before Nov 2010 (b) After Nov 2010



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



Google to stop automated import of Gmail contacts to Facebook

5 NOV 2010 115 VIEWS

Google To Facebook: You Can't Import Our User Data Without Reciprocity

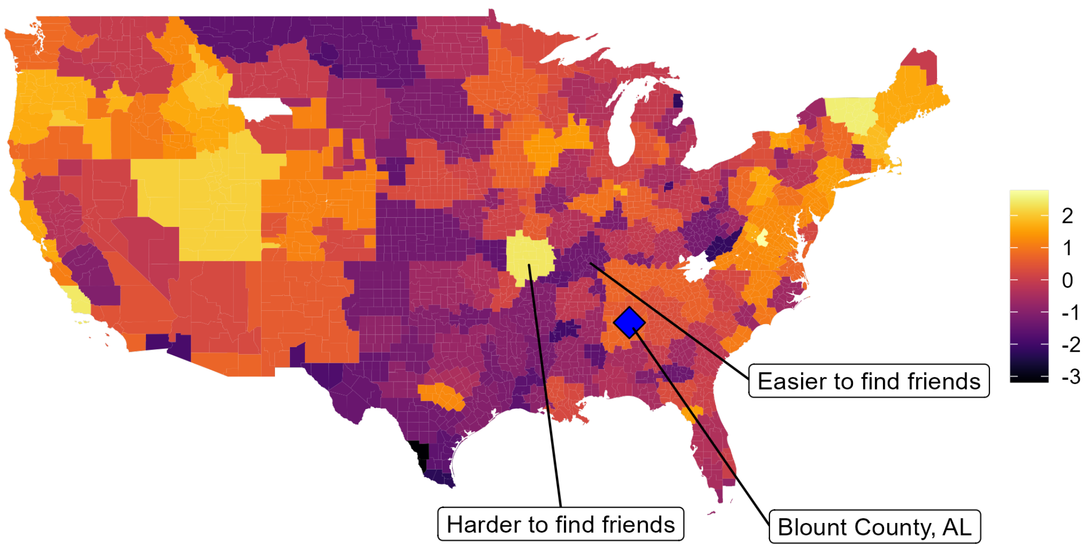
Jason Kincaid @jasonkincaid / 3:04 AM GMT+1 • November 5, 2010 Comment

The war between Google and Facebook is heating up: Google just made one small tweak to its Terms of Service that will have a big impact on the world's biggest social network. From now on, any service that accesses Google's Contacts API — which makes it easy to import your list of friends' and coworkers' email addresses into another service — will need to offer reciprocity. Facebook doesn't, so it's going to lose access to this key piece of the social graph.

So what does that mean in layman's terms? When you initially sign up for Facebook, you're run through a series of prompts asking you to enter your Google account information so that Facebook can import the email addresses of your contacts. This is a very powerful feature because it helps new users instantly connect with dozens of their friends. And Google is turning it off, because it thinks Facebook isn't playing fair.

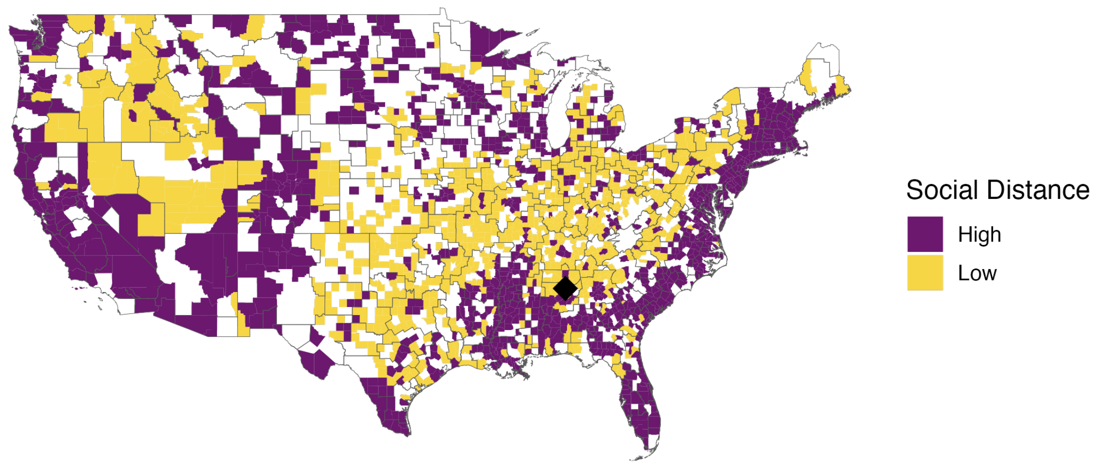
Notes: The figure shows two headlines contemporary to the incident between Google and Facebook giving details on the source of the conflict. The source of the headlines is the Tech blog TechCrunch.com (https://techcrunch.com/2010/11/04/facebook-google-contacts/?guccounter=1)

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



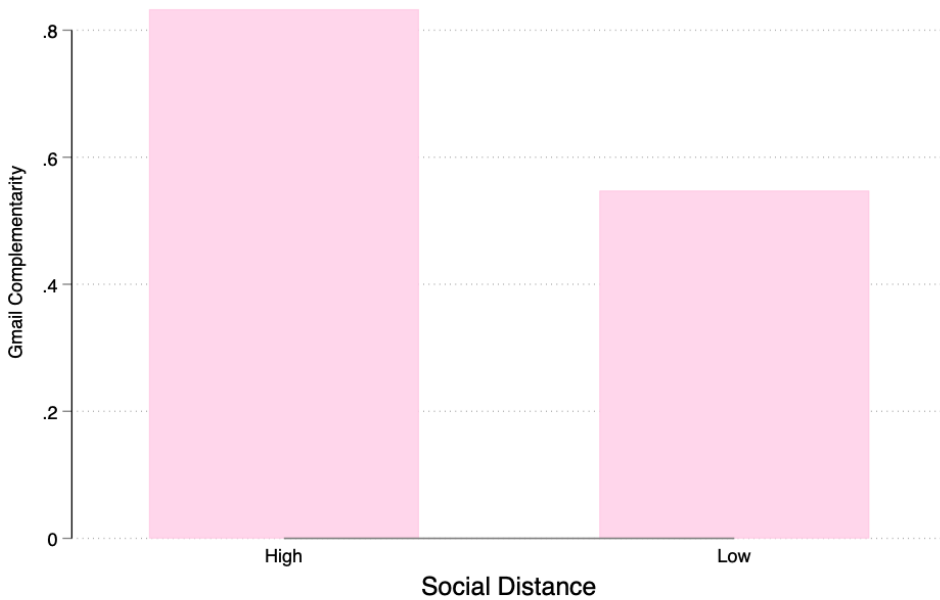
Notes: The map plots the geographic distribution of the cumulative Gmail complementarity in the six quarters post-API change between Blount county, AL, and the rest of the counties in the US. The data source of email client usage is Google Trends and comes at the DMA level. Lighter (darker) color indicates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

Figure A.5: Counties with High and Low Social Distance to Blount County, AL



Notes: The map plots the geographic distribution of high and low social distance counties with respect to Blount County. We compute the distance between US counties using the same twelve characteristics we used to construct our online homophily measure, see subsection 2.2 for more details. We divide US counties into high and low social distance counties splitting the data between the top and bottom terciles of the distribution.

Figure A.6: Gmail Complementarity by High- and Low-Social Distance to Blount County, Alabama

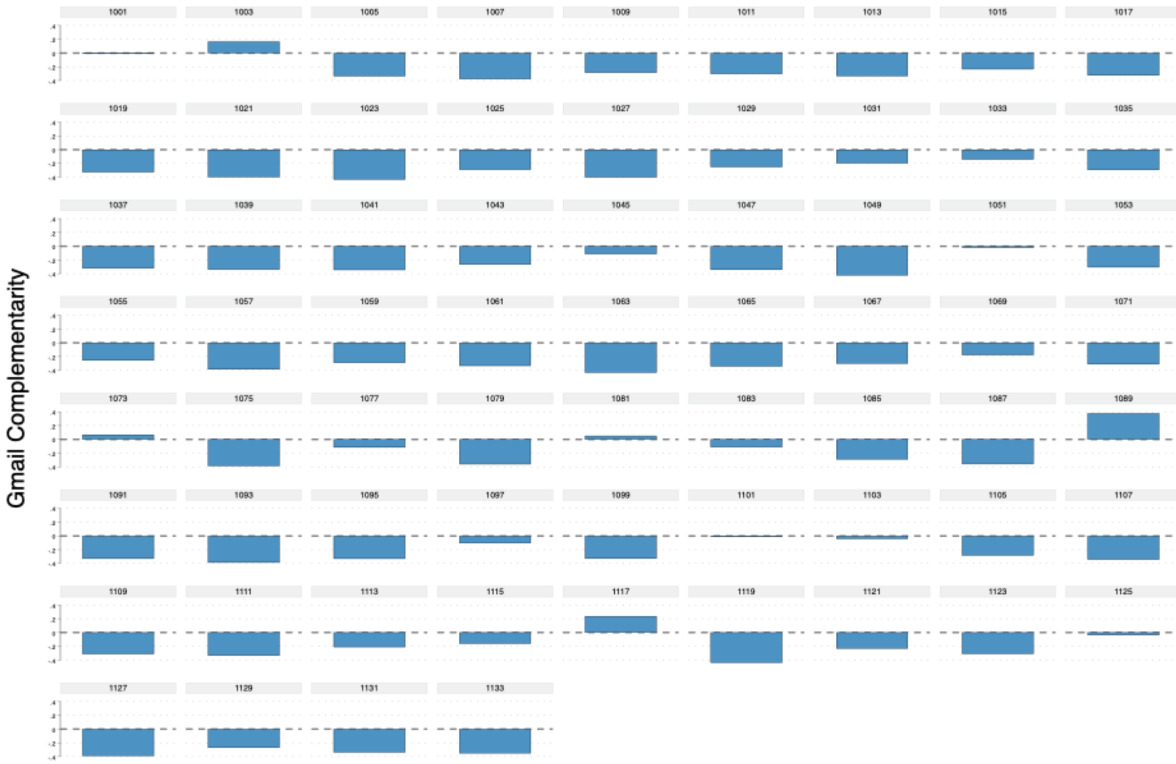


Notes: The figure plots the average Gmail complementarity by social distance for Blount County, Alabama. We denote a county to have high social distance if it belongs to the top tercile of the social distance distribution, whereas we denote the county to have low social distance if it belongs to the 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 2 for more details.

Figure A.7: Gmail Complementarity by High- and Low-Social Distance by County in Alabama

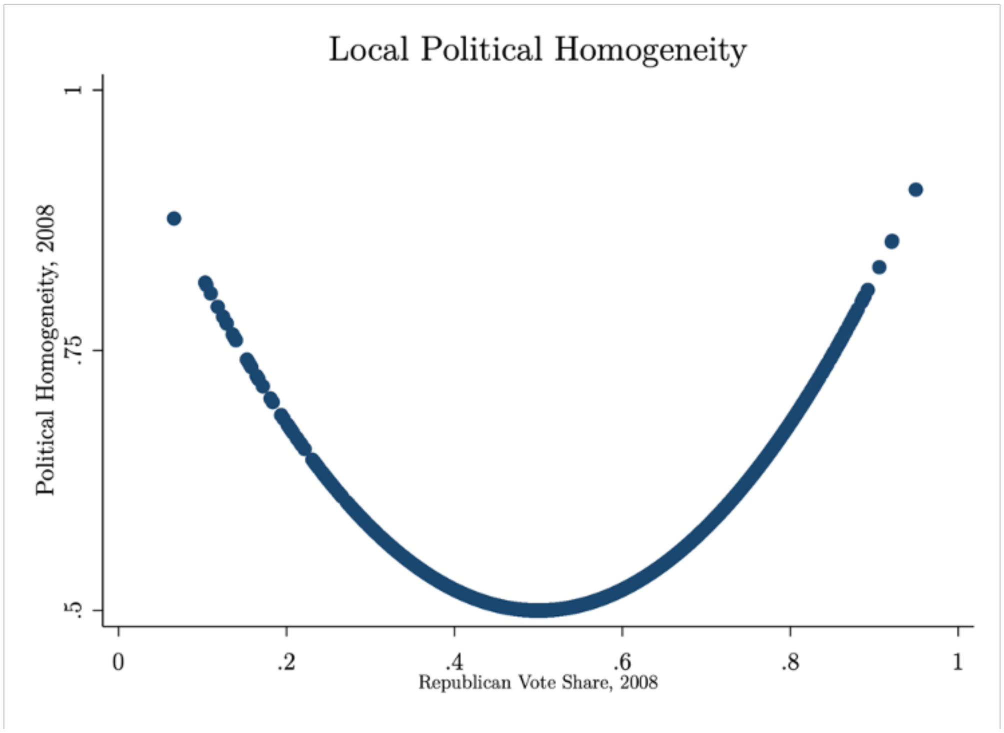
Notes: The figure plots the average Gmail complementarity by social distance for all the counties in Alabama. We denote a county to have high social distance if it belongs to the top tercile of the social distance distribution, whereas we denote the county to have low distance if it belongs to the bottom 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 2 for more details.

Figure A.8: Gmail Homophily Shock by County in Alabama



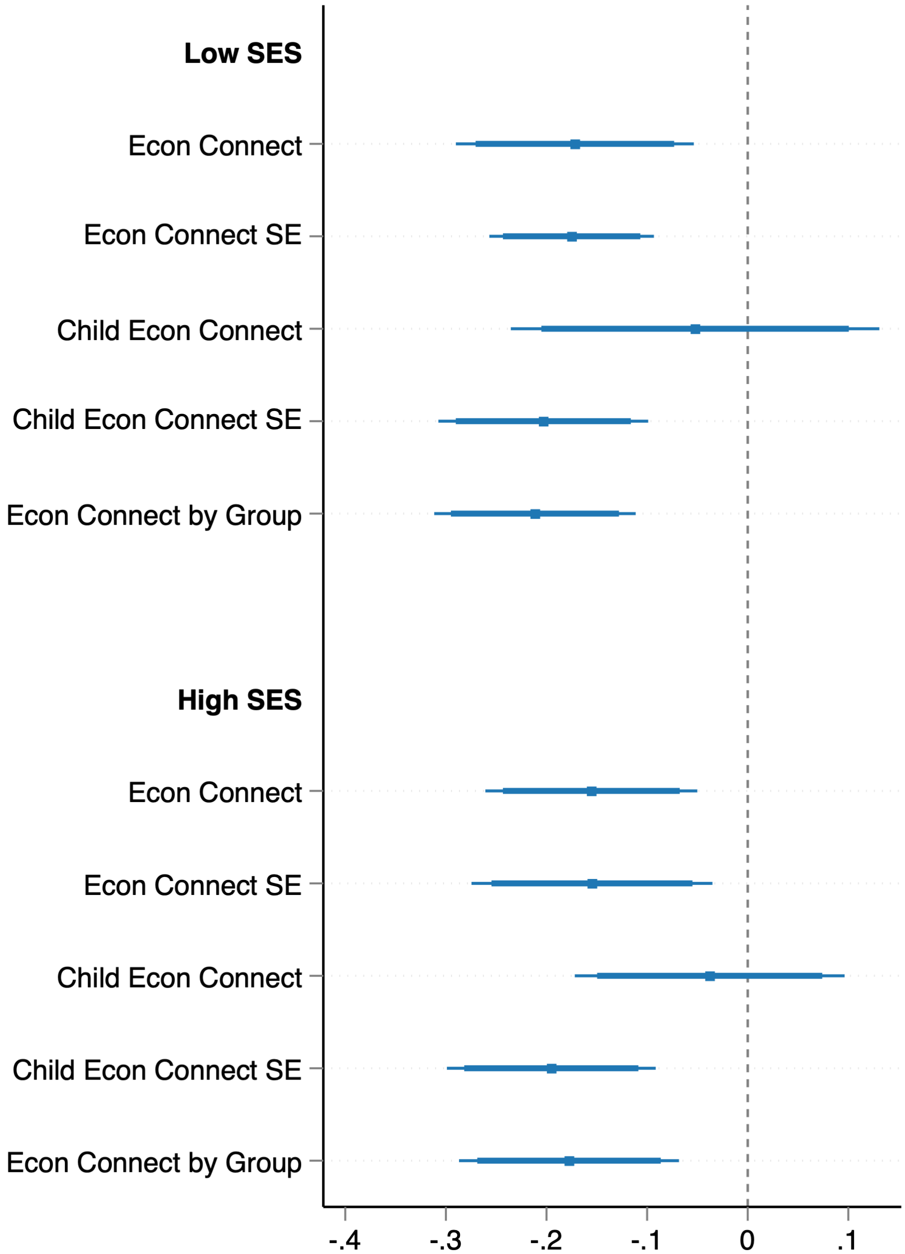
Notes: The figure plots the Gmail Homophily Shock for each county in Alabama. The Gmail Homophily Shock is calculated as the difference in the average Gmail complementarity among high- and low-social distance counties. The average Gmail complementarity by social distance is plotted in Figure A.7 for all the counties in Alabama.

Figure A.9: Political Homogeneity and Vote Shares



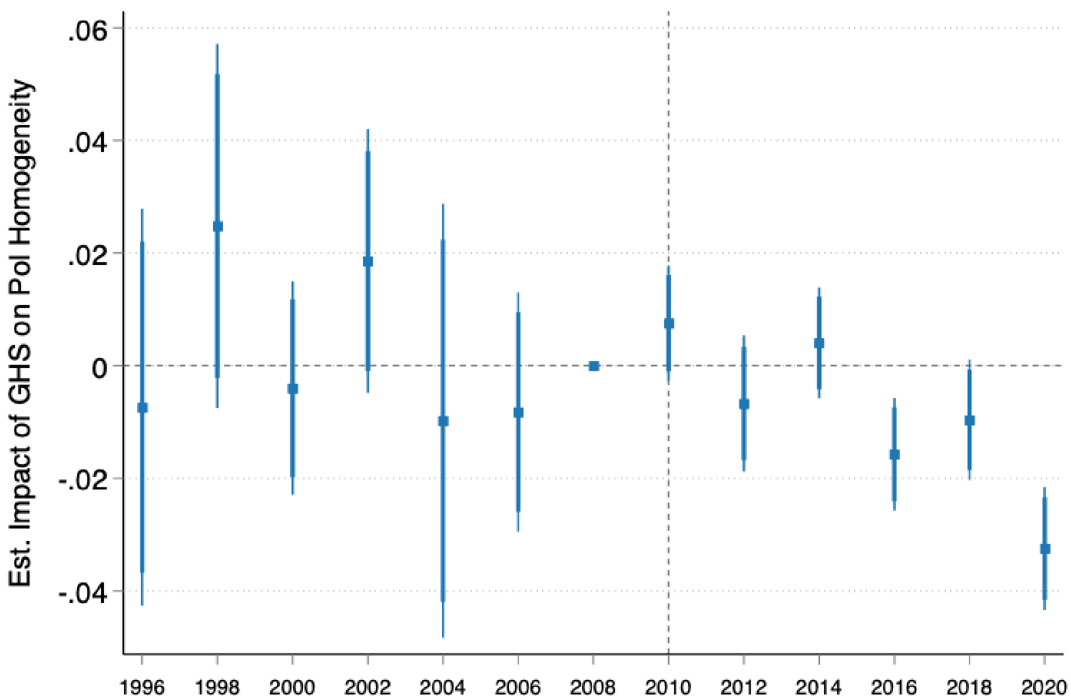
Notes: The figure plots the functional relationship between political homogeneity and the vote share for the Republican party (or any bipartisan electoral system). We construct political homogeneity as the opposite of local political fractionalization, and it is computed as $PolHomogeneity_{it} = 1 - 2r_{it}(1 - r_{it})$, where r_{it} is the Republican vote share at time t in county i.

Figure A.10: Gmail Complementarity Has Similar Impact on Other Economics Connectedness Measures



Notes: The figure plots the impact of our Gmail Homophily Shock on all variables measuring economic connectedness from Chetty et al. (2022a,b). We plot point estimates and confidence intervals from the most saturated specification which includes: log population, share White, share with at least some college, 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; the pre-period Gmail complementarity and DMA fixed effects. We cluster standard errors at the state level.

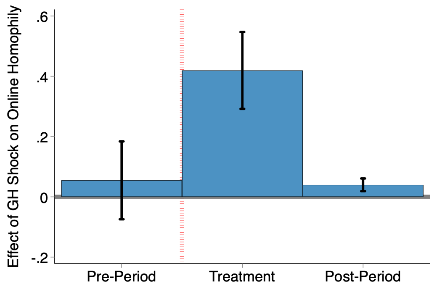
Figure A.11: Gmail Homophily Shock and Political Homogeneity, House Elections



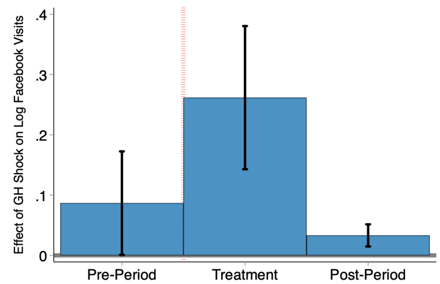
Notes: Event study analysis of the impact of Gmail Homophily Shock on political homogeneity in congressional house elections. We plot the estimated coefficients associated with the effect of one standard deviation increase in the Gmail Homophily Shock on political homogeneity every two years between 1996 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 measures the differential Gmail complementarity in the six quarters following the API change. Controls include: log population, share White, share with at least some college, 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; the pre-period Gmail complementarity and DMA fixed effects. We cluster standard errors at the state level.

Figure A.12: Gmail Complementarity Has No Effect Outside Treatment Window, Sparse Specification

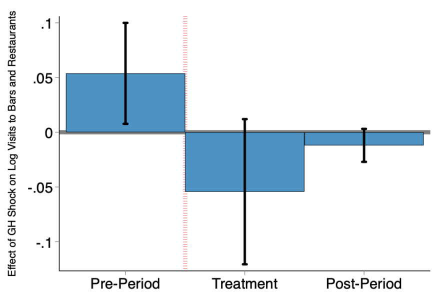
(a) First-Stage



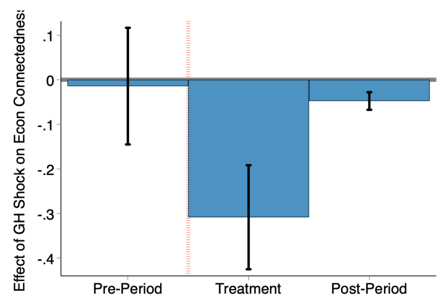
(b) Log Facebook Visits



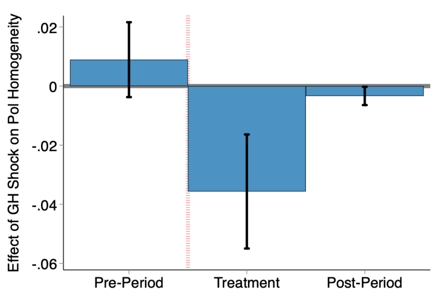
(c) Log Bars and Restaurants Visits



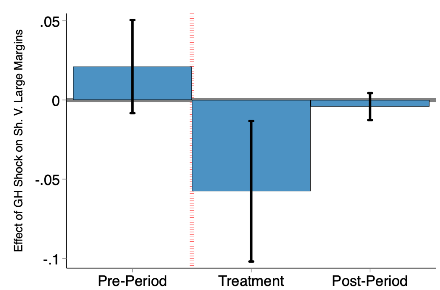
(d) Economic Connectedness



(e) Political Homogeneity



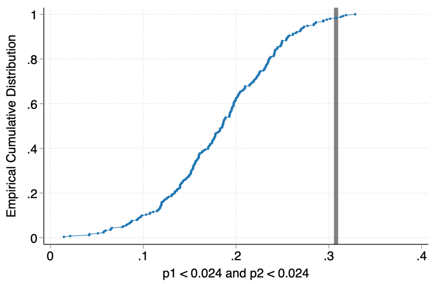
(f) Share Extreme Margins



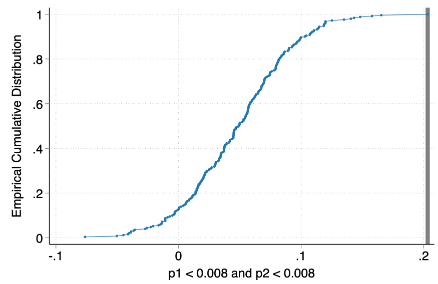
Notes: The figure plots the effect of the cumulative Gmail complementarity on our outcome variables in three different windows: the pre-period, the treatment period and the post-period. We build the cumulative Gmail complementarity in the pre-period using the six quarters prior to the API change. The 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 6 displays results using our most saturated specification (same as in column 6 of Table 1)

Figure A.13: Randomization Analysis, Main Results

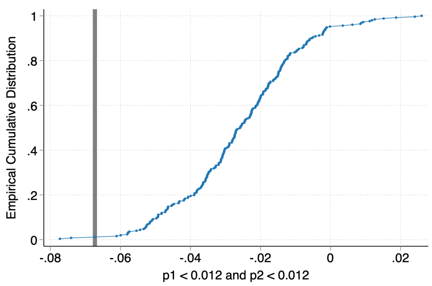
(a) Online Homophily



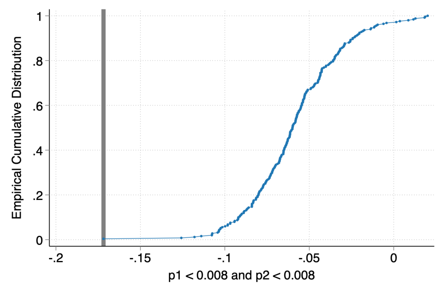
(b) Log Facebook Visits



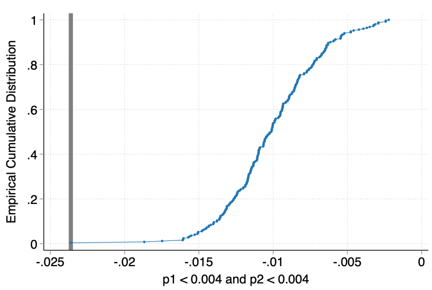
(c) Log Bars and Restaurants Visits



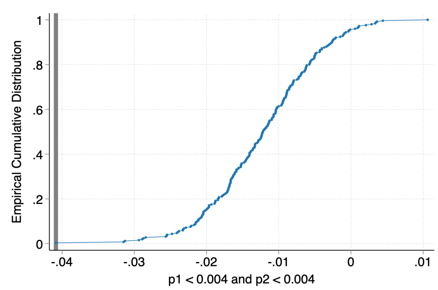
(d) Economic Connectedness



(e) Political Homogeneity



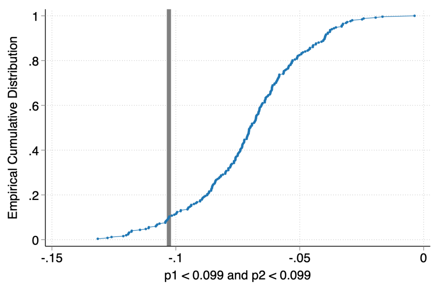
(f) Share Extreme Margins (70 points)



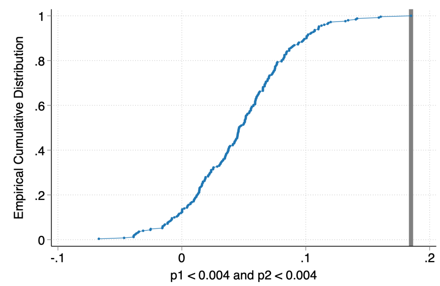
Notes: The figure plots the results of our randomization analysis for our main results. We plot the empirical cumulative distribution obtained by randomly shuffling the observed email distribution and computing our reduced-form analysis on the simulated data 250 times. The black solid line indicates our estimate in the observed data. Under each graph we present two summary statistics: $p1$ is the fraction of placebo estimates with larger magnitude than our true estimate when the true estimate is positive. When our true estimate is negative $p1$ is the fraction of placebo estimates with lower magnitude than our true estimate. $p2$ is the fraction of placebo estimates with larger magnitude than our true estimate in absolute value.

Figure A.14: Randomization Analysis, Additional Results

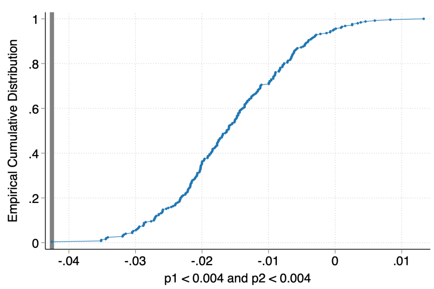
(a) Log Other Social Media Visits



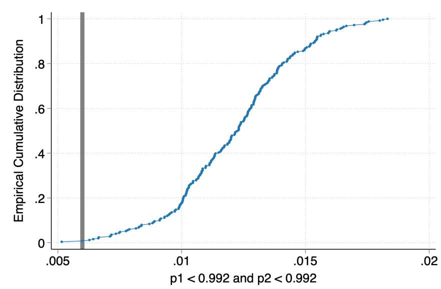
(b) Log Any Social Media Visits



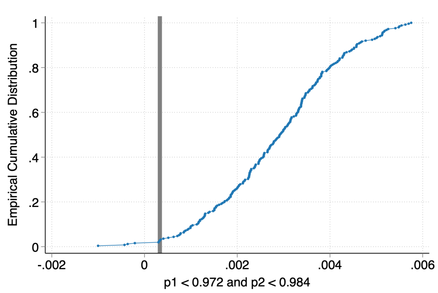
(c) Vote Margins



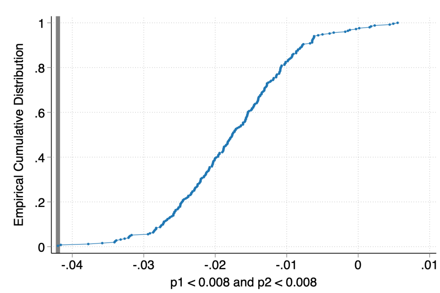
(d) Standard Deviation Republican Vote Share



(e) Standard Deviation Political Homogeneity



(f) Share Extreme Margins (50 points)



Notes: The figure plots the results of our randomization analysis for additional results of the paper. We plot the empirical cumulative distribution obtained by randomly shuffling the observed email distribution and computing our analysis on the simulated data 250 times. The black solid line indicates our estimate in the observed data. Under each graph we present two summary statistics: $p1$ is the fraction of placebo estimates with larger magnitude than our true estimate when the true estimate is positive. When our true estimate is negative $p1$ is the fraction of placebo estimates with lower magnitude than our true estimate. $p2$ is the fraction of placebo estimates with larger magnitude than our true estimate in absolute value.

Table A.1: Descriptive Statistics for Independent and Control Variables

Variable	Mean	Std. Dev.	Min	Max	N. Obs
Gmail Homophily Shock	0.00	1.00	-4.66	5.10	3042
Pre-period Gmail Complementarity	0.00	1.00	-4.41	8.32	3042
Online Homophily	0.00	1.00	-5.02	1.68	3042
Total Facebook Links (000s)	3532.69	3310.99	210.83	44647.79	3042
Share Facebook Links Outside County	1.00	0.00	1.00	1.00	3042
Main Demographic Variables (2010)					
Population (000s)	89.41	208.39	0.29	2504.70	3042
Share White	0.83	0.16	0.10	0.99	3042
Share Black	0.09	0.15	0.00	0.86	3042
Share Hispanic	0.08	0.13	0.00	0.96	3042
Median Income (000s)	45.44	11.55	19.62	122.07	3042
Share Some College	0.33	0.07	0.13	0.66	3042
Share Labor Force	0.48	0.06	0.21	0.72	3042
Share Unemployed	0.09	0.04	0.00	0.27	3042
Share Rural	0.59	0.31	0.00	1.00	3042
Median Age	40.42	4.98	22.60	62.70	3042
Main Demographic Variables in Changes (2010-2000)					
Δ Population (000s)	8.05	30.28	-240.58	644.25	3042
Δ Share White	-0.01	0.03	-0.31	0.23	3042
Δ Share Black	0.00	0.02	-0.14	0.28	3042
Δ Share Hispanic	0.02	0.02	-0.05	0.22	3042
Δ Median Income (000s)	10.18	5.06	-6.12	41.42	3042
Δ Share Some College	0.05	0.02	-0.07	0.22	3042
Δ Share Labor Force	0.01	0.03	-0.15	0.25	3042
Δ Share Unemployed	0.03	0.03	-0.12	0.18	3042
Δ Share Rural	-0.01	0.06	-0.80	0.65	3042
Δ Median Age	3.01	1.81	-5.10	13.50	3042
Main Political Variables (2008)					
Political Homogeneity	0.55	0.06	0.50	0.90	3042
Republican Share	0.58	0.14	0.07	0.95	3042
Turnout (000s)	38.85	87.26	0.16	926.46	3042

Notes: Descriptive Statistics of all independent and control variables used in the analysis.

Table A.2: Descriptive Statistics for Outcome Variables

Variable	Mean	Std. Dev.	Min	Max	N. Obs
Online Visits and Time Consumption					
Total Minutes Spent Online (000s)	27.19	79.83	0.00	1190	2872
Facebook Visits (000s)	1.28	3.28	0.00	50	2872
Other Social Media Visits (000s)	0.16	0.62	0.00	10	2872
Total Internet Visits (000s)	301.92	901.69	0.00	14815	2872
Minutes Spent on Facebook (000s)	26.26	67.22	0.00	1323	2872
Minutes Spent on other Social Media (000s)	1.35	6.39	0.00	134	2872
Offline Visits					
Any Venue (000s)	318.42	872.59	0.07	12777	36564
Bars	604.07	2704.91	0.00	83050	36564
Restaurants	68175.91	205511.46	0.00	5034405	36564
Theatres	2.46	68.13	0.00	3991	36564
Live Sport	1101.25	5253.80	0.00	127361	36564
Amusement Parks	1444.56	31251.21	0.00	2063582	36564
Golf Facilities	4095.40	14985.50	0.00	430902	36564
Skiing Facilities	270.23	3193.08	0.00	226112	36564
Fitness Centers	11619.58	38797.47	0.00	867381	36564
Bowling Centers	919.16	2852.10	0.00	53369	36564
Other Recreational	719.29	7074.38	0.00	337198	36564
Religious Org.	5699.44	13576.57	0.00	272436	36564
Voluntary Org.	3.43	46.23	0.00	2202	36564
Social Capital and Long Ties					
Econ Connect	0.00	1.00	-2.94	3.10	2943
Econ Connect SE	0.00	1.00	-1.47	6.01	2943
Child Econ Connect	0.00	1.00	-2.75	3.61	2664
Child Econ Connect SE	0.00	1.00	-1.54	4.44	2664
Econ Connect by Group	0.00	1.00	-3.11	2.95	2937
Econ Connect for High SES	0.00	1.00	-3.13	2.62	2943
Econ Connect SE for High SES	0.00	1.00	-1.61	5.85	2943
Child Econ Connect for High SES	0.00	1.00	-2.83	2.95	2664
Child Econ Connect SE for High SES	0.00	1.00	-1.57	4.28	2664
Econ Connect by Group for High SES	0.00	1.00	-3.28	2.51	2937
Fraction Long Ties	0.00	1.00	-2.64	4.15	3036
Electoral Outcomes, 2020 Presidential Elections					
Political Homogeneity	0.60	0.09	0.50	0.94	3042
Vote Margin	0.40	0.22	0.00	0.94	3042
Republican Share	0.66	0.16	0.06	0.97	3042
Turnout	46481.81	106697.22	159.00	1210507.00	3042
Electoral Outcomes, 2020 House Elections					
Political Homogeneity	0.62	0.11	0.50	1.00	3040
Vote Margin	0.42	0.24	0.00	1.00	3040
Republican Share	0.67	0.18	0.00	1.00	3040
Turnout	44876.20	103032.78	151.00	1170400.00	3040
Precinct Level Electoral Outcomes					
Std. Dev. Republican Share, 2016	0.11	0.07	0.00	0.40	2729
Iqr Republican Share, 2016	0.15	0.11	0.00	0.80	2729
Share Margins Above 50	0.49	0.33	0.00	1.00	2729
Share Margins Above 70	0.22	0.26	0.00	1.00	2729
CCES Outcome Variables					
Extreme Partisanship	0.42	0.49	0.00	1.00	391880
Extreme Ideology	0.21	0.41	0.00	1.00	391880

Notes: Descriptive Statistics of all the dependent variables used in the analysis.

Table A.3: Homophily Shock and Total Internet Usage

Dep. Variable:	Log Tot Visits				Log Tot Minutes			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GH Shock	0.047 (0.089)	0.114 (0.079)	0.126 (0.083)	0.138* (0.078)	0.030 (0.106)	0.105 (0.094)	0.119 (0.099)	0.128 (0.091)
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*
Mean of Dep. Var.	7.973	7.973	7.973	7.973	10.278	10.278	10.278	10.278
Adj R2	0.727	0.731	0.731	0.732	0.694	0.698	0.698	0.698
Observations	2872	2872	2872	2872	2872	2872	2872	2872

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables measure log online visits (columns 1-4) and log minutes spent online (columns 5-8). The data on web searches comes from ComScore and refers to the first trimester of 2016. Gmail Homophily Shock (GH Shock) measures 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.4: 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)
GH 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)
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*
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

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables measure log minutes spent on Facebook (columns 1-4) and log minutes spent on other social media (columns 5-8). Other social media include Instagram, Twitter and Reddit. The data on web searches comes from ComScore and refers to the first trimester of 2016. Gmail Homophily Shock (GH Shock) measures 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.5: Homophily Shock and Social Media Visits (IHS)

Dep. Variable:	IHS Facebook Visits				IHS Other SM Visits			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GH 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)
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*
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

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, Twitter and Reddit. The data on web searches comes from ComScore and refers to the first trimester of 2016. Gmail Homophily Shock (GH Shock) measures 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.6: 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)
GH 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)
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*
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

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, Twitter and Reddit. The data on web searches comes from ComScore and refers to the first trimester of 2016. Gmail Homophily Shock (GH Shock) measures 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.7: Homophily Shock 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)
GH Shock	0.255***	0.222***	0.199***	0.201***	0.302***	0.283***	0.241***	0.247***
	(0.060)	(0.056)	(0.058)	(0.062)	(0.081)	(0.080)	(0.082)	(0.090)
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*
Mean of Dep. Var.	5.540	5.540	5.540	5.540	8.019	8.019	8.019	8.019
Adj R2	0.875	0.875	0.875	0.876	0.795	0.796	0.796	0.796
Observations	2872	2872	2872	2872	2872	2872	2872	2872

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, Twitter and Reddit. The data on web searches comes from ComScore and refers to the first trimester of 2016. Gmail Homophily Shock (GH Shock) measures 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.8: Homophily Shock and Offline Visits to Any Venue, 2019

Dep. Variable:	Log Total Visits					
	(1)	(2)	(3)	(4)	(5)	(6)
GH Shock	-0.005 (0.045)	0.036 (0.045)	-0.007 (0.044)	0.031 (0.044)	0.030 (0.045)	0.048 (0.041)
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*
Month 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.	11.090	11.090	11.090	11.090	11.090	11.090
Adj R2	0.924	0.936	0.957	0.963	0.963	0.964
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 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 (GH Shock) measures 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.9: Homophily Shock and Bars and Restaurants Visits (IHS)

Dep. Variable:	IHS Bars and Restaurants Visits					
	(1)	(2)	(3)	(4)	(5)	(6)
GH 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)
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*
Month 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.	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

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 (GH Shock) measures 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.10: Homophily Shock and Visits to Any Venues (IHS)

Dep. Variable:	IHS Total Visits					
	(1)	(2)	(3)	(4)	(5)	(6)
GH Shock	-0.005	0.036	-0.007	0.031	0.030	0.048
	(0.045)	(0.045)	(0.044)	(0.044)	(0.045)	(0.041)
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*
Month 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.	11.783	11.783	11.783	11.783	11.783	11.783
Adj R2	0.924	0.936	0.957	0.963	0.963	0.964
Observations	36564	36564	36564	36564	36564	36564

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 (GH Shock) measures 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.11: Homophily Shock and the Total Number of Facebook Connections, 2016

Dep. Variable:	Log All Connections, 2016					
	(1)	(2)	(3)	(4)	(5)	(6)
GH Shock	-0.028	-0.026	-0.029	-0.028	-0.023	-0.019
	(0.019)	(0.022)	(0.025)	(0.020)	(0.020)	(0.018)
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.	15.397	15.397	15.397	15.397	15.397	15.397
Adj R2	0.928	0.930	0.939	0.951	0.951	0.953
Observations	3042	3042	3042	3042	3042	3042

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the log number of total Facebook connections in 2016. Gmail Homophily Shock (GH Shock) measures 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.12: Homophily Shock and the Total Number of Facebook Connections, 2016 (IHS)

Dep. Variable:	IHS All Connections, 2016					
	(1)	(2)	(3)	(4)	(5)	(6)
GH Shock	-0.028 (0.019)	-0.026 (0.022)	-0.029 (0.025)	-0.028 (0.020)	-0.023 (0.020)	-0.019 (0.018)
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.	16.091	16.091	16.091	16.091	16.091	16.091
Adj R2	0.928	0.930	0.939	0.951	0.951	0.953
Observations	3042	3042	3042	3042	3042	3042

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the inverse hyperbolic sine of the number of total Facebook connections in 2016. Gmail Homophily Shock (GH Shock) measures 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.13: Main Results Controlling for Total Number of Facebook Links

	Online Homophily		Log Facebook Visits		Log Bars and Rest. Visits		Econ Connect		Political Homogeneity		Extreme Margins	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
GH Shock	0.308***	0.300***	0.203***	0.205***	-0.067**	-0.065**	-0.172***	-0.170***	-0.024***	-0.024***	-0.042***	-0.044***
	(0.079)	(0.077)	(0.061)	(0.061)	(0.032)	(0.031)	(0.059)	(0.059)	(0.006)	(0.006)	(0.011)	(0.011)
Log All Links, 2016		0.206**		-0.036		-0.050		-0.067		0.005*		0.046***
		(0.084)		(0.066)		(0.066)		(0.046)		(0.003)		(0.016)
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*
Adj R2	0.769	0.772	0.867	0.867	0.954	0.954	0.872	0.872	0.876	0.876	0.757	0.759
Mean of Dep. Var.	0.000	0.000	4.853	4.853	9.318	9.318	0.000	0.000	0.605	0.605	0.357	0.357
Observations	3042	3042	2872	2872	36564	36564	2943	2943	3042	3042	2729	2729

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 connections in 2016. The dependent variables are the six main outcomes of the paper: Online Homophily 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 70 points in 2016 in columns 11 and 12. Gmail Homophily Shock (GH Shock) measures 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. In line with their respective specifications, columns 3, 4, 5 and 6 also control for the total number of online and offline visits respectively. In columns 5 and 6 we also account for month fixed effects together with DMA fixed effects. Robust standard errors clustered by state in parentheses.

Table A.14: Gmail Complementarity Shock Reduces Extreme Ideology

Dep. Variable:	Extreme Ideology					
	(1)	(2)	(3)	(4)	(5)	(6)
GH Shock × Post	-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. We build this indicator by transforming the answer to a survey question in the Cooperative Election Study (CES) asking respondents to place themselves on a ideology scale of five possible alternatives ranging from "Very Conservative" to "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.15: Homophily Shock Has No Impact on Republican Vote Share, 2020

Dep. Variable:	Republican Share, 2020					
	(1)	(2)	(3)	(4)	(5)	(6)
GH 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)
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.	0.664	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

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 (GH Shock) measures 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: Homophily Shock Has No Impact on Turnout, 2020

Dep. Variable:	Log Turnout, 2020					
	(1)	(2)	(3)	(4)	(5)	(6)
GH 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)
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.	9.562	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

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the log turnout in the 2020 presidential election. Gmail Homophily Shock (GH Shock) measures 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: Homophily Shock and the Share of Facebook Links Outside the County, 2016

Dep. Variable:	Share Out Connections, 2016					
	(1)	(2)	(3)	(4)	(5)	(6)
GH Shock	-0.029***	-0.001	-0.001	0.007	0.004	0.004
	(0.010)	(0.010)	(0.007)	(0.009)	(0.008)	(0.008)
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.	0.561	0.561	0.561	0.561	0.561	0.561
Adj R2	0.628	0.691	0.792	0.800	0.801	0.808
Observations	3042	3042	3042	3042	3042	3042

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the share of Facebook connections outside the county in 2016. Gmail Homophily Shock (GH Shock) measures 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.18: Homophily Shock Reduces the Fraction of Long Ties

Dep. Variable:	Fraction of Long Ties					
	(1)	(2)	(3)	(4)	(5)	(6)
GH Shock	-0.431***	-0.149**	-0.231***	-0.176***	-0.169***	-0.155***
	(0.100)	(0.067)	(0.053)	(0.047)	(0.048)	(0.045)
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	0.465	0.650	0.854	0.865	0.865	0.883
Observations	3036	3036	3036	3036	3036	3036

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the county-level fraction of long ties between users with zero mutual friends, where at least one user resides in a given county (Jahani et al., 2023). We standardize the dependent variable to have zero mean and standard deviation equal to one. Gmail Homophily Shock (GH Shock) measures 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.19: Impact of the Gmail Homophily Shock by Share Urban

	Online Homophily	Log Facebook Visits	Log Bars and Rest. Visits	Econ Connect	Political Homogeneity	Extreme Margins
	(1)	(2)	(3)	(4)	(5)	(6)
GH Shock	0.183**	0.101	-0.084**	-0.127**	-0.015***	-0.025*
	(0.090)	(0.061)	(0.039)	(0.062)	(0.005)	(0.013)
– × Share Urban	0.130***	0.102***	0.019	-0.041***	-0.009***	-0.018***
	(0.027)	(0.024)	(0.017)	(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	0.780	0.868	0.954	0.873	0.882	0.760
Mean of Dep. Var.	0.000	4.853	9.318	0.000	0.605	0.218
Observations	3042	2872	36564	2943	3042	2729

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: Online Homophily in column 1, Log Facebook visits in column 2, Log Bars and Restaurants 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 70 points in 2016 in column 6. Gmail Homophily Shock (GH Shock) measures 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 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. In line with their respective specifications, columns 2 and 3 also control for the total number of online and offline visits respectively. In column 3 we also account for month fixed effects together with DMA fixed effects. Robust standard errors clustered by state in parentheses.

Table A.20: Main Results are Robust to Alternative Splits in the Construction of GH Shock

	Online Homophily	Log Facebook Visits	Log Bars and Rest. Visits	Econ Connect	Political Homogeneity	Extreme Margins
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Median Split					
GH Shock	0.343^{***}	0.128^{**}	-0.079^{***}	-0.072^{*}	-0.022^{***}	-0.040^{***}
	(0.066)	(0.049)	(0.023)	(0.041)	(0.004)	(0.006)
	Panel B: Quartile Split					
GH Shock	0.315^{***}	0.103^{*}	-0.036	-0.103^{***}	-0.015^{***}	-0.038^{***}
	(0.058)	(0.052)	(0.027)	(0.033)	(0.004)	(0.009)
	Panel C: Continuous Measure					
GH Shock	0.269^{***}	0.121^{***}	-0.038^{**}	-0.067^{***}	-0.013^{***}	-0.017^{***}
	(0.032)	(0.027)	(0.016)	(0.017)	(0.003)	(0.006)
Adj R2	0.801	0.868	0.952	0.872	0.879	0.761
Mean of Dep. Var.	-0.000	4.853	9.316	0.001	0.605	0.218
Observations	3044	2874	36588	2944	3044	2730
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*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table presents the main results of the paper with alternative construction of the GH Shock. Panel A shows main results where we define high and low social distance using above and below median values; in Panel B we split the data using top and bottom quartiles; In Panel C we present results where we employ a continuous measure that weights the Gmail complementarity by the distance to the median value of the social distance distribution. The dependent variables are the six main outcomes of the paper: Online Homophily in column 1, Log Facebook visits in column 2, Log Bars and Restaurants 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 70 points in 2016 in column 6. Gmail Homophily Shock (GH Shock) measures 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 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. Columns 2 and 3 also control for the total number of online and offline visits respectively. In column 3 we account for month fixed effects and DMA fixed effects. Robust standard errors clustered by state in parentheses.

Table A.21: 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	0.174***	0.193***	0.183***	0.193***	-0.045	-0.028	-0.052	-0.051
	(0.034)	(0.035)	(0.037)	(0.039)	(0.028)	(0.029)	(0.033)	(0.034)
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*
Mean of Dep. Var.	4.853	4.853	4.853	4.853	2.391	2.391	2.391	2.391
Adj R2	0.866	0.867	0.867	0.868	0.841	0.842	0.842	0.842
Observations	2872	2872	2872	2872	2872	2872	2872	2872

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables are the 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. Robust standard errors clustered by state in parentheses.

Table A.22: 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	0.263***	0.291***	0.269***	0.285***	-0.117**	-0.091*	-0.131**	-0.132**
	(0.048)	(0.049)	(0.049)	(0.048)	(0.047)	(0.048)	(0.052)	(0.051)
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*
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	0.790	0.762	0.763	0.763	0.763
Observations	2872	2872	2872	2872	2872	2872	2872	2872

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables are the log minutes spent on Facebook (columns 1-4) and on other social media (columns 5-8). Other social media include Instagram, Twitter and Reddit. 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 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.23: Homophily Shock and Visits to Bars and Restaurant, 2019, OLS

Dep. Variable:	Log Bars and Restaurants Visits					
	(1)	(2)	(3)	(4)	(5)	(6)
Online Homophily	-0.034**	-0.064**	-0.027	-0.070**	-0.075**	-0.083**
	(0.015)	(0.025)	(0.031)	(0.033)	(0.037)	(0.037)
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.	9.318	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

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. 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.24: Homophily Shock and Economic Connectedness, OLS

Dep. Variable:	Econ Connectedness					
	(1)	(2)	(3)	(4)	(5)	(6)
Online Homophily	0.098*	-0.113**	-0.136***	-0.105***	-0.098***	-0.071***
	(0.053)	(0.044)	(0.028)	(0.021)	(0.023)	(0.022)
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	0.355	0.697	0.814	0.863	0.863	0.871
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.25: Online Homophily and Political Homogeneity, 2020, OLS

Dep. Variable:	Political Homogeneity, 2020					
	(1)	(2)	(3)	(4)	(5)	(6)
Online Homophily	0.006	-0.024***	-0.025***	-0.029***	-0.003	-0.006**
	(0.005)	(0.005)	(0.005)	(0.004)	(0.002)	(0.002)
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	0.257	0.536	0.631	0.654	0.863	0.873
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.26: 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	0.006**	-0.003**	0.011**	-0.006*	-0.060***	-0.023**	-0.063***	-0.021**
	(0.003)	(0.002)	(0.005)	(0.003)	(0.016)	(0.009)	(0.011)	(0.009)
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	0.645	0.707	0.560	0.638	0.643	0.768	0.561	0.757
Mean of Dep. Var.	0.114	0.114	0.150	0.150	0.495	0.495	0.218	0.218
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 50 points (cols 5-6) and the share of precincts with vote margins of at least 70 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.