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

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

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

September 5, 2024

Abstract

We examine the long-run effect of homophily in online social networks on interpersonal interactions in local communities. We measure online homophily across counties in the US using Facebook data. For identification, we exploit a conflict between Facebook and Google over data sharing of user information during the early expansion phase of Facebook. We find evidence that homophilic connections led to increased social media usage but reduced *offline* socialization. This shift was accompanied by deterioration of local social cohesion, as individuals became less connected across income strata and less likely to share the same political opinions with others in their counties.

Keywords: Social Media, Networks, Homophily, Social Capital

*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 Postdoctoral Fellow at the Barcelona School of Economics and at the Institute of Economic Analysis. 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 Ministerio de Ciencia e Innovacion grant AEI/10.13039/501100011033.

1 Introduction

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

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* U.S. counties. Our goal is to document the effects in counties that ended up with Facebook links to other socially similar counties, as opposed to socially dissimilar.[1] The key empirical challenge is that these friendship links are driven by endogenous selection; counties where people tend to have friends in very similar places will differ along many dimensions to begin with.[2] An experiment which randomly assigns whether counties are connected to socially similar or dissimilar counties and track outcomes over many years is, obviously, not practically feasible.

To overcome this challenge, we make use of a unique natural experiment by exploiting a data-sharing conflict between Facebook and Google that started in 2010, a period during which many new users joined the social network. Before the conflict, with the "Find Friends" function provided by Facebook, new users of the network were offered to import their email contact data to Facebook

[1]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 in weighted by their Facebook friendship links. It is quite intuitive, capturing the social similarity of the Facebook connections emanating out of a county. We will call this variable Online Homophily.

[2]To identify the causal impact is difficult since friendship links are highly endogenous to the preferences of users, who self-select into networks characterized by high degrees of homophily. In fact, homophily is one of the most salient features of social networks, both offline and online, as people are more likely to form connections with those who resemble them in terms of race, socioeconomic status, political preferences, and other attributes (see Jackson (2008) for an overview).

and easily connect to their contacts who were already on Facebook.[3] This was possible as Facebook was effectively given permission to cross-reference all Facebook users with all user email addresses. 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 a Gmail account since the "Find Friends" function would be deactivated. Through several data construction steps we create the county-level variation of interest. First, we posit that the deactivation of the "Find Friends" function gave rise to a form of Gmail complementarity in connectivity costs between pair of counties where Gmail was popular, but only during the relevant time period. We make use of data on the relative popularity of Gmail compared to other email clients (Yahoo! and Windows Live Hotmail, the two other dominant clients at the time) in respective counties at different moments in time. We document that there is a persistent decline in bilateral county-to-county Facebook connections (as measured in 2016 and 2020) when Gmail complementarity is computed after, but not before November 2010. This establishes that the Google-Facebook conflict indeed had a long-term effect on the patterns of bilateral friendship links. Since the average level of popularity of Gmail compared to other email clients can be correlated with potentially important characteristics of counties, we use the difference in Gmail complementarity before and after the conflict as a source of variation in the likelihood of forming the connections *between pairs of counties.* Finally, to create a source of exogenous variation in the homophily of Facebook network connections *at the county level* we compare Gmail complementarity before and after the conflict for counties with high and low social distance from each particular county. The intuition is that the data sharing conflict between Facebook and Google induced some counties to get relatively more connections with more similar counties (i.e. low social distance counties), while other counties were induced to form disproportionately more connections with less similar counties. More specifically, we create a variable called Gmail Homophily Shock (GH Shock), which subtracts the average Gmail complementarity with socially similar counties from that

[3] As an example of the user interface when signing up for Facebook, requiring new users to share their email login information, including the password to their email accounts, see Figure A.2. The data sharing conflict was reported on in mass media at the time; see Figure A.3.

with socially dissimilar counties, for the period between November 2010 and April 2012. Simply put, while the data construction is rather involved, the variation implies a plausibly exogenous shock that shifted some counties to have greater degree of homophily in their connections to other counties.

With this county-level data at hand, we first test and confirm that the GH Shock had a persistent "first stage" effect on the degree of online Facebook homophily in US counties, detectable in both 2016 and 2020. We provide numerous robustness checks to establish this shock can be deemed as a source of credible variation for capturing causal relationships. Most importantly, as a placebo test, we show that the Gmail complementarity in the period before the conflict is irrelevant for homophily.[4] This indicates that we have a valid first-stage regression, with a shock to online homophily driven by the temporary conflict between Facebook and Gmail, rather than the differences in the average popularity of different email clients.[5]

We use the shock to study the long-term effect of online homophily on social interactions. 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, such as Twitter, Instagram, Reddit, etc. 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 speaks to the well-known demand for online homophily; the overall online experience becomes more attractive to users.

With this shift in greater social media usage in online networks that are more socially similar, what are the consequences for traditional community bonds? This is an important potential conse-

[4]In addition to checking whether pre-determined variables and trends are correlated with the shock, we document that Gmail complementarity *before* November 2010 did not have a significant impact on online homophily in 2016, but GH Shock *between* November 2010 and April 2012 had a strong positive effect on subsequent online homophily. This effect becomes much smaller quantitatively if Gmail complementarity is computed *after* April 2012, when "Find Friends" by email was discontinued, but long-term consequences are, nevertheless, likely to remain.

[5]One question is whether our GH Shock variable is a shock to homophily or a shock to the total number of friends. By construction, we always control for the preferences for email clients in the pre-2010 time period, and we look at the difference between high- and low-social distance counties. However, it could still be the case that our GH Shock also positively predicts the total number of connections. Empirically, we do not find a significant effect of GH Shock on the number of connections in either direction, while the impact of GH Shock on the measure of online homophily is positive and highly significant 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.

quence to study, especially since the presence of externalities can have welfare implications. Recent work by Bursztyn et al. (2023) highlights how social media can impose negative externalities on non-users, resulting in reduced overall welfare despite high levels of usage. Users, while consuming it heavily, may even prefer coordination among friends so that the online option does not even exist. This parallel suggests that increased online homophily, while boosting social media engagement, might similarly have unintended negative consequences on offline socialization and the strength of local communities. For example, if you are preoccupied with online content, you may not be interested in going to a restaurant with your friend. Moreover, if you would like to go but your friends are not interested, your demand for the restaurant is lower. You may then decide to spend time on social media for the simple reason there is nobody to go to the restaurant with. Over time, offline socialization erodes. To test this hypothesis, we examine the effects of increased Facebook homophily on offline interactions within the county. Using SafeGraph mobility data for 2019 (capturing 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, restaurants, and 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, less time on other social media, more time on social media in total, and 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.

Finally, we document that these effects were accompanied with important impacts on local social cohesion along a number of dimensions. 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. We also find evidence of implications for the *distribution of political opinions*. Online homophily can affect the distribution of political opinions in two ways. First, by focusing on communication with like-minded people online, and getting constant reinforcement of their pre-existing political preferences, politics could become more extreme, and within-county voting behavior becomes more one-sided. Alternatively, by reducing offline within-county communications – social interactions in another realm with a high baseline degree of homophily – voting preferences within counties could become less similar, more diverse, and less polarized. We find that higher online homophily made local political opinions more diverse starting in 2016. Within-county homogeneity is reduced, as measured by the probability that two randomly picked county residents vote for the same party. We also find that exposure to the online homophily shock decreased the probability of extreme vote margins and increased within-county measures of dispersion of political opinions, such as inter-quartile range and standard deviation of vote shares. Furthermore, online homophily made people less extreme in answering survey questions: an increase in online homophily made Cooperative Election Study respondents less likely to say that they are "Strong Democrats" or "Strong Republicans".[6]

As an important robustness test, for all the outcomes affected by online homophily, we document an inverted U-pattern for the effect of the GH Shock over time, similar to the results for the first stage: the effects are not significant for the GH Shock before November 2010, they are significant for the GH Shock between November 2010 and April 2012, and they are small, though showing some degree of policy persistence, for the GH Shock after 2012.

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

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

[6] It is noteworthy that online homophily did *not* disproportionately benefit one 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. One sanity check that we do is to show that most of our estimates are weaker in places with a larger share of Facebook connections coming from within their own county. This finding is consistent with the idea that Gmail Homophily Shock should have a stronger impact on the network of connections if more of these connections are coming from other counties. Note that the share of own county links does not significantly depend on the GH Shock, so the results are not simply driven by a non-linear effect of GH Shock. 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.

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

Second, our work relates closely to the literature on networks and homophily in the networks. 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). 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 studying the causal impact of online homophily.

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

The rest of the paper is organized as follows. Section 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 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. (2018b).[7] The measures of connectedness are available 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 across time and space, we use Google's Search Volume Index (SVI) at the DMA-level at the quarterly level 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 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 (2021). Precinct-level vote shares for the 2016 presidential election come from Kaplan et al. (2022). To measure the ideology of US counties, we exploit polls from US Tracker Gallup (Gallup 2017). We collapse individual-level self-assessed ideology (ranging from 1 "very liberal" to 5 "very conservative") from 2008, 2009, and 2010 at the county level.

We leverage survey data from the Cooperative Election Studies (CES) to reconstruct variation overtime in the intensity of political preferences.[8]

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

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

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

Offline Activity. Data on visits to different establishments comes from SafeGraph.[9] The data we obtain 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 county pair, we calculate differences in terms of their demographic characteristics (as measured by the percentage of Whites, Blacks, Hispanics, those with at least college education, median income, total population, percentage of rural population, median age, percentage in labor force, and percentage unemployed), their political preferences (as measured by the share of Republican votes in 2004), and their ideology score (as measured by the county average self-assessed ideology from Gallup polls of 2008-2010 that ranges from 1 very liberal through 5 very conservative). We 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, $Social\_Similarity_{ij}$.

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 it is connected to, using the share of

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

Facebook connections as weights. Formally, we compute

$$Online\_Homophily_i = \sum_{j=1}^{J} \pi_{ij} Social\_Similarity_{ij} \tag{1}$$

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

[Figure 1 about here.]

3 Empirical Strategy

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

3.1 Empirical Challenge and Construction of the Instrument

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 other's opinions, and the extent of inter-group contact. These factors might also separately affect our outcomes of interest. Thus, we need an identification strategy 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 a friendship with more similar individuals after meeting with them). To identify the effect of online homophily we exploit variation in exposure to potential Facebook friends caused by the conflict between Google and Facebook in 2010, which changed the way Facebook suggested friends to new joining users, 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.[10]

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

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

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

[Figure 2 about here.]

Gmail Complementarity Shock. Before constructing our instrumental variable, we start by investigating if indeed Facebook connections responded to the Gmail-Facebook conflict. We hypothesize that Gmail users had a smaller probability of forming a friendship connection with other

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

Gmail users after November 2010. Hence, we expect that 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, this 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, γ_i and μ_j are county fixed effects, and ϵ_{ij} is an error term.[11]

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

[Figure 3 about here.]

Construction of the Instrumental Variable. We now exploit the Gmail complementarity shock for the county-pair connections to construct our instrumental variable for 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 Gmail Homophily Shock (GH_shock_i) in two steps. First, we compute

[11] 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

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

a social distance between US counties using the same twelve characteristics we used to construct our online homophily measure. Employing this distance, and for each county, we divide US counties between high and low social distance counties (see equation (1)).[12] 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.

[Figure 4 about here.]

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 4. The left bar on the graph represents the value of $\sum_{t=1}^{6} \overline{gmail_{it}}^{HI}$ term from equation (3), while the right bar represents $\sum_{t=1}^{6} \overline{gmail_{it}}^{LO}$ term from equation (3). Since the first term in equation (3) is higher than the second term and since 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.6 shows the value of the two different terms in equation (3) in every county in Alabama.[13] As one can see, Facebook-Google conflict introduced heterogenous 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

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

[13] The variation of the final variable we construct, the difference between the two bars for each county, is shown in Appendix Figure A.7.

shock with similar counties than with socially distant counties for the first group and vice-versa for the second group.

We map the residual variation of this measure controlling for Baseline Controls and DMA fixed effects in Figure 5. As one can see, there is a high degree of heterogeneity in this variable both within and across American states.

3.2 Implied First Stage: Gmail Homophily Shock and Online Homophily

In this sub-section, we check whether Gmail Homophily Shock, the variable we just constructed, is a good predictor of online homophily – our main independent variable of interest described in subsection 2.3. Formally, we estimate equation (4).

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

where $Online\_Homophily_i$, 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.

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 further include political homogeneity in 2008 to the list of controls.[14] Demographic Trends include differences between 2010 and 2000 for all the baseline, demographic, and political controls. 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 identification comes from the *change* in Gmail complementarity during the Facebook-Google conflict. Finally, to facilitate the interpretation of coefficients we standardize our independent variables. Standard errors are clustered at the state level.

In column 1, where we only control for log population and pre-period Gmail complementarity, the relationship between Online Homophily and the Gmail Homophily Shock is 0.625, positive and significant at the 1% level. The magnitude of the effect decreases to 0.277 after the addition of

[14]Political homogeneity is defined as one minus the Herfindal Index calculated using Democrat and Republican vote shares. In this case, assuming the Republican and Democrat vote share sum up to 1, the formula boils down $1 - 2r(1 - r)$, where r is the Republican vote share. See more details about this variable and the rationale to use it in section 4.4.

Baseline Controls but it increases after adding more controls and in the most saturated specifications converges to 0.344, 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 0.305, which is slightly smaller than the estimate in column 6.

The magnitude of our preferred specification (column 6) implies that one standard deviation increase in the Gmail Homophily Shock increases Online Homophily by about 34% of a standard deviation. Kleibergen-Paap F-statistics in all specifications that include Baseline Controls are always above 30, so we do not need to use weak instrument robust methods.

[Table 1 about here.]

[Figure 5 about here.]

For interpretation of our results, it is important to understand what kind of connections are affected by our instrument. 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 search for 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 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. 2018a). 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 the independent variables when we use them as outcome variables). Figure 6 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. We still observe some small significant relationship with the share of unemployed (negative) and change in median income (positive). 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 6 about here.]

[Figure 7 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 7a shows these results for the implied first stage as three bar graphs, with coefficients for the pre-period, during the treatment period, and after the treatment period shown side by side. As one can see, the relationship between GH Shock and online homophily in pre-period is 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, in what follows, we always control for pre-period, to take into account possible average differences across counties with different baseline levels of Gmail complementarity, i.e. pre-existing Gmail complementarity before policy change.

In what follows, we will focus on the reduced-form relationship between Gmail Homophily Shock, social capital, and local political cohesion, as well as IV estimation where Online Homophily is instrumented by GH Shock. For the reduced form relationship, we estimate equation (4) using as the dependent variable one of our outcomes of interest. 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 randomization inference (see Figures A.13 and A.14 in the Online Appendix).

In addition, we provide the result of an instrumental variable approach where we use the Gmail Homophily Shock as the IV and our measure of online homophily as the endogenous variable. More precisely, we estimate the following second-stage regression

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

where $Outcome_{it}$ is one of our outcomes variables, $Online\_Homophily_i$ is the endogenous variable we construct measuring online homophily according to equation (1) instrumented by GH_shock_i defined in equation (3); X_i is the same set of controls as in the (implied) first stage presented above in equation (4); ε_i is an error term (as in equation (4) we use standard errors clustered at the state level).

4 Results

4.1 Online Homophily and Usage of Social Media

We start our analysis by 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; the bottom panel of this Table summarizes the results of the IV specification presented in equation (5). Columns (1)-(4) show the results for the (log) number of Facebook visits, while columns (5)-(8) summarize the results for the (log) visits to other social media, that is Twitter, Instagram, and Reddit.[15]

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 activities 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 20.3% increase in the number of Facebook visits and 10.3% decrease in the number of other social media visits, with the former coefficient significant at 1% and the latter at 5%. Similarly, our results in the bottom panel imply that one standard deviation higher *Online Homophily* increases visits to Facebook by 66.1% (column 4), significant at 1%, whereas it decreases visits to other social media by 33.4% (column 8), significant at 5%. The Kleibergen-Paap F-stat is consistently above 28, ruling out weak instrument concerns.

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

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

number of minutes spent on Facebook and reduces the number of minutes spent on other social media. The magnitudes in the most saturated specifications (columns 4 and 8) imply that one standard deviation increase in GH Shock leads to 25.5% increase in the number of minutes spent on Facebook and to 22.4% decrease in the number of minutes spent on other social media.

Finally, Table 3 presents the results for total visits and total time spent on social media activity, which combines Facebook and other social media. The results imply that one standard deviation increase in Gmail Homophily Shock increases the total number of social media visits by 18.5% and the total number of minutes spent on any social media by 24.2%. The implied effect suggests that a one-standard deviation increase in online homophily increases total time spent on social media by about 0.7 log-points, more than a 70% increase.

Another way of seeing these results is to look at the coefficients for Gmail Homophily Shock, computed during, before, and after 6 quarters of Google-Facebook conflict (Figure 7b). As one can see, most of the effect happens during the treatment period, with the magnitudes for pre- and post-period being 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 their network patterns into account. Finally, we'd like to emphasize that qualitatively, Figure 7b resembles Figure 7a, i.e. the effect on the outcome variable of interest (Facebook visits) seems to mirror the variation in the 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 in total.

4.2 Online Homophily and Offline Activity

In this subsection, we test if 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 you are preoccupied with online content, you may not be interested in going to a restaurant with your friend. Moreover, if you would like to go but your friends are not interested, your demand for the restaurant is lower. You may then decide to spend time on social media for the simple reason 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 were people are likely to socialize, such as bars, restaurants or live sporting events. We leverage data from Safegraph covering all months of 2019 at the county-by-month level. We estimate specifications (4) and (5) 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). Table 4 presents our results on bars and restaurants where we progressively include controls in a way that mirrors our implied first-stage table (Table 1).

In all specifications presented in Table 4, the coefficients for Gmail Homophily Shock (Online Homophily in the second panel) 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 implies that a one standard deviation increase in Gmail Homophily Shock leads to 6.7% fewer visits to bars and restaurants, significant at 5% level, based on the most saturated specification (column 6). According to the IV estimates reported in the same table, a one standard deviation increase in online homophily reduces our proxy of offline interactions by 25.1%, significant at the 10% level (column 6). While the F-stats presented here are different from the previous exercise, given the different samples, they are still sufficiently high to rule out the weak instruments problem.

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

So far, we 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 bars and restaurants separately, theaters, live sports events, cultural sites, amusement parks, recreational venues, religious, and voluntary organizations (Figure 8). We document the negative and significant effects of the homophily shock on visits to bars, restaurants, live sports events, and amusement parks, and a marginally significant negative coefficient for voluntary organizations. All these results are consistent with a reduction in offline social interactions. The only positive coefficient that we document is for recreational venues, most of which are gyms, i.e. establishments that people mostly visit alone.

[Table 4 about here.]

[Figure 8 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 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 data exists.

4.3 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 was 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 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.

As in the rest of our results, the bottom panel of Table 5 reports the IV estimates using online homophily as the endogenous variable and the Gmail Homophily Shock as the instrumental variable. The point estimates are, again, consistent with the reduction in economic social capital across US communities. The magnitude in the IV estimates is .511, implying a reduction of approximately 50% of a standard deviation for each standard deviation increase in online homophily.[16] The results presented in Figure A.10 show that the results are similar if we use alternative measures of economic connectedness from Chetty et al. (2022a,b).

As in the case of other outcome variables, we report the coefficients for Gmail Homophily Shock computed during, before, and after the treatment period (see Figure 7d). The pre-coefficient is not statistically significant and has a different sign from the negative and significant coefficient in the

[16] 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.

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.]

4.4 Online Homophily and Political Opinions

4.4.1 Hypotheses

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

On the other hand, people exposed to a more homogeneous network spend more time on social media, switching away from other forms of political communication and resulting in fewer interactions within local communities (e.g., going to restaurants, bars and cultural events) which were shown to be important determinants of voting behavior (Cantoni and Pons 2022). Thus, an increase in online homophily can result in the divergence of preferences within counties.

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

4.4.2 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 a county level. This measure captures the degree of local political homogeneity (consensus), i.e., the opposite of local political fractionalization, and is defined as

$$PolHomogeneity_{it} = 1 - 2r_{it}(1 - r_{it}) \tag{6}$$

Here r_{it} is the vote share of a Republic party in county i in Presidential or House elections at time t. Figure A.8 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 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. The results based on IV estimation are numerically and qualitatively similar to the reduced form.

[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 9 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 of social media at that time. From 2016 onward, we observe a jump in the point estimates indicating a significant reduction of 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 results of congressional elections (see Figure A.11 in the Online Appendix).

Figure 7e 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 7e is in line with similar tests for other outcomes.

[Figure 9 about here.]

4.4.3 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, that include 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 10 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 or more percent.

[Figure 10 about here.]

We check if 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 7f summarizes these results. As one can see, 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 7 are, thus, consistent with each other and with our general claim: that the identifying variation comes from our treatment period and not before (even though we control for pre-period in all the 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).

4.4.4 Online Homophily and Intensity of Political Preferences

So far, we have looked at the dispersion or convergence of political preferences but not at their intensity. However, the effect of online homophily on the intensity of political preferences may have important implications, as it speaks more directly on 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.[17] We create a similar measure for extreme ideology for the respondents who defined themselves as either strongly liberal or strongly conservative.

[17] 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.

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

$$Extreme\_id_{it} = \beta_0 + \beta_1 GH\_shock_i + \beta_2 GH\_shock_i \times post_t + \beta_3 X_i + \delta_t + \epsilon_{it} \quad (7)$$

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 similar effect for extreme ideological positions). These results seem puzzling in the 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). However, it is important to note that we are looking at the effect of online homophily rather 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 media exposure.

[Table 7 about here.]

4.4.5 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 in our paper: Table A.15 documents no significant effects for voting for Republicans/Democrats in the most saturated specifications. With the magnitude of the effect being 0.001, we can rule out the effects of up to 0.57%, with a mean of dependent variable being 66.4%, thus our results are close to being precisely estimated zeroes.

Similarly, we do not find any significant evidence for turnout. However, the results, reported in Table A.16, are pretty noisy and are far from 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 the reduction in extreme margins of voting and extreme political preferences.

4.5 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 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.

4.6 Gmail Homophily Shock and Long Ties

One potential positive effect of social media is that they help create "long ties" which have been associated with important 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).

Our results in Table A.18 indicate that our Gmail Homophily Shock leads to a reduction in the fraction of long ties. To ease interpretation we express the dependent variable in standard deviations. All columns in Table A.18 point to a negative and statistically significant impact of the Gmail Homophily Shock on the fraction of long ties in a county. Once we control for 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 our Gmail Homophily Shock. These results rule out that the negative impact of online homophily on Facebook might have positive effects through the creation of long ties.

4.7 Heterogeneity of Effects

By construction, our main variable of interest, Gmail Homophily Shock, presumes that at least some connections that people have on Facebook are out-of-county connections. Moreover, our results, theoretically, should be stronger if the share of out-of-county connections is higher. In this subsection, we formally test this claim, with a caveat that the share of online connections might be an endogenous variable.

We 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. As one can see, in the most saturated specifications (columns 3-6), the relationship between Gmail Homophily Shock and the share of outside connections is small and insignificant. In column 6, we can rule out effects of up to 0.47%, with a mean of the dependent variable being 56.1%. Thus, even though in principle the share of outside connection 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 the first stage and major outcomes of our analysis (Facebook visits, bar visits, economic connectedness, political homogeneity, extreme vote margin). In all the specifications, the interaction term with the share of outside links has the same sign 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 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 1% level.

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

5 Discussion and Conclusion

In this paper, we examine what happens to communities that are exposed to more like-minded communities through their online network. We exploit a conflict about data sharing between Gmail and Facebook that happened in 2010-2012 to construct exogenous variation in the degree of online homophily between different counties in the US. 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 fundamentally 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 but increases the overall usage of social media. Second, higher online homophily decreases interpersonal contact, as proxied by visits to bars and restaurants, and other locations where people socialize. Third, it leads to a reduction in local social capital. The impact of online homophily in the political arena 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. Importantly, it also leads to a reduction in the prevalence of extreme political positions.

From a social policy standpoint, our results uncover an important and often ignored aspect of higher online homophily. “Death of distance” technologies (Cairncross 2002) are responsible for transforming the world into a global village (Alstyne and Brynjolfsson 2005), increasing the diversity of networks (Eagle et al. 2010) and facilitating the creation of beneficial “long ties” (Jahani et al. 2023). However, the effect of social media crucially depends on the structure of online networks. As long as they promote the creation of homophilic networks, all the potential positive effects may come at the cost of undermining local communities and their social and political cohesion. Policymakers wanting to bring people closer online should keep in mind the trade-offs such technologies introduce for the traditional structure of communities.

References

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. "The Welfare Effects of Social Media." *American Economic Review* 110:629–76.

Alstyne, Marshall Van and Erik Brynjolfsson. 2005. "Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities." *Management Science* 51:851–868.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel. 2018a. "The Economic Effects of Social Networks: Evidence from the Housing Market." *Journal of Political Economy* 126:2224–2276.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018b. "Social Connectedness: Measurement, Determinants, and Effects." *Journal of Economic Perspectives* 32:259–80.

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. 2015. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science* 348:1130–1132.

Bakshy, Eytan, Itamar Rosenn, Cameron Marlow, and Lada Adamic. 2012. "The Role of Social Networks in Information Diffusion." In *Proceedings of the 21st International Conference on World Wide Web*, pp. 519–528.

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

Barrera, Oscar, Sergei Guriev, Emeric Henry, and Ekaterina Zhuravskaya. 2020. "Facts, Alternative Facts, and Fact Checking in Times of Crisis." *American Economic Review* 110:3256–3275.

Bauernschuster, Stefan, Oliver Falck, and Ludger Woessmann. 2014. "Surfing Alone? The Internet and Social Capital: Evidence from an Unforeseeable Technological Mistake." *Journal of Public Economics* 117:73–89.

Bodle, Robert. 2011. "Regimes of Sharing." *Information, Communication & Society* 14:320–337.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. "A 61-Million-Person Experiment In Social Influence And Political Mobilization." *Nature* 489:295–298.

Boxell, Levi, Matthew Gentzkow, and Jesse 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* 114:10612–10617.

Braghieri, Luca, Ro'ee Levy, and Alexey Makarin. 2022. "Social Media and Mental Health." *American Economic Review* 112:3660–93.

Bramoullé, Yann, Sergio Currarini, Matthew Jackson, Paolo Pin, and Brian Rogers. 2012. "Homophily and Long-Run Integration in Social Networks." *Journal of Economic Theory* 147:1754–

1786.

Bursztyn, Leonardo, Georgy Egorov, Ruben Enikolopov, and Maria Petrova. 2024. “Social Media and Xenophobia: Evidence from Russia.” *NBER Working Paper* .

Bursztyn, Leonardo, Benjamin R Handel, Rafael Jimenez, and Christopher Roth. 2023. “When Product Markets Become Collective Traps: The Case of Social Media.” *NBER Working Paper* .

Cairncross, Frances. 2002. “The Death of Distance.” *RSA Journal* 149:40–42.

Campante, Filipe, Ruben Durante, and Francesco Sobbrio. 2017. “Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation.” *Journal of the European Economic Association* 16:1094–1136.

Campante, Filipe, Ruben Durante, and Andrea Tesei. 2022. “Media and Social Capital.” *Annual Review of Economics* 14:69–91.

Cantoni, Enrico and Vincent Pons. 2022. “Does Context Outweigh Individual Characteristics In Driving Voting Behavior? Evidence From Relocations Within The United States.” *American Economic Review* 112:1226–1272.

Chetty, Raj, Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, et al. 2022a. “Social Capital I: Measurement and Associations with Economic Mobility.” *Nature* 608:108–121.

Chetty, Raj, Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, et al. 2022b. “Social Capital II: Determinants of Economic Connectedness.” *Nature* 608:122–134.

ComScore. 2016. “Comscore, [Metric], 1/01/2016 - 31/03/2016 [Database].”

Conover, Michael, Jacob Ratkiewicz, Matthew Francisco, Bruno Goncalves, Filippo Menczer, and Alessandro Flammini. 2021. “Political Polarization on Twitter.” *Proceedings of the International AAAI Conference on Web and Social Media* 5:89–96.

Currarini, Sergio, Matthew O. Jackson, and Paolo Pin. 2009. “An Economic Model of Friendship: Homophily, Minorities, and Segregation.” *Econometrica* 77:1003–1045.

Eagle, Nathan, Michael Macy, and Rob Claxton. 2010. “Network Diversity And Economic Development.” *Science* 328:1029–1031.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. 2020. “Social Media and Protest Participation: Evidence from Russia.” *Econometrica* 88:1479–1514.

Enikolopov, Ruben, Maria Petrova, and Adam Szeidl. 2021. “Homophily Offline and Online Social Networks.” *Working paper* .

Falck, Oliver, Robert Gold, and Stephan Heblich. 2014. “E-lections: Voting Behavior and the Internet.” *American Economic Review* 104:2238–65.

Feld, Scott L. 1982. “Social Structural Determinants of Similarity Among Associates.” *American Sociological Review* pp. 797–801.

Fujiwara, Thomas, Karsten Müller, and Carlo Schwarz. 2023. “The Effect of Social Media on Elections: Evidence from the United States.” *Journal of the European Economic Association* p. jvad058.

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

Gentzkow, Matthew. 2006. “Television and Voter Turnout.” *The Quarterly Journal of Economics* 121:931–972.

Gentzkow, Matthew and Jesse M. Shapiro. 2011. “Ideological Segregation Online And Offline.” *The Quarterly Journal of Economics* 126:1799–1839.

Geraci, Andrea, Mattia Nardotto, Tommaso Reggiani, and Fabio Sabatini. 2022. “Broadband Internet and Social Capital.” *Journal of Public Economics* 206:104578.

Granovetter, Mark S. 1973. “The Strength of Weak Ties.” *American Journal of Sociology* 78:1360–1380.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2004. “The Role of Social Capital in Financial Development.” *American Economic Review* 94:526–556.

Guiso, Luigi, Luigi Zingales, and Paola Sapienza. 2016. “Long-Term Persistence.” *Journal of the European Economic Association* 14:1401–1436.

Guriev, Sergei, Emeric Henry, Théo Marquis, and Ekaterina Zhuravskaya. 2023. “Curtailing False News, Amplifying Truth.” Pse working papers, HAL.

Guriev, Sergei, Nikita Melnikov, and Ekaterina Zhuravskaya. 2021. “3G Internet and Confidence in Government.” *The Quarterly Journal of Economics* 136:2533–2613.

Halberstam, Yosh and Brian 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, Emeric, Ekaterina Zhuravskaya, and Sergei Guriev. 2022. “Checking And Sharing Alt-Facts.” *American Economic Journal: Economic Policy* 14:55–86.

Jackson, Matthew O. 2008. *Social and Economic Networks*. Princeton University Press.

Jahani, Eaman, Samuel P. Fraiberger, Michael Bailey, and Dean Eckles. 2023. “Long Ties, Disruptive Life Events, And Economic Prosperity.” *Proceedings Of The National Academy Of Sciences* 120:e2211062120.

Kaplan, Ethan, Jörg L. Spenkuch, and Rebecca Sullivan. 2022. “Partisan Spatial Sorting in the United States: A Theoretical and Empirical Overview.” *Journal of Public Economics* 211:104668.

Langtry, Alastair. 2023. “More Connection, Less Community: Network Formation and Local Public Goods Provision.” *Working paper* .

Leip, David. 2021. *Dave Leip's Atlas of U.S. Presidential Elections.*

Levy, Ro'ee. 2021. "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment." *American Economic Review* 111:831–70.

Manacorda, Marco and Andrea Tesei. 2020. "Liberation Technology: Mobile Phones and Political Mobilization in Africa." *Econometrica* 88:533–567.

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2021. "IPUMS National Historical Geographic Information System: Version 16.0 [Database]."

Melnikov, Nikita. 2022. "Mobile Internet and Political Polarization." Technical report.

Müller, Karsten and Carlo Schwarz. 2020. "Fanning The Flames Of Hate: Social Media And Hate Crime." *Journal Of The European Economic Association* 19:2131–2167.

Müller, Karsten and Carlo Schwarz. 2023. "From Hashtag To Hate Crime: Twitter And Antiminority Sentiment." *American Economic Journal: Applied Economics* 15:270–312.

Muraskin, William. 1974. "The Moral Basis of a Backward Sociologist: Edward Banfield, the Italians, and the Italian-Americans." *American Journal of Sociology* 79:1484–1496.

Nyhan, Brendan, Jaime Settle, Emily Thorson, Magdalena Wojcieszak, Pablo Barberá, Annie Y. Chen, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, Sandra González-Bailón, Andrew M. Guess, Edward Kennedy, Young Mie Kim, David Lazer, Neil Malhotra, Devra Moehler, Jennifer Pan, Daniel Robert Thomas, Rebekah Tromble, Carlos Velasco Rivera, Arjun Wilkins, Beixian Xiong, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker. 2023. "Like-Minded Sources on Facebook Are Prevalent but Not Polarizing." *Nature* 620:137–144.

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

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

Rajkumar, Karthik, Guillaume Saint-Jacques, Iavor Bojinov, Erik Brynjolfsson, and Sinan Aral. 2022. "A Causal Test of the Strength of Weak Ties." *Science* 377:1304–1310.

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

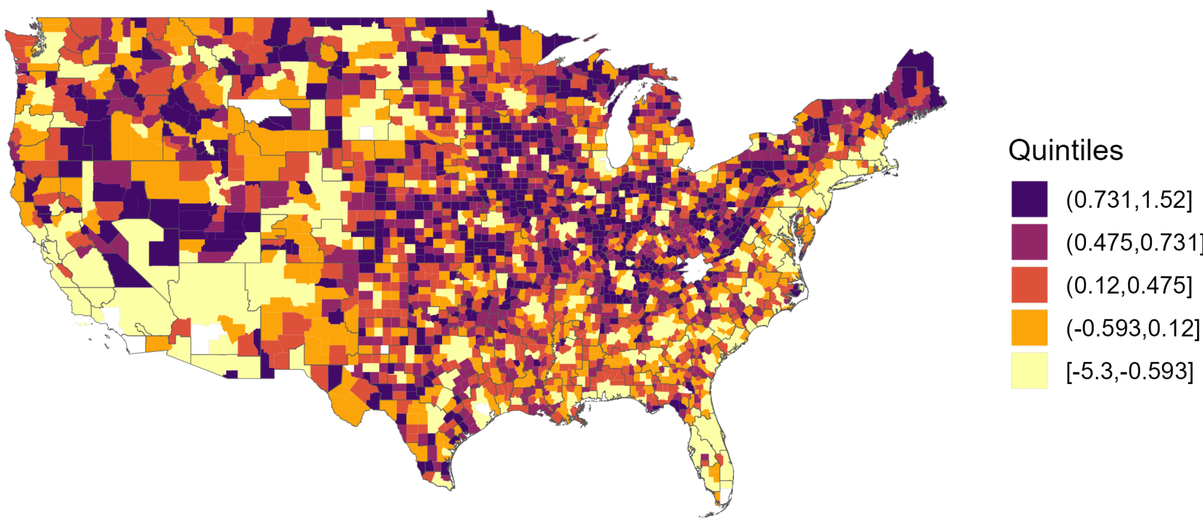
Sunstein, Cass R. 2007. *Republic.com 2.0*. Princeton University Press.

Tarbush, Bassel and Alexander Teytelboym. 2012. "Homophily In Online Social Networks." In *Internet And Network Economics*, edited by Paul W. Goldberg, pp. 512–518, Berlin, Heidelberg. Springer Berlin Heidelberg.

Ugander, Johan, Lars Backstrom, Cameron Marlow, and Jon Kleinberg. 2012. "Structural Diversity in Social Contagion." *Proceedings of the National Academy of Sciences* 109:5962–5966.

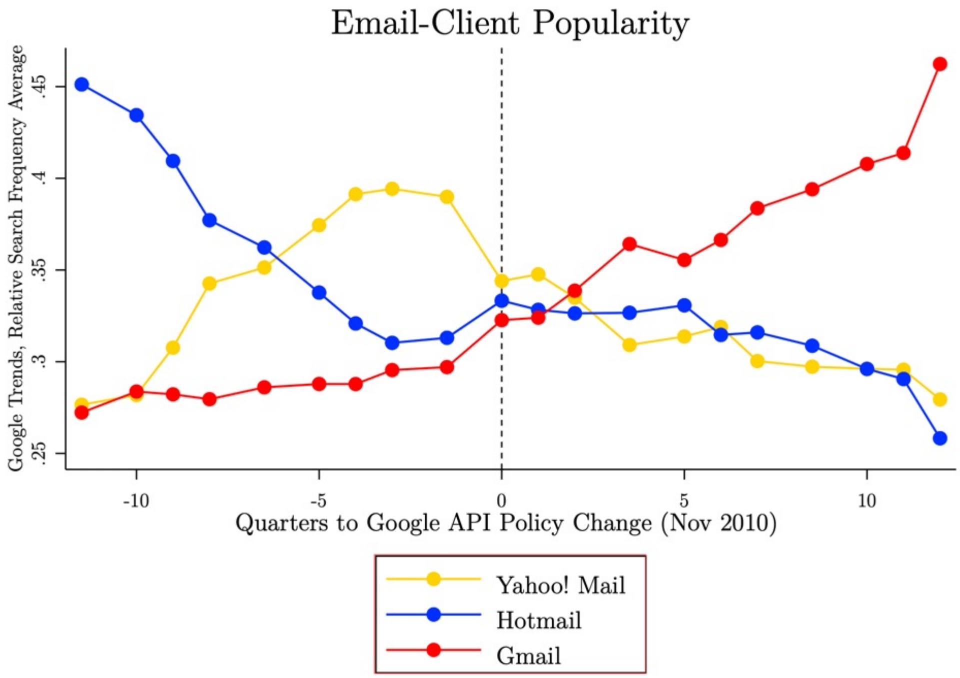
Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. 2020. "Political Effects of the Internet and Social Media." *Annual Review of Economics* 12:415–438.

Figure 1: Online Homophily, 2016



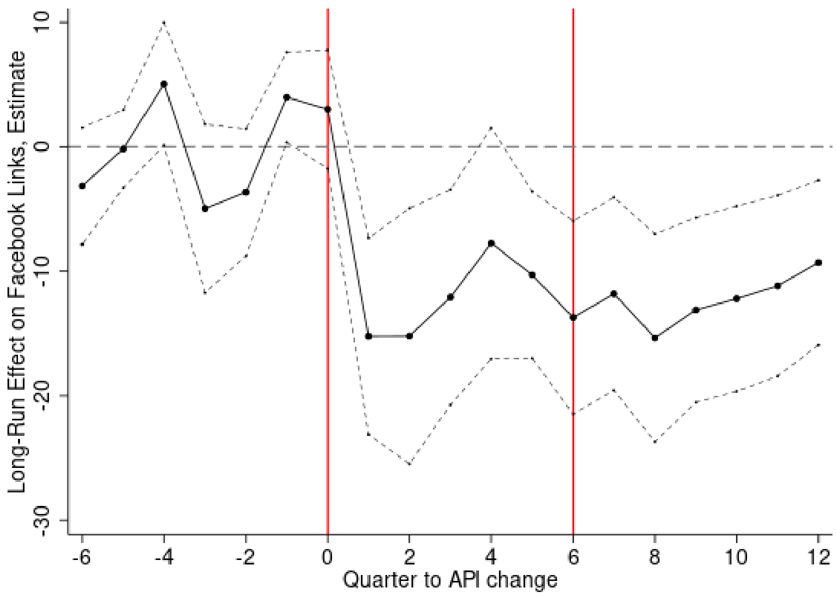
Notes: The map plots the geographic distribution of Online Homophily in 2016 across US counties. We construct Online Homophily in two steps. First, we generate an index of social similarity between US counties by taking the inverse of the principal component analysis of twelve variables: log population, 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



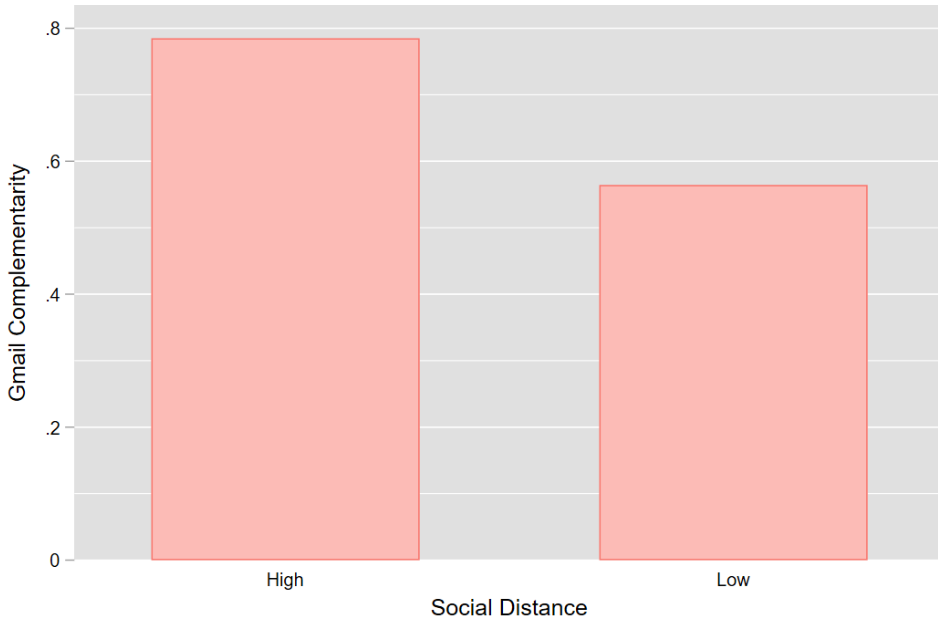
Notes: The figure plots the popularity of the main email clients in the US, quarterly between 2006 and 2016. The source of the data is Google Trends and popularity is measured as the average search frequency of a given email client in a DMA. We focus on the three largest email clients: Yahoo! Mail, Hotmail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

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



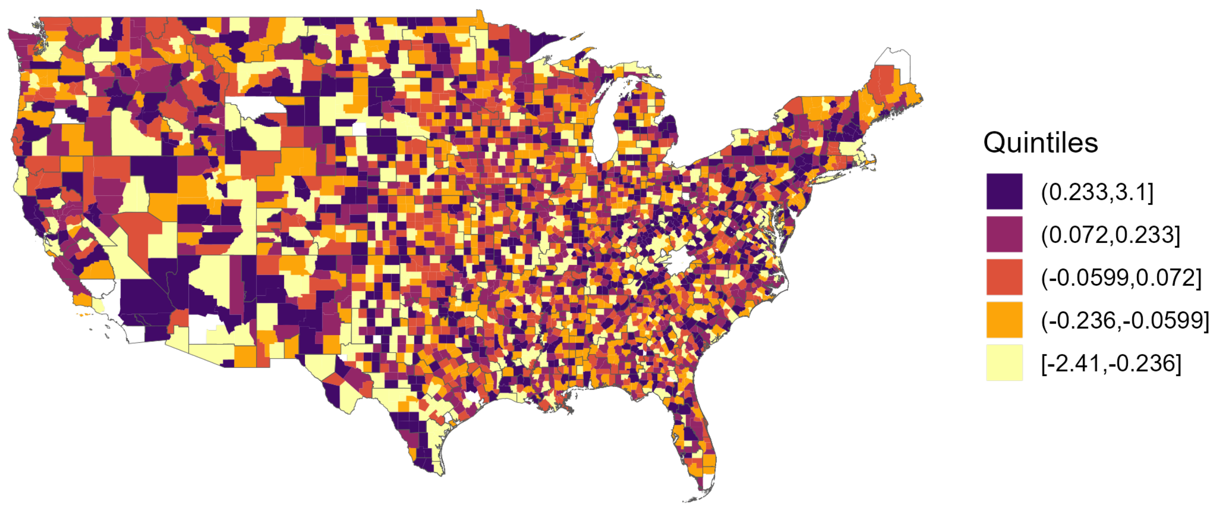
Notes: The graph plots the effect of Gmail complementarity by quarter on 2016 county-pairs Facebook links. We regress and plot the estimates of separate linear models where the outcome variable is the inverse hyperbolic sine of the relative friendship index between county pairs in 2016. Each estimated 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 Complementarity by High- and Low-Social Distance in Blount County, Alabama



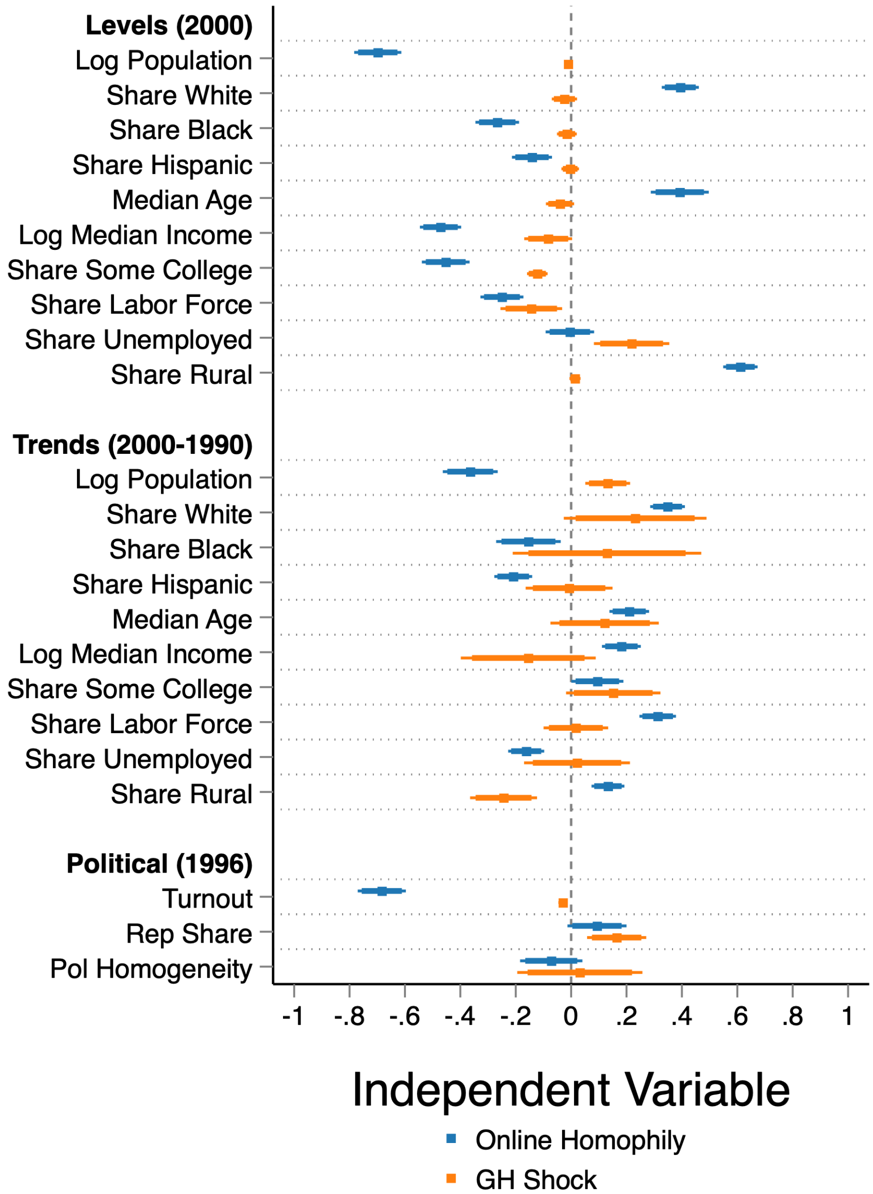
Notes: The figure plots the average Gmail complementarity by social distance for Blount County, Alabama. We denote a county to have high social distance if it belongs to the top tercile of the social distance distribution, whereas we denote the county to have low social distance if it belongs to the bottom tercile of the social distance distribution. We define social distance as the principal component of the difference in absolute value of twelve variables: log population, share White, share Black, share Hispanic, share with at least some college, share unemployed, log median income, share in labor force, share rural, median age in 2010, share of Republican votes in 2004, and their ideology score (county average self-assessed ideology from Gallup polls of 2008-2010 that ranges from 1 very liberal through 5 very conservative). See section 2 for more details.

Figure 5: Gmail Homophily Shock



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

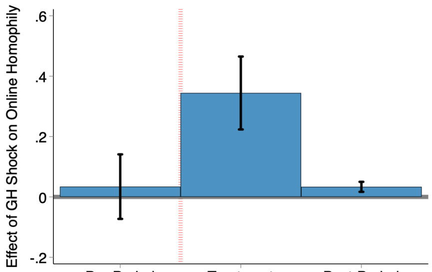
Figure 6: Balance Tests of the Gmail Homophily Shock



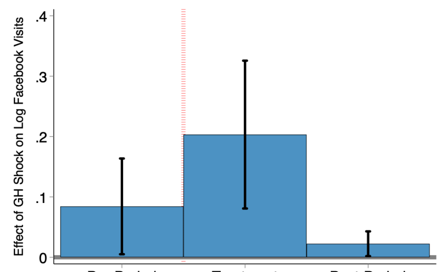
Notes: The figure plots balance tests of our Gmail Homophily Shock compared to the correlation between our endogenous variable, Online Homophily, and pre-determined county observables. The blue bars show the correlation between Online Homophily and a host of predetermined socio-economic and political county-level controls shown on the Y-axis. The orange bars report balance tests of the Gmail Homophily Shock, our standardized excluded instrument. The specification for the balance tests 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. We cluster standard errors at the state level.

Figure 7: Gmail Complementarity Has No Effect Outside Treatment Window

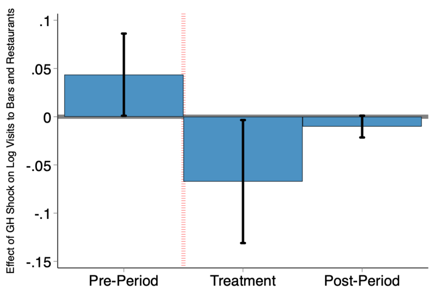
(a) First-Stage



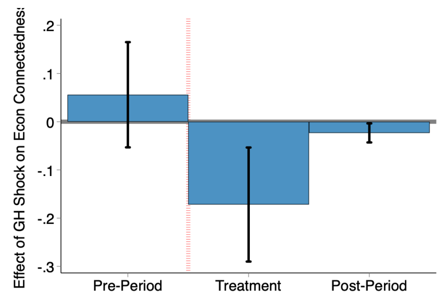
(b) Log Facebook Visits



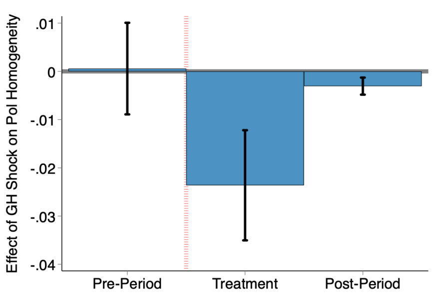
(c) Log Bars and Restaurants Visits



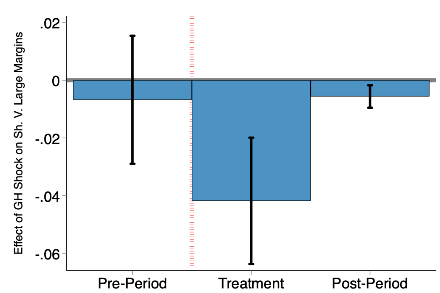
(d) Economic Connectedness



(e) Political Homogeneity

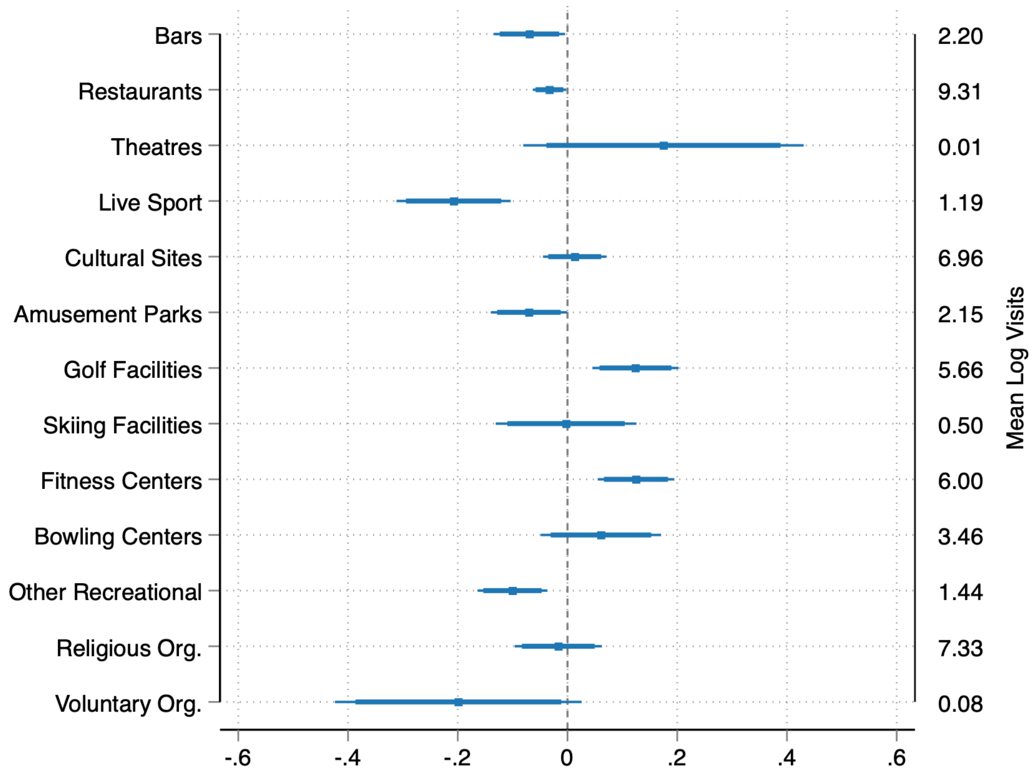


(f) Share Extreme Margins



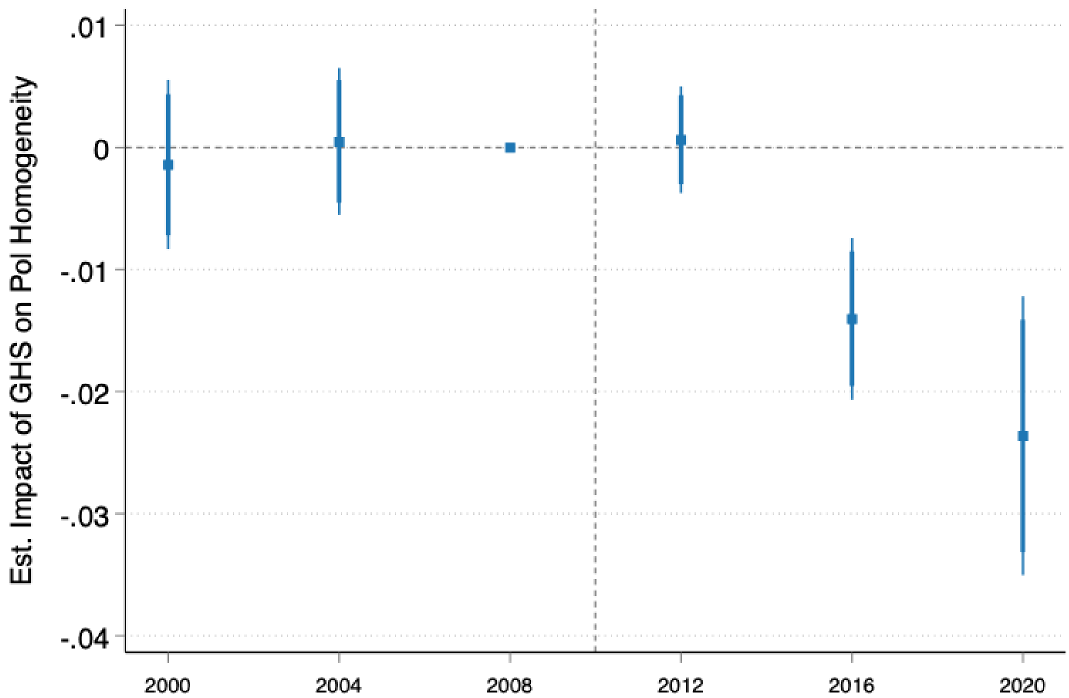
Notes: The figure plots the effect of the cumulative Gmail complementarity on our outcome variables in three different windows: the pre-period, the treatment period, and the post-period. We build the cumulative Gmail complementarity in the pre-period using the six quarters before the API change. The cumulative Gmail complementarity in the treatment window is our Gmail Homophily Shock which is constructed using the six quarters after API change. Similarly, we define the post-period window using the six quarters following the end of the Google-Facebook incident. We plot bars and confidence intervals from our most saturated specification which includes the following controls: log population, share White, share with at least some college and share unemployed, share Black, share Hispanic, log median income, share in labor force, share rural and median age as of 2010; turnout, political homogeneity and Republican shares as of 2008; all trends defined as the difference for all socio-demographic controls between 2000 and 2010 as well as DMA fixed effects. We include the pre-period Gmail complementarity as a control when we regress our outcomes on the treatment and post-period Gmail complementarity. We cluster standard errors at the state level. Figure A.12 displays results using a sparser specification including only DMA FEs and baseline controls (same as in column 3 of Table 1).

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



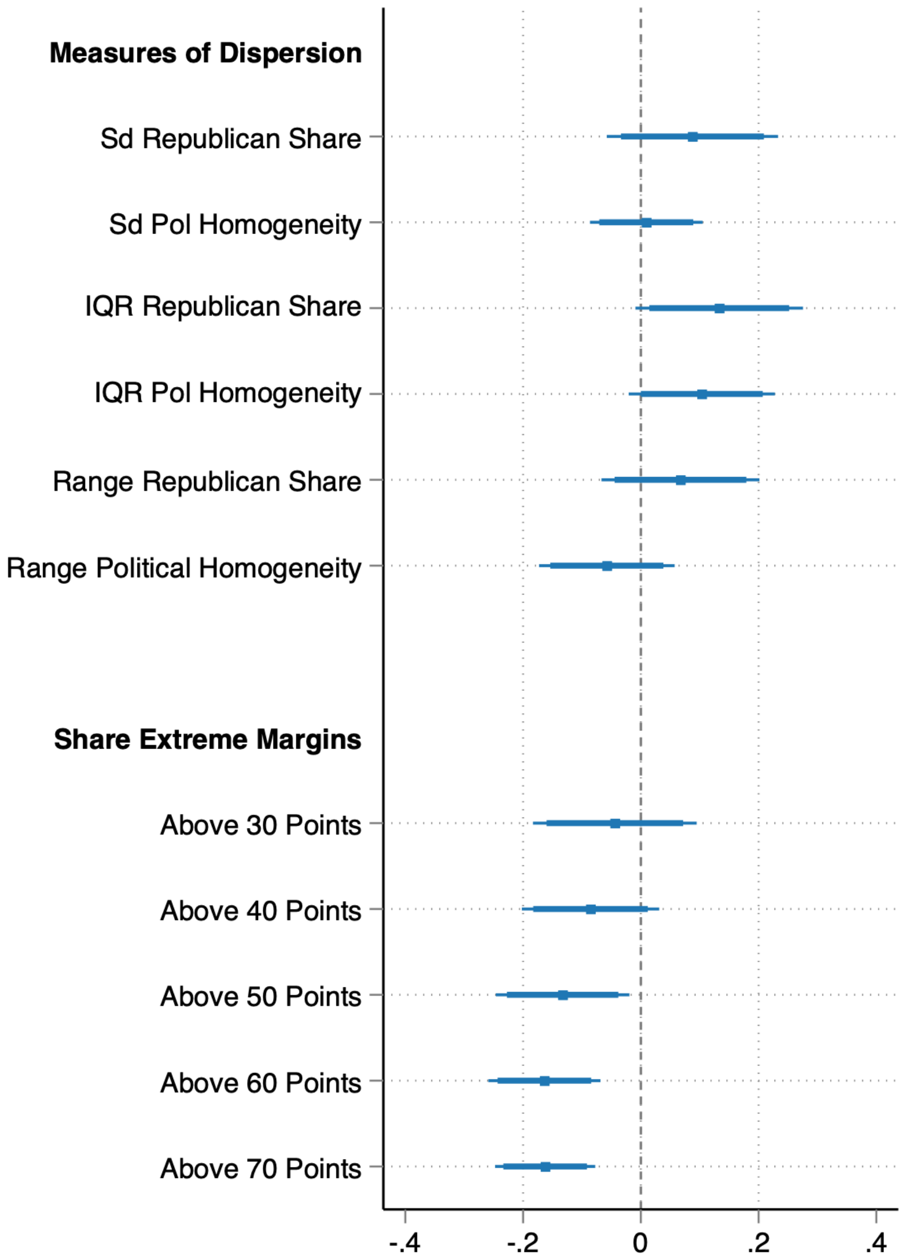
Notes: The figure plots the estimated effect of the Gmail Homophily Shock on the log number of visits for several venues of interaction. The left y-axis shows the venue of interaction while the right y-axis presents the average number of visits for each venue in logs (Appendix Table A.2 shows means in levels). The source of the data is Safegraph and we reconstruct 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 Association (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 9: 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 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 is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Controls include the pre-period Gmail complementarity using the last six quarters before the API changed; log population, share White, share with at least some college and share unemployed, share Black, share Hispanic, log median income, share in labor force, share rural and median age as of 2010; turnout, political homogeneity and Republican shares as of 2008; 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 10: 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						
	2016						2020
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Gmail Homophily Shock	0.625***	0.277***	0.419***	0.411***	0.357***	0.344***	0.305***
	(0.148)	(0.084)	(0.063)	(0.069)	(0.058)	(0.060)	(0.055)
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*	*Yes*
KP F-Stat	17.732	10.914	43.657	35.280	37.285	32.856	30.920
Adj R2	0.572	0.719	0.810	0.819	0.827	0.834	0.844
Mean of Dep. Var.	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	3042	3042	3042	3042	3042	3042	3042

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. 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 is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications, we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Baseline Controls include basic demographic and political county characteristics: share White, share 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)
Gmail Homophily Shock	0.262***	0.226***	0.205***	0.203***	-0.078**	-0.081*	-0.099**	-0.103**
	(0.059)	(0.055)	(0.056)	(0.061)	(0.035)	(0.044)	(0.045)	(0.046)
Mean of Dep. Var.	4.853	4.853	4.853	4.853	2.391	2.391	2.391	2.391
Adj R2	0.866	0.866	0.867	0.867	0.841	0.842	0.842	0.842
Observations	2872	2872	2872	2872	2872	2872	2872	2872
Instrumental Variable Estimates								
Online Homophily	0.690***	0.629***	0.658***	0.661***	-0.206**	-0.225*	-0.319**	-0.334**
	(0.127)	(0.145)	(0.170)	(0.182)	(0.102)	(0.129)	(0.154)	(0.162)
F-stat	44.568	35.157	31.467	28.251	44.568	35.157	31.467	28.251
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*Yes*	*Yes*	*Yes*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*Yes*	*Yes*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*Yes*	*No*	*No*	*No*	*Yes*

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

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

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

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

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

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

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, covers all of 2019 and varies at the county-by-month level. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications, we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online Homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. All specifications control for the total number of 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:	Economic Connectedness					
	(1)	(2)	(3)	(4)	(5)	(6)
Gmail Homophily Shock	-0.516***	-0.312***	-0.308***	-0.197***	-0.184***	-0.172***
	(0.098)	(0.056)	(0.058)	(0.058)	(0.063)	(0.059)
Mean of Dep. Var.	0.000	0.000	0.000	0.000	0.000	0.000
Adj R2	0.396	0.705	0.816	0.863	0.863	0.872
Observations	2943	2943	2943	2943	2943	2943
Instrumental Variable Estimates						
Online Homophily	-0.881***	-1.158***	-0.749***	-0.485**	-0.535**	-0.511**
	(0.333)	(0.264)	(0.202)	(0.188)	(0.230)	(0.230)
F-stat	17.199	11.266	48.436	38.553	36.399	30.942
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*

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

Table 6: Online Homophily and Political Homogeneity, 2020

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

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

Table 7: Gmail Complementarity Shock Reduces Extreme Partisan Identity

Dep. Variable:	Extreme Partisanship					
	(1)	(2)	(3)	(4)	(5)	(6)
Gmail Homophily 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 strong democrat or 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.

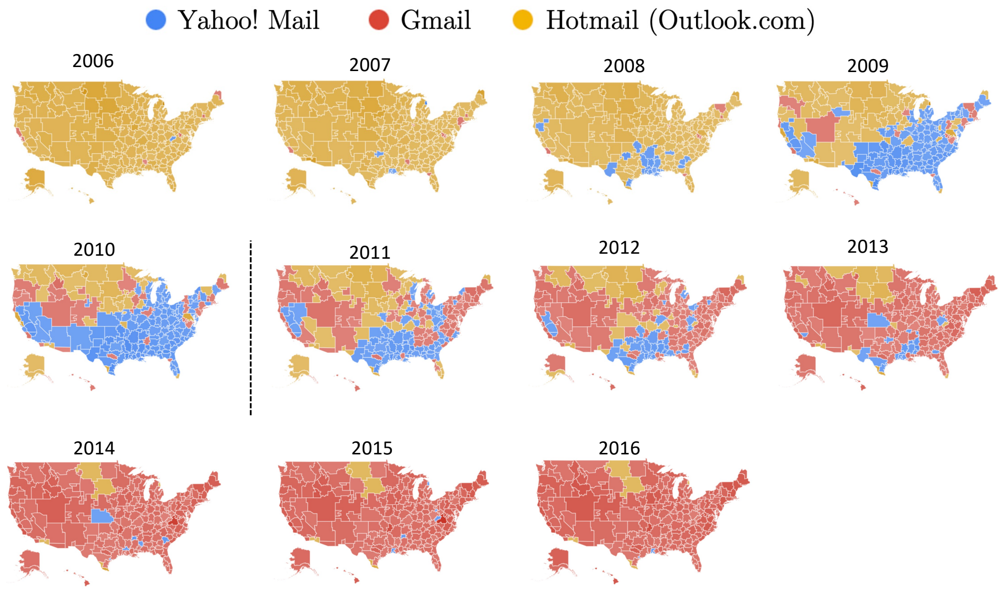
Table 8: The Impact of the Gmail Homophily Shock by Share of Links Outside the County

	First-Stage	Log Facebook Visits	Log Bars and Rest. Visits	Econ Connect	Political Homogeneity	Extreme Margins
	(1)	(2)	(3)	(4)	(5)	(6)
Gmail Homophily Shock	0.234***	0.084	-0.049	-0.123*	-0.019***	-0.032***
	(0.061)	(0.060)	(0.036)	(0.064)	(0.006)	(0.010)
– × Share Out Connections	0.112***	0.121***	-0.021	-0.049***	-0.005***	-0.011***
	(0.012)	(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	0.902	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 connection outside the county. The dependent variables are the six main outcomes of the paper: first-stage in column 1, Log Facebook visits in column 2, Log 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 is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. In line with their respective specifications, column 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)

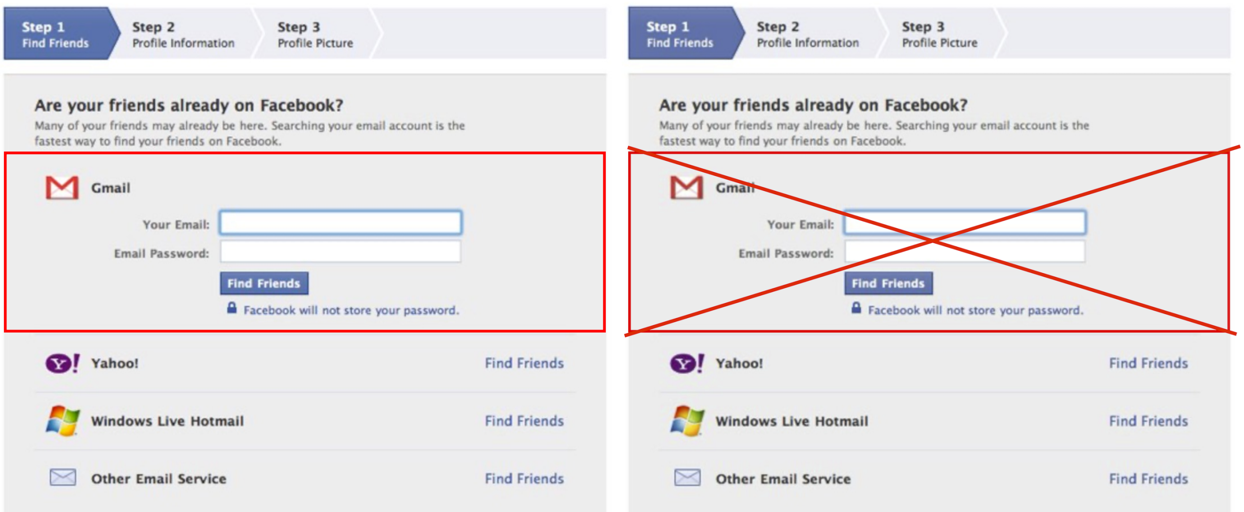
Figure A.1: Geographic Distribution of Email Clients Popularity Across US DMAs



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.

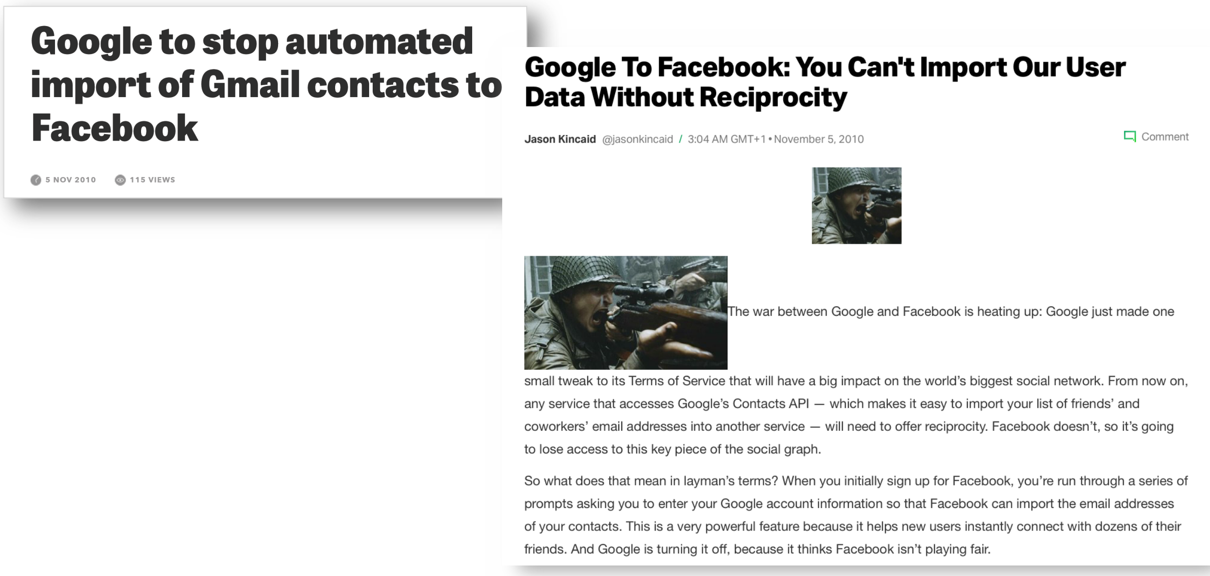
Figure A.2: Exterior Look of Join-Facebook Window, 2009-2013

(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

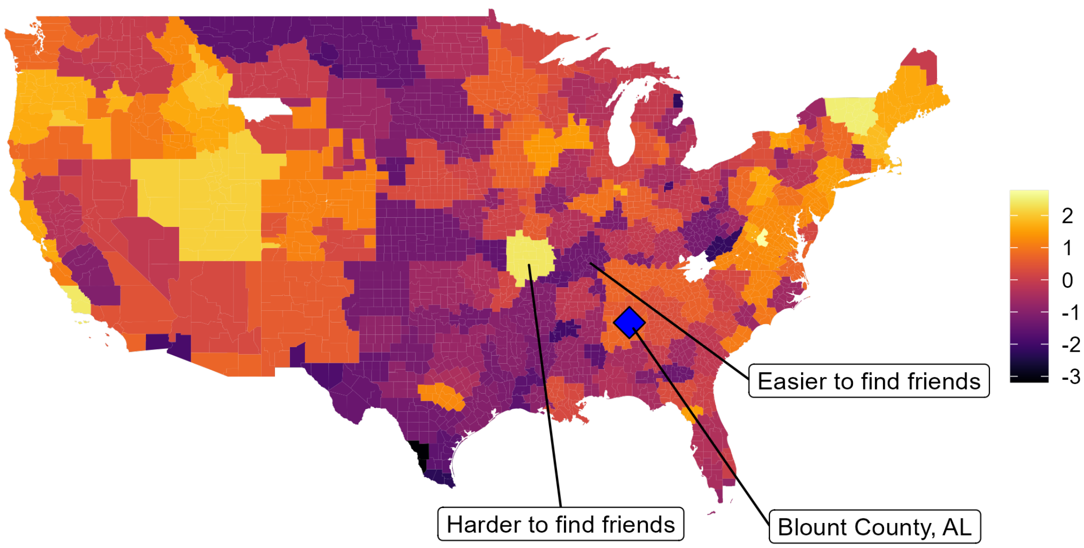
Comment

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.

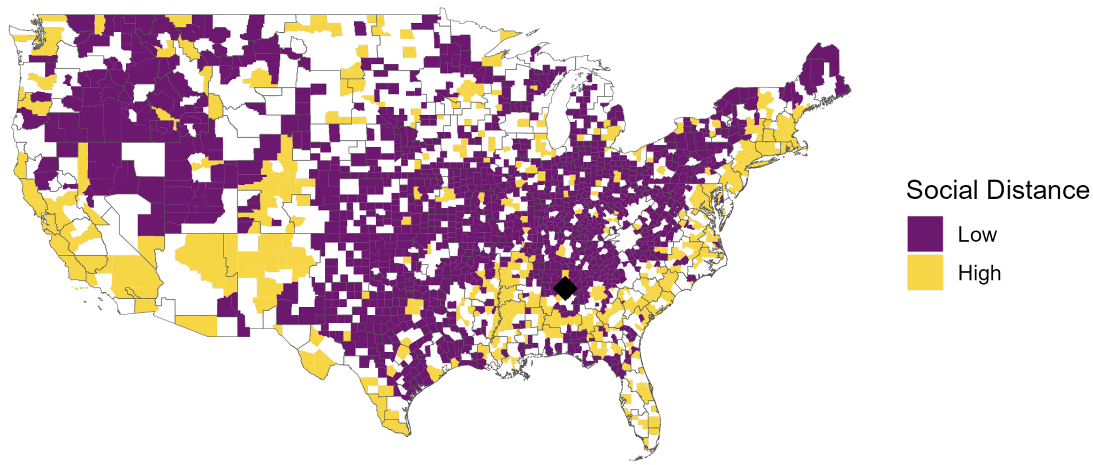
Notes: The figure shows two headlines contemporary to the incident between Google and Facebook giving details on the source of the conflict. The source of the headlines is the Tech blog TechCrunch.com (https://techcrunch.com/2010/11/04/facebook-google-contacts/?guccounter=1)

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 using the top and bottom tercile of the distance distribution.

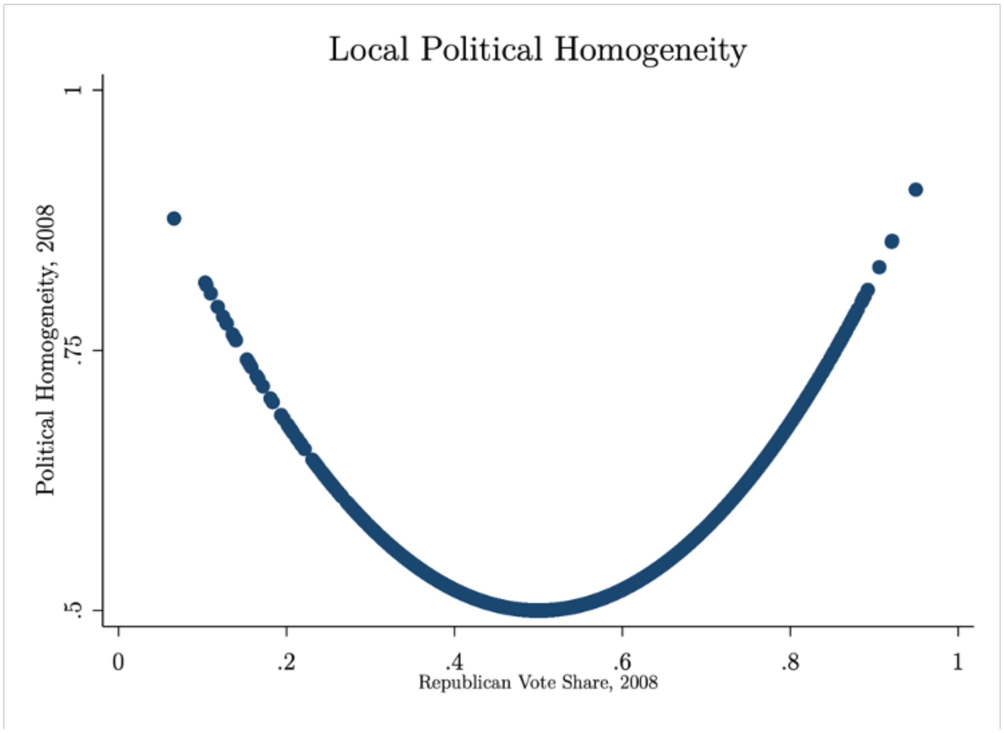
Figure A.6: Gmail Complementarity by High- and Low-Social Distance by County in Alabama

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.7: Gmail Homophily Shock by County in Alabama

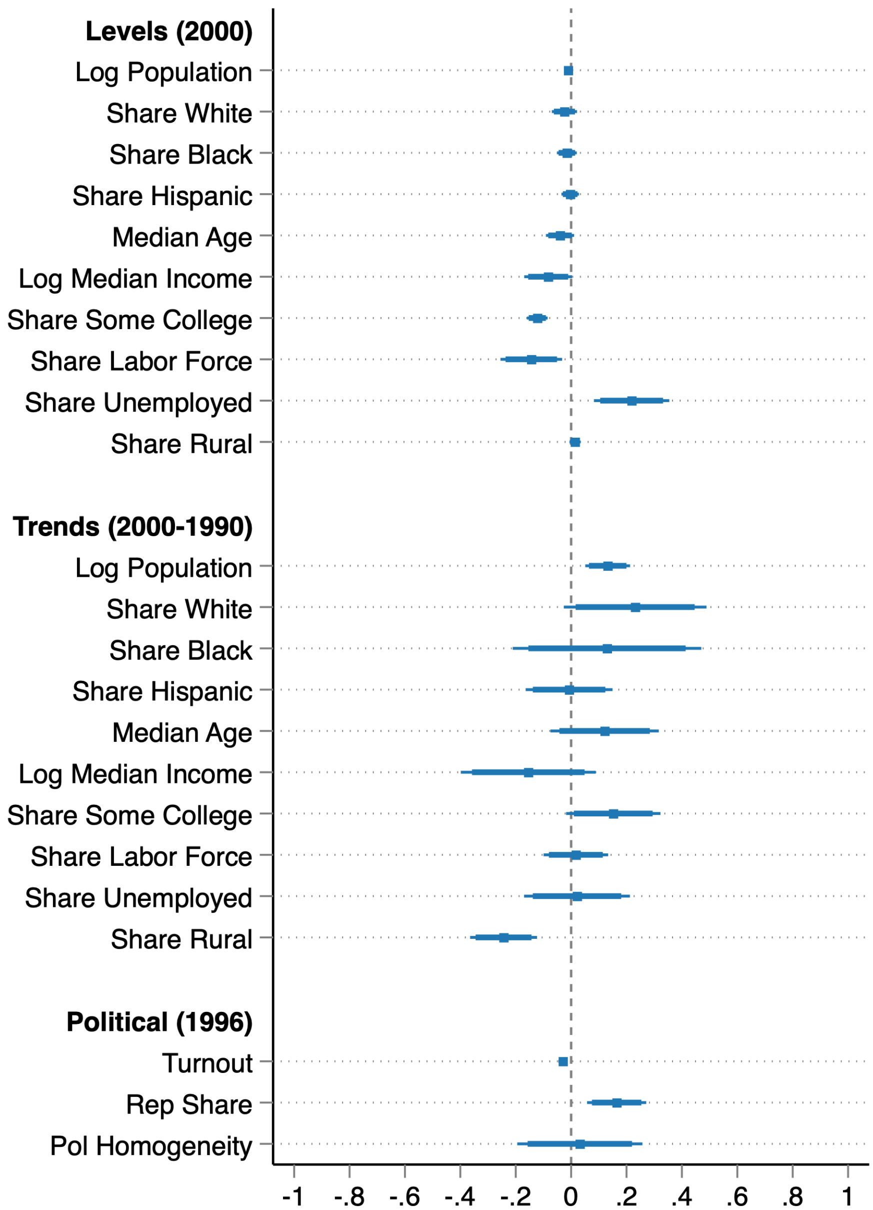
Notes: The figure plots the Gmail Homophily Shock for each county in Alabama. The Gmail Homophily Shock is calculated as the difference in the average Gmail complementarity among high- and low-social distance counties. The average Gmail complementarity by social distance is plotted in Figure A.6 for all the counties in Alabama.

Figure A.8: Political Homogeneity and Vote Shares



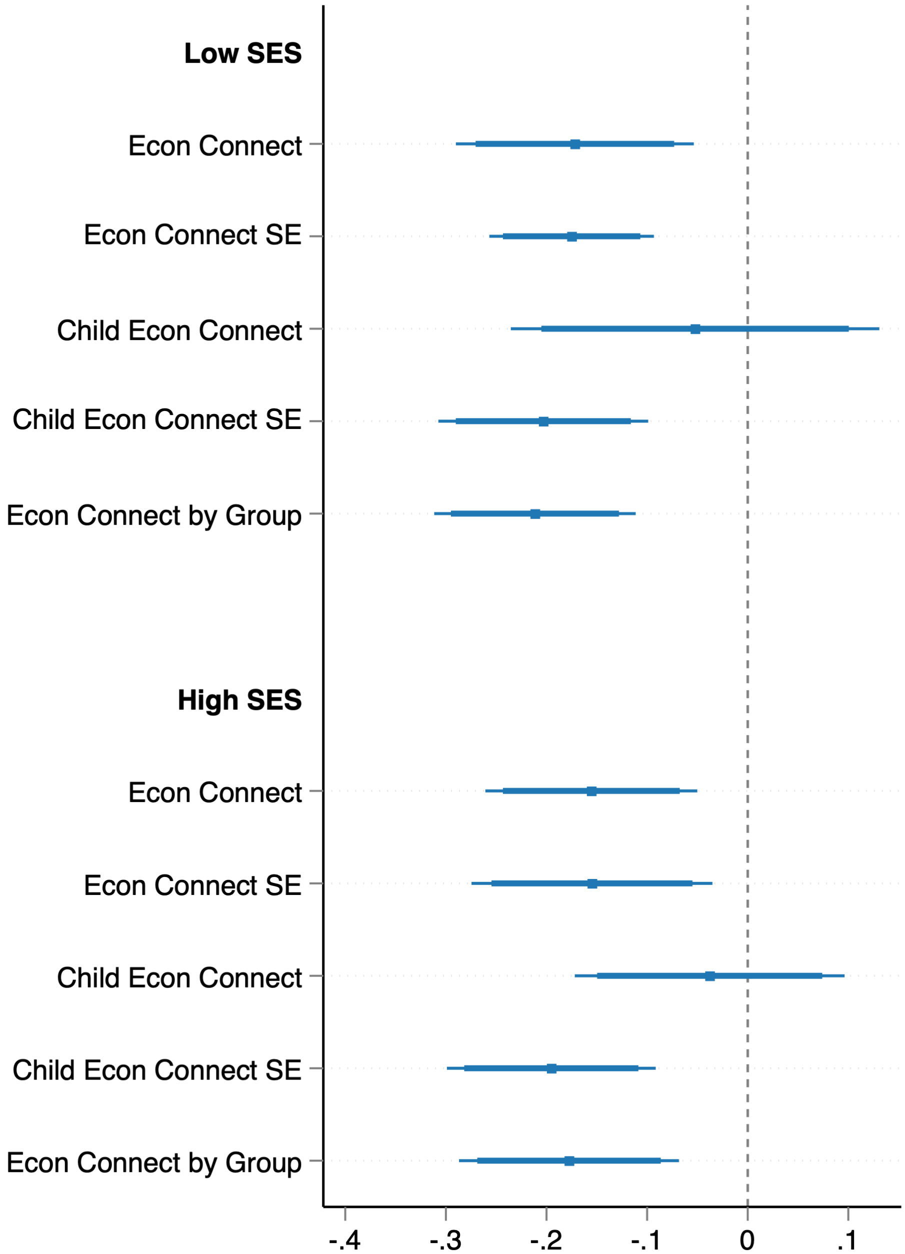
Notes: The figure plots the functional relationship between political homogeneity and the vote share for the Republican party (or any bipartisan electoral system). We construct political homogeneity as the opposite of local political fractionalization, and it is computed as $PolHomogeneity_{it} = 1 -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

Figure A.9: Balance Tests of the Gmail Homophily Shock



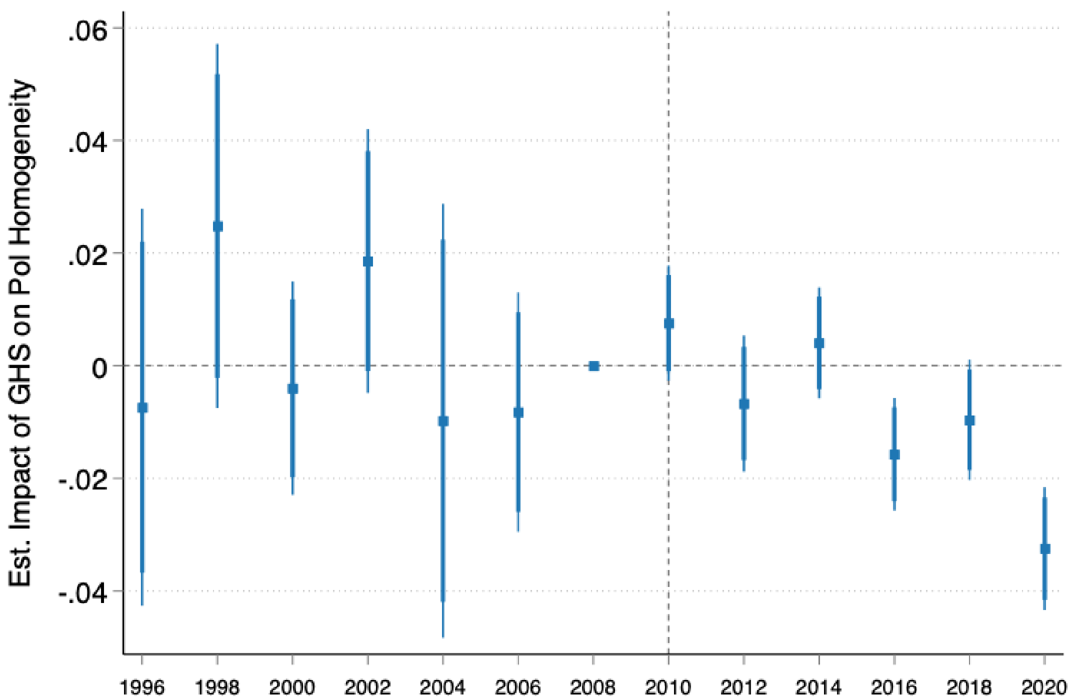
Notes: Balance tests of our Gmail Homophily Shock. We plot the estimated coefficients from separate regressions of our baseline model in equation (4) where we test the balancedness of our Gmail Homophily Shock on a host of predetermined socio-economic and political county-level controls shown on the Y-axis. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. We show the results from 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. We cluster standard errors at the state level.

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.

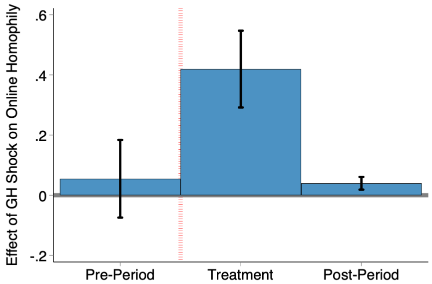
Figure A.11: Gmail Homophily Shock and Political Homogeneity, House Elections



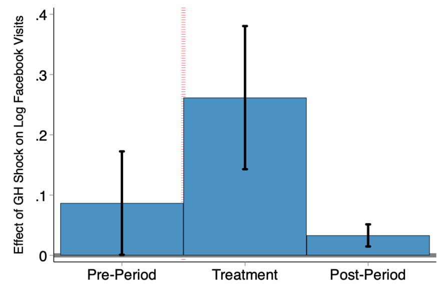
Notes: Event study analysis of the impact of Gmail Homophily Shock on political homogeneity in congressional house elections. We plot the estimated coefficients associated with the effect of one standard deviation increase in the Gmail Homophily Shock on political homogeneity every two years between 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 is our standardized excluded instrument measuring 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

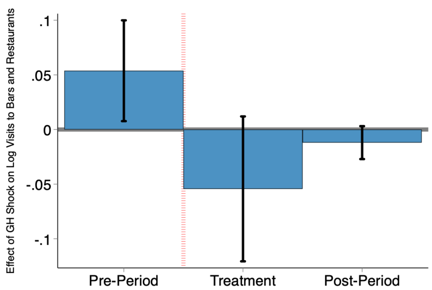
(a) First-Stage



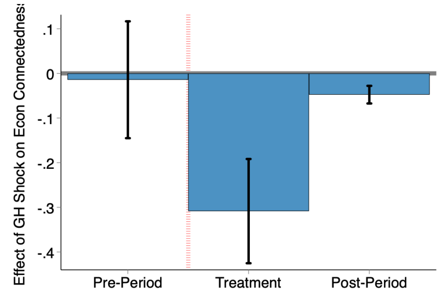
(b) Log Facebook Visits



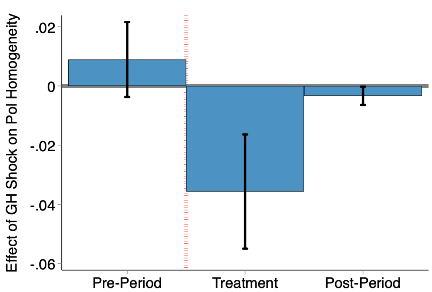
(c) Log Bars and Restaurants Visits



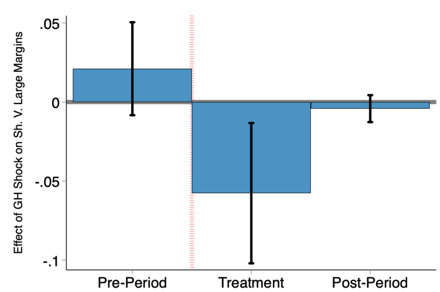
(d) Economic Connectedness



(e) Political Homogeneity



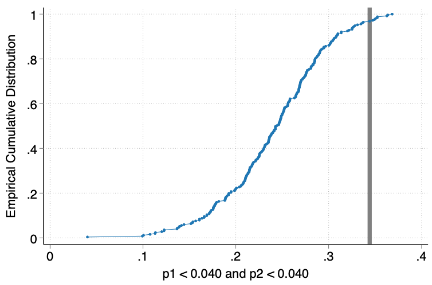
(f) Share Extreme Margins



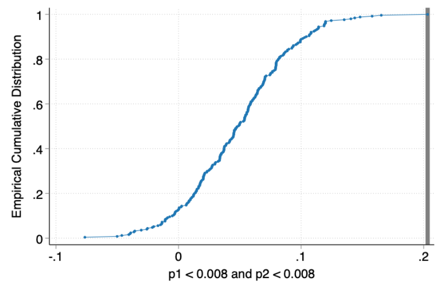
Notes: The figure plots the effect of the cumulative Gmail complementarity on our outcome variables in three different windows: the pre-period, the treatment period and the post-period. We build the cumulative Gmail complementarity in the pre-period using the six quarters prior to the API change. The 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 7 displays results using our most saturated specification (same as in column 6 of Table 1)

Figure A.13: Randomization Analysis, Main Results

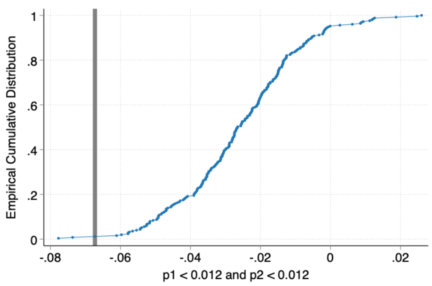
(a) First-Stage



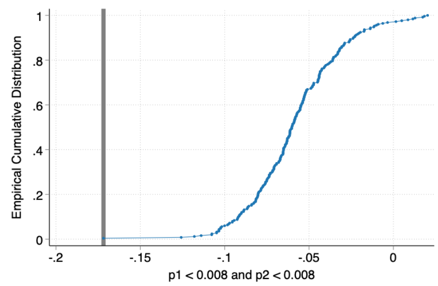
(b) Log Facebook Visits



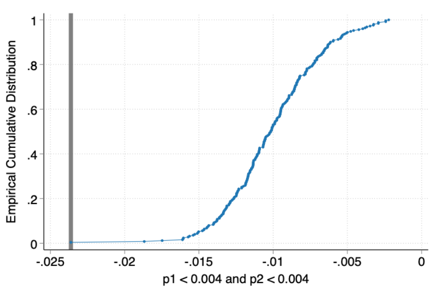
(c) Log Bars and Restaurants Visits



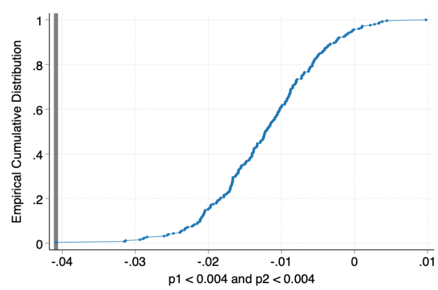
(d) Economic Connectedness



(e) Political Homogeneity



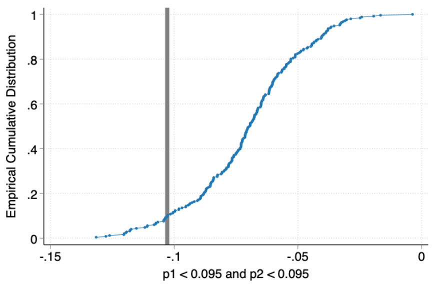
(f) Share Extreme Margins (70 points)



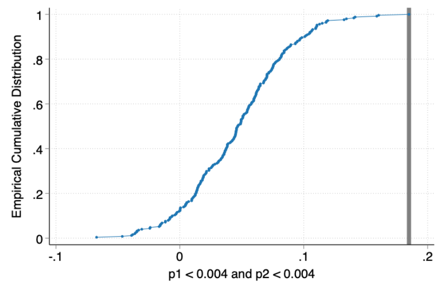
Notes: The figure plots the results of our randomization analysis for our main results. We plot the empirical cumulative distribution obtained by randomly shuffling the observed email distribution and computing our reduced-form analysis on the simulated data 250 times. The black solid line indicates 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

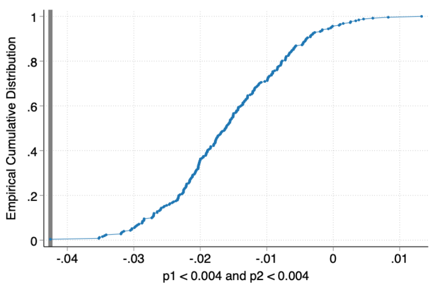
(a) Log Other Social Media Visits



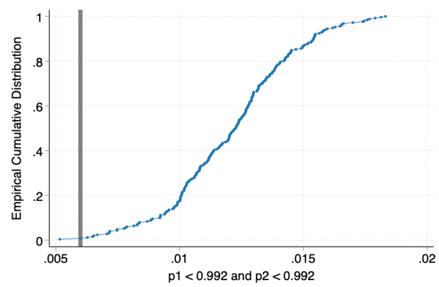
(b) Log Any Social Media Visits



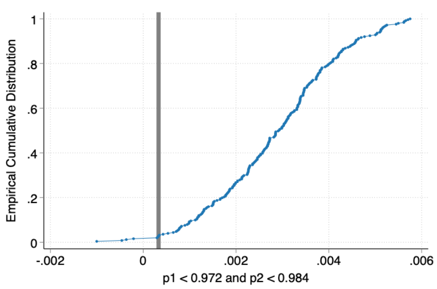
(c) Vote Margins



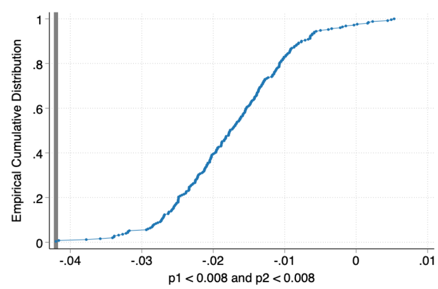
(d) Standard Deviation Republican Vote Share



(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
Main Independent and Control Variables					
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	-5.31	1.52	3042
Total Facebook Links (000s)	7166.38	7377.48	282.45	85417.46	3042
Share Facebook Links Outside County	0.56	0.16	0.09	0.96	3042
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
2000-2010 Changes					
Population (000s)	8.05	30.28	-240.58	644.25	3042
Share White	-0.01	0.03	-0.31	0.23	3042
Share Black	0.00	0.02	-0.14	0.28	3042
Share Hispanic	0.02	0.02	-0.05	0.22	3042
Median Income (000s)	10.18	5.06	-6.12	41.42	3042
Share Some College	0.05	0.02	-0.07	0.22	3042
Share Labor Force	0.01	0.03	-0.15	0.25	3042
Share Unemployed	0.03	0.03	-0.12	0.18	3042
Share Rural	-0.01	0.06	-0.80	0.65	3042
Median Age	3.01	1.81	-5.10	13.50	3042
Political Homogeneity, 2008	0.55	0.06	0.50	0.90	3042
Republican Share, 2008	0.58	0.14	0.07	0.95	3042
Turnout, 2008 (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 (000s)	46.48	106.70	0.16	1210.51	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 (000s)	44.88	103.03	0.15	1170.40	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
CES 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)
Gmail Homophily Shock	0.047	0.114	0.126	0.138*	0.030	0.105	0.119	0.128
	(0.089)	(0.079)	(0.083)	(0.078)	(0.106)	(0.094)	(0.099)	(0.091)
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
Instrumental Variable Estimates								
Online Homophily	0.109	0.269	0.344	0.382*	0.069	0.248	0.323	0.355
	(0.201)	(0.178)	(0.218)	(0.213)	(0.241)	(0.210)	(0.256)	(0.243)
F-stat	50.706	40.147	36.980	33.044	50.706	40.147	36.980	33.044
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*Yes*	*Yes*	*Yes*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*Yes*	*Yes*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*Yes*	*No*	*No*	*No*	*Yes*

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

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

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

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

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

Table A.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)
Gmail Homophily Shock	0.321***	0.297***	0.255***	0.256**	-0.167**	-0.151*	-0.184**	-0.197**
	(0.088)	(0.087)	(0.088)	(0.096)	(0.064)	(0.083)	(0.087)	(0.090)
Mean of Dep. Var.	7.960	7.960	7.960	7.960	3.478	3.478	3.478	3.478
Adj R2	0.782	0.783	0.783	0.783	0.761	0.763	0.763	0.763
Observations	2872	2872	2872	2872	2872	2872	2872	2872
Instrumental Variable Estimates								
Online Homophily	0.807***	0.783***	0.777***	0.793**	-0.420**	-0.400*	-0.561**	-0.609**
	(0.215)	(0.247)	(0.286)	(0.316)	(0.157)	(0.216)	(0.260)	(0.270)
F-stat	46.340	36.590	32.640	28.763	46.340	36.590	32.640	28.763
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*Yes*	*Yes*	*Yes*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*Yes*	*Yes*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*Yes*	*No*	*No*	*No*	*Yes*

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

Table A.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)
Gmail Homophily 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)
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
Instrumental Variable Estimates								
Online Homophily	0.663***	0.608***	0.630***	0.642***	0.761***	0.746***	0.734***	0.764**
	(0.127)	(0.141)	(0.166)	(0.178)	(0.196)	(0.225)	(0.263)	(0.295)
F-stat	45.242	35.665	31.973	28.722	46.340	36.590	32.640	28.763
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*Yes*	*Yes*	*Yes*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*Yes*	*Yes*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*Yes*	*No*	*No*	*No*	*Yes*

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

Table A.8: Homophily Shock and Offline Visits to Any Venue, 2019

Dep. Variable:	Log Total Visits					
	(1)	(2)	(3)	(4)	(5)	(6)
Gmail Homophily 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)
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
Instrumental Variable Estimates						
Online Homophily	-0.007	0.131	-0.016	0.076	0.084	0.139
	(0.071)	(0.140)	(0.107)	(0.101)	(0.120)	(0.111)
F-stat	17.532	10.861	46.545	37.687	39.842	35.224
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
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*

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

Table A.9: Homophily Shock and Bars and Restaurants Visits, 2019 (IHS)

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

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

Table A.10: Homophily Shock and Visits to Any Venue, 2019 (IHS)

Dep. Variable:	IHS Total Visits					
	(1)	(2)	(3)	(4)	(5)	(6)
Gmail Homophily 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)
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
Instrumental Variable Estimates						
Online Homophily	-0.007	0.131	-0.016	0.076	0.084	0.139
	(0.071)	(0.140)	(0.108)	(0.101)	(0.120)	(0.111)
F-stat	17.532	10.861	46.545	37.687	39.842	35.224
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
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*

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

Table A.11: Homophily Shock and the Total Number of Facebook Connections, 2016

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

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

Table A.12: Homophily Shock and the Total Number of Facebook Connections, 2016 (IHS)

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

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

Table A.13: Main Results Controlling for Total Number of Facebook Links

	First-Stage		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)
Gmail Homophily Shock	0.344***	0.340***	0.203***	0.207***	-0.067**	-0.070**	-0.172***	-0.170***	-0.024***	-0.024***	-0.042***	-0.044***
	(0.060)	(0.060)	(0.061)	(0.061)	(0.032)	(0.030)	(0.059)	(0.058)	(0.006)	(0.006)	(0.011)	(0.011)
Log All Links, 2016		0.174**		-0.119*		0.107		-0.161***		0.010***		0.049***
		(0.082)		(0.065)		(0.086)		(0.051)		(0.003)		(0.015)
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Political Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Trends	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
KP F-Stat	32.856	32.259	28.251	28.396	31.939	30.952	30.942	30.793	32.856	32.259	28.428	29.232
Adj R2	0.834	0.836	0.867	0.867	0.954	0.954	0.872	0.874	0.876	0.877	0.757	0.760
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: first-stage in columns 1 and 2, Log Facebook visits in column 3 and 4, Log Bar visits in columns 5 and 6, Economic Connectedness in columns 7 and 8, political homogeneity in columns 9 and 10 and the share of precinct within a county with electoral margins larger than 70 points in 2016 in columns 11 and 12. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online Homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. 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)
Gmail Homophily 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)
Gmail Homophily Shock	0.111^{***}	0.017^{***}	0.010^{***}	0.004	0.003	0.001
	(0.023)	(0.006)	(0.004)	(0.003)	(0.003)	(0.003)
Mean of Dep. Var.	0.664	0.664	0.664	0.664	0.664	0.664
Adj R2	0.407	0.934	0.960	0.967	0.968	0.971
Observations	3042	3042	3042	3042	3042	3042
Instrumental Variable Estimates						
Online Homophily	0.178^{***}	0.061^{**}	0.024^{***}	0.011	0.008	0.003
	(0.021)	(0.024)	(0.007)	(0.008)	(0.009)	(0.009)
F-stat	17.732	10.914	43.657	35.280	37.285	32.856
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the Republican vote share in the 2020 presidential election. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online Homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. 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)
Gmail Homophily Shock	-0.084***	-0.013	-0.013	0.009	0.012*	0.003
	(0.011)	(0.012)	(0.008)	(0.006)	(0.007)	(0.007)
Mean of Dep. Var.	9.562	9.562	9.562	9.562	9.562	9.562
Adj R2	0.987	0.995	0.997	0.998	0.998	0.998
Observations	3042	3042	3042	3042	3042	3042
Instrumental Variable Estimates						
Online Homophily	-0.134***	-0.045	-0.031	0.023	0.033*	0.008
	(0.032)	(0.039)	(0.020)	(0.016)	(0.018)	(0.021)
F-stat	17.732	10.914	43.657	35.280	37.285	32.856
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the log turnout in the 2020 presidential election. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online Homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. 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)
Gmail Homophily Shock	-0.022*	0.017*	-0.007	-0.002	-0.004	-0.003
	(0.011)	(0.010)	(0.007)	(0.008)	(0.008)	(0.007)
Mean of Dep. Var.	0.561	0.561	0.561	0.561	0.561	0.561
Adj R2	0.543	0.644	0.786	0.794	0.794	0.801
Observations	3042	3042	3042	3042	3042	3042
Instrumental Variable Estimates						
Online Homophily	-0.036***	0.062	-0.018	-0.005	-0.011	-0.009
	(0.012)	(0.051)	(0.016)	(0.019)	(0.022)	(0.021)
F-stat	17.732	10.914	43.657	35.280	37.285	32.856
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*

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

Table A.18: Homophily Shock Reduces the Fraction of Long Ties

Dep. Variable:	Fraction of Long Ties					
	(1)	(2)	(3)	(4)	(5)	(6)
Gmail Homophily 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)
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
Instrumental Variable Estimates						
Online Homophily	-0.689**	-0.538*	-0.541***	-0.422***	-0.457***	-0.437**
	(0.285)	(0.297)	(0.149)	(0.151)	(0.170)	(0.168)
F-stat	17.615	10.403	40.017	35.087	36.581	30.988
Log Pop, 2010	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
DMA FE	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*No*	*No*	*Yes*	*Yes*	*Yes*
Political Controls	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Demographic Trends	*No*	*No*	*No*	*No*	*No*	*Yes*

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

Table A.19: Impact of the Gmail Homophily Shock by Share Urban

	First-Stage	Log Facebook Visits	Log Bars and Rest. Visits	Econ Connect	Political Homogeneity	Extreme Margins
	(1)	(2)	(3)	(4)	(5)	(6)
Gmail Homophily Shock	0.193***	0.101	-0.084**	-0.127**	-0.015***	-0.025*
	(0.065)	(0.061)	(0.039)	(0.062)	(0.005)	(0.013)
– × Share Urban	0.157***	0.102***	0.019	-0.041***	-0.009***	-0.018***
	(0.015)	(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	0.849	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: first-stage in column 1, Log Facebook visits in column 2, Log Bar visits in column 3, Economic Connectedness in column 4, political homogeneity in column 5 and the share of precinct within a county with electoral margins larger than 70 points in 2016 in column 6. Gmail Homophily Shock is our standardized excluded instrument measuring the differential Gmail complementarity in the six quarters following the API change. Across all specifications we also control for the pre-period Gmail complementarity using the last six quarters before the API changed. Online Homophily is the standardized inverse of the social diversity index. The social diversity index is the average socio-economic distance to all counties in a county network, weighted by 2016 Facebook links. Baseline Controls include log population, share White, share with at least some college and share unemployed in 2010; turnout and Republican shares as of 2008. Demographic Controls include share Black, share Hispanic, log median income, share in labor force, share rural and median age in 2010. Political Controls add political homogeneity in 2008 to the list of controls. Demographic Trends include differences for all Demographic Controls between 2000 and 2010. 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: Homophily Shock and Social Media Visits, OLS

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

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables are 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.21: Homophily Shock and Time Spent on Social Media, OLS

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

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables are 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.22: 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	0.026 (0.019)	0.023 (0.019)	0.035 (0.031)	-0.006 (0.032)	-0.005 (0.035)	-0.008 (0.036)
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 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.23: Homophily Shock and Economic Connectedness, OLS

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

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

Table A.24: Online Homophily and Political Homogeneity, 2020, OLS

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

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

Table A.25: Online Homophily and Dispersion of Electoral Results, OLS

Dep. Variable:	Sd Trump Share		Iqr Trump Share		Share Extreme Margins		Share V. Extreme Margins	
Online Homophily	0.010***	0.002	0.016**	0.001	-0.080***	-0.040***	-0.082***	-0.038***
	(0.003)	(0.002)	(0.007)	(0.005)	(0.023)	(0.012)	(0.017)	(0.010)
DMA FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Baseline Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Demographic Controls	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*
Political Controls	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*
Demographic Trends	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*
Adj R2	0.647	0.707	0.560	0.637	0.645	0.769	0.564	0.759
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.