Social Networks and Voter Information

Victoria Mooers*

April 7, 2026

Updated frequently. Click here for latest version.

Abstract

Informed voters are essential for government accountability, and social networks are an important avenue through which voters acquire political information. However, U.S. House of Representatives districts do not necessarily align with social networks. This misalignment potentially impacts the ease with which voters learn about their representatives, by altering the chance of encountering friends who provide relevant political information. I study whether the alignment between district boundaries and social networks affects voter knowledge, turnout, and campaign contributions in congressional elections. Using Facebook's Social Connectedness Index and an event study design, I find that an increase in the share of friends living in the same district increases voters' knowledge about their representative. For example, a 10 percentage point increase in this share raises the probability that a voter knows their representative's party by 3.3 percentage points; this represents a 5% increase over the mean. Additionally, a higher share of friends in the same district increases voter turnout in House elections and shifts campaign contributions towards own-district House candidates. These findings suggest that aligning political boundaries with social networks can enhance democratic engagement.

*Email: mooers@uchicago.edu. I would like to thank Andrea Prat, Alessandra Casella, Suresh Naidu, Ebonya Washington, and Gemma Dipoppa for their guidance and support. I am grateful to Jack Blumenau, Sarah Eichmeyer, Navin Kartik, Mate Malashkhia, John Marshall, Tereza Petrovicova, Evan Sadler, and workshop and conference participants at Columbia University, the 2024 WEAI Graduate Student Workshop, the 2024 Washington University in St. Louis Economics Graduate Student Conference, University of Arkansas, University of Oklahoma, University of Alabama-Birmingham, the 2025 CEPR Political Economy Symposium, the 2025 LSE Behavioural Political Economy Workshop, the APSA Annual Meeting, the University of California, Berkeley, Harvard University, the University of Chicago, and the MEA Annual Meeting for their thoughtful comments. Special thanks to Nicolas Longuet-Marx for help with data on campaign contributions. I acknowledge the Program for Economic Research at Columbia University and the Provost Diversity Fellows Small Grants Program for support, as well as computing resources from Columbia University's Shared Research Computing Facility project, which is supported by NIH Research Facility Improvement Grant 1G20RR030893-01, and associated funds from the New York State Empire State Development, Division of Science Technology and Innovation (NYSTAR) Contract C090171, both awarded April 15, 2010. All remaining errors are my own.

1 Introduction

Every ten years, the United States redraws its congressional districts, reshaping the political boundaries that define constituencies for the U.S. House of Representatives. While the U.S. Constitution mandates that these districts have roughly equal populations, it places few other restrictions on how these lines are drawn. Literatures spanning the social sciences, mathematics, and computer science have debated the most appropriate way to measure the fairness of these districts (e.g., Stephanopoulos and McGhee 2015, McCartan and Imai 2023). But the proposed measurements typically assume that changes in district boundaries do not change voter turnout: In a given location on the map, the expected number of voters from each political party is fixed—regardless of which other voters are grouped in their district.

However, voters do not act in isolation. Rather, many voters primarily learn political information through their friends, families, neighbors, and coworkers—i.e., their social networks (Lazarsfeld et al. 1944). The structure of social networks has been shown to determine how information spreads, influencing a variety of economic outcomes (Conley and Udry 2010, Banerjee et al. 2013, Beaman et al. 2021). However, causal estimates of how social networks impact political knowledge are limited (Fowler et al. 2011). In particular, the role of political boundaries—which group some friends together while separating others—in social learning has not been causally explored.

How does the alignment between social networks and political boundaries impact voters' political knowledge and behavior? Voters who live in the same district as a larger share of their friends may be more likely to hear about their representative through their social network. In the aggregate, such incidental exposure could lead to sizable differences in voter knowledge between areas where social networks are more or less concentrated within district boundaries. Additionally, more informed constituents may be more likely to vote (e.g., Snyder and Strömberg 2010), suggesting that changing district borders could also change who turns out to vote. Without understanding this relationship between district borders and social networks, existing measures of gerrymandering and electoral fairness remain incomplete, overlooking the role of social networks in shaping political knowledge and participation.

In this paper, I estimate the causal impact of the county-level share of same-district friends (which I refer to as "district homophily") on voters' knowledge, turnout, and campaign contributions in U.S. House of Representatives elections. For the average person in a county, district homophily is the share of their friends that live in their district. Using an event study design, I leverage changes in district homophily resulting from redistricting, which I demonstrate are plausibly exogenous.

I find that voters are more informed when their social networks better align with their congressional districts. For example, I find that a 10 percentage point (slightly less than one standard deviation) increase in district homophily raises the probability that a voter knows their representative's party by 3.3 percentage points, from a mean of 62% (a 5% increase).[1] The same increase in district homophily raises the probability that a voter recognizes their representative's name by 0.7 percentage points, from a mean of 93%. By Snyder and Strömberg 2010's estimates, this is equivalent to the effect of publishing 15 additional newspaper articles about the representative in the local newspaper (from a mean of 101 articles per congressional term). I find no effect on knowledge of governors and senators—placebo outcomes, since these statewide offices are unaffected by district borders—suggesting that the increase in knowledge of House representatives does not reduce attention to other elected officials.

I also find that district homophily increases voter turnout in House elections and redirects campaign contributions toward own-district House candidates. A 10 percentage point increase in district homophily raises turnout in House elections (relative to turnout in the top-of-ticket election) by 0.4 percentage points among general election voters, or 0.2 percentage points among the voting age population (roughly equivalent to publishing 49 additional newspaper articles, by Snyder and Strömberg 2010's estimates). For the full voting age population, this implies that for every four additional people who hear about their representative through their social network, one more person votes in the House election. Additionally, a 10 percentage point increase in district homophily raises the share of dollars contributed to in-district House candidates (as a share of the county's total contributions to all House candidates) by 7.4 percentage points, from a mean of 51%.

This paper makes two primary contributions:

First, I contribute to the literature on how voters learn about politics by providing causal estimates of the extent to which social networks impact voter knowledge at the scale of nearly the entire U.S. A large literature on government accountability emphasizes the importance of informed voters in government oversight, with informed voters receiving more public spending. This literature has emphasized traditional media—such as TV, newspapers, and radio—as key sources of political information (e.g., Strömberg 2004, Eisensee and Strömberg 2007, Ferraz and Finan 2008). A related literature examines how the internet and social media heighten responsiveness to government effectiveness by increasing voters' access to information and easing coordination (e.g., Manacorda and Tesei 2020, Guriev, Melnikov, and Zhuravskaya 2021; but contrast to Falck et al. 2014). Relatedly, Guriev, Melnikov, Silva, and Zhuravskaya 2026 show that mobile broadband

[1] In 2012, the year of redistricting that I leverage, Americans had on average 245 Facebook friends (Hampton et al. 2012, February). Around this same period, Americans were estimated to have on average 600 acquaintances (McCormick et al. 2010), though only 10-25 people they trust (DiPrete et al. 2011). A 10 percentage point increase in friends is then, on average, an increase of about 25 Facebook friends or 60 acquaintances.

expansion reduced the incumbency advantage in U.S. House elections, with evidence that spillovers between voters in the same district may account for part of the effect, a pattern consistent with the social network mechanism I study.

Nonetheless, as media choices have expanded, people can easily avoid political content if they are not interested (Prior 2007, Prior 2019), and few people closely follow the news (Delli Carpini and Keeter 1996). Consequently, incidental exposure to political information through social ties may be of increasing importance. As such, I build on this work by highlighting the role of social networks as a key source of information for voters. A rich literature in political science carefully studies the role of social networks in the spread of political information, but calls for more causal estimates (Baybeck and Huckfeldt 2002, Fowler et al. 2011, Sokhey and Djupe 2011, Sokhey and McClurg 2012, Campbell 2013). Several studies have taken up this challenge through lab experiments (e.g., Klar and Shmargad 2017, Druckman et al. 2018). Outside the lab, through field experiments Fafchamps et al. 2019 and Arias et al. 2019 study consolidating democracies, Mozambique and Mexico, respectively, and collect detailed data on social networks to analyze the consequences of both incentives and social network structure on the spread of political information. In a recent RCT, Egorov et al. 2026 show in Argentina that network spillovers can dominate and reverse the direct effect of a political information campaign: leaflets describing the likely negative consequences of an outsider candidate's proposals reduced his support among recipients but raised it among nearby non-recipients, producing a net gain for the candidate. A closely related literature studies the role of peer effects in political behavior (Sinclair 2012), which also explores other mechanisms such as social pressure (Gerber et al. 2008, Sinclair et al. 2012) and recruitment (Klofstad 2007). Many of these studies focus directly on how behaviors, such as voting, are transmitted: For example, Nickerson 2008 shows that in a door-to-door canvassing campaign, not only are people directly treated (i.e., who answered the door) more likely to vote, but so are other members of their households. Pons 2018 expands to a nationwide field experiment in France to analyze the effects of door-to-door canvassing on persuasion and vote shares. Additionally, Cantoni and Pons 2022 and Brown et al. 2023 estimate the causal effects of place on political behavior, finding that the state in which voters live explains much of the variation in voter turnout; their estimates capture peer effects and the effects of state institutions (such as same-day registration or voter ID laws).

I build on these studies by estimating how social network structure impacts voter knowledge, using data from across the continental U.S. By employing national data on social ties, my analysis comprehensively captures social networks, minimizing biases that may arise from excluding some ties, much like Alt et al. 2022, who use administrative data for all of Denmark to estimate the diffusion of economic information through networks. Similarly comprehensive, Bond et al. 2012 analyzes the national-level impact on U.S. voter turnout of a Facebook Election Day reminder that displayed pictures of friends who clicked an "I Voted" button, demonstrating the role of social influence in voters' turnout decisions, though not focusing directly on information transmission within networks. My estimates, by contrast, leverage the geographic variation in the U.S. social network, making my findings especially relevant for redistricting.

Second, I bridge the literature on social learning with the literature on models of political geography by providing the first causal estimates of the extent to which the match between social networks and political boundaries affects voters' political knowledge and behavior.

Existing models of strategic partisan redistricting (or "gerrymandering") generally assume that changes to district boundaries do not affect the distribution of partisans in a given area. These models take the perspective of a strategic gerrymanderer who has the goal of maximizing their party's influence in a legislature, such as by maximizing the expected share of legislature seats won, by strategically allocating voters to districts under the constraint that districts have equal populations (Owen and Grofman 1988, Friedman and R. T. Holden 2008, Gul and Pesendorfer 2010, Friedman and R. Holden 2020). The aim is to find the gerrymanderer's optimal strategy, which generally takes some form of "packing" a district with voters of a single type and "cracking" other districts by mixing voters of different types. The gerrymanderer may face some uncertainty over voter preferences, and the structure of the uncertainty influences the optimal strategy (Kolotilin and Wolitzky 2020). However, in these models the choice of how the boundaries are drawn does not impact the uncertainty: Voters' decisions are independent of the district map. A recent exception is Bouton et al. 2023, which develops a model of strategic gerrymandering that accounts for heterogenous turnout rates, and which allows for voters' turnout decisions to endogenously respond to the turnout decisions of others in their district. This model does not include information or social networks, so turnout decisions are driven by changes in the expected benefit of voting. In a similar vein, measures of gerrymandering seek to detect unfairly drawn maps through various measures of partisan bias, which also are constructed assuming that the distribution of partisan voters is fixed (e.g., McCartan, Kenny, Simko, Ebowe, et al. 2024, Stephanopoulos and McGhee 2015).

However, once social learning is taken into account, my findings suggest that voters' turnout decisions will indeed depend on how district borders are drawn, particularly how much of their social network is grouped within their own district. To date, the literature on gerrymandering has largely overlooked the role of social learning—primarily due to lack of social network data at sufficient scale. Though my findings are specific to the U.S., this approach may hold relevance in any context where political boundaries are drawn.

The remainder of this paper is organized as follows. In section 2, the empirical strategy is presented, with a discussion of the construction of district homophily, as well as the event study design. Section 3 presents the outcomes data. Section 4 presents and discusses the findings on voters' information, voter turnout, and campaign contributions, while section 5 provides robustness checks. Section 6 outlines a conceptual framework, illustrating the model of information diffusion within districts that underlies my approach. Section 8 concludes with policy recommendations and suggestions for future research.

2 Empirical Strategy and Networks Data

2.1 District Homophily: A Measure of District and Network Alignment

Voters learn political information through their social networks (Lazarsfeld et al. 1944, Druckman et al. 2018). For this to happen, though, voters' friends need to have relevant political information themselves. Alt et al. 2022, for example, show in Danish data that people become more pessimistic about the economy when a friend of a friend loses their job—presumably because they have heard the bad news from their immediate friend. In the U.S. context, with 435 representatives in the House, people may pay more attention to their own representative rather than others. As a result, it is plausible that a friend is more likely to know and share useful information about a representative if they live in the same district. Accordingly, information about representatives is likely to spread more quickly when people are more likely to interact with others from the same district.

Assuming that information about representatives spreads more readily between friends in the same district, the way district boundaries are drawn can shape how political information flows through social networks. I study how geographic mismatches between social networks and political boundaries influence this flow, focusing specifically on the alignment between county-level social networks and congressional districts.[2] To capture this alignment, I construct "district homophily": the probability that a randomly chosen person from a county and a randomly chosen one of their friends both live in the same congressional district.[3]

In order to construct district homophily, I use data representing the Facebook friendship graph, which is one of the best available proxies for real-world social networks. However, I also construct an alternative measure based on commuting flows, and with this measure I find qualitatively similar results. I construct district homophily for counties in the 48 contiguous U.S. states, and I show that district homophily varies

[2]I focus on counties to facilitate linking to county-level outcomes data for vote counts and campaign contributions. However, results on voter information and self-reported turnout are similar when instead using zip code-level social networks.

[3]Friendship shares of a similar flavor underpin Echenique and Fryer 2007.

substantially across the U.S. I demonstrate that district homophily is (not surprisingly) correlated with many determinants of social networks and district borders, but *changes* in district homophily due to redistricting largely are not. Consequently, plausibly causal identification of the impacts of district homophily can leverage these changes over time, as I discuss further in Section 2.2.

2.1.1 Definition of District Homophily

To construct district homophily, let $\mathcal{C} = [1, ..., C]$ be the set of all counties and $\pi_{c,k}$ be the share of county c's friends that live in county k. The matrix of county friendship shares can be represented as

$$\Pi = \begin{pmatrix} \pi_{1,1} & \cdots & \pi_{1,C} \\ \vdots & \ddots & \vdots \\ \pi_{C,1} & \cdots & \pi_{C,C} \end{pmatrix} \tag{1}$$

Suppose, for a moment, that every county is fully contained in one congressional district. Then, district homophily for county c is simply the sum of friendship shares across other counties in the district c is in, i.e.

$$\bar{\pi}_c = \sum_{k:d_c=d_k} \pi_{c,k} \tag{2}$$

where $\bar{\pi}_c$ represents district homophily and d_c is the district county c is in.

However, district borders do not necessarily follow county borders: counties can be fully contained within a single district, but they can also intersect multiple districts.[4] I adjust for this by taking population-weighted averages. Let $q_{(c,d)}$ represent the share of county c's population that lives in district d. The probability that a randomly chosen individual from county c lives in district d is then $q_{(c,d)}$. If we choose at random one of this person's friends, the probability that friend lives in county k is $\pi_{c,k}$; conditional on living in county k, the probability that friend lives in district d is $q_{(k,d)}$. Accordingly, the overall share of the first person's friends that also live in district d (regardless of county) is $\sum_{k \in C} \left(\pi_{c,k} \times q_{(k,d)}\right)$. Then, taking the population-weighted average across all districts county c intersects, represented by $D(c)$, gives district homophily:

$$\bar{\pi}_c = \sum_{d \in D(c)} \sum_{k \in \mathcal{C}} \left(\pi_{c,k} \times q_{(c,d)} \times q_{(k,d)}\right) \tag{3}$$

[4] This is not unique to counties: federal law does not require districts within a state to follow the borders of any other geographic unit.

2.1.2 Proxy for Social Networks: Facebook Social Connectedness Index

For data on social networks, I use the Facebook Social Connectedness Index (SCI)—one of the best existing proxies for real-world social networks.

In essence, the SCI aggregates the Facebook friendship graph to provide a measure of the strength of social connection between two locations (such as counties) (Bailey, Cao, et al. 2018). For each pair of counties, $SCI_{c,k}$ is constructed as the relative probability of a friendship link between users in county c and county k:

$$\mathrm{SCI}_{c,k} = \frac{\text{Friendship Links}_{c,k}}{\text{Facebook Users}_c \times \text{Facebook Users}_k} \tag{4}$$

That is, the SCI is the number of friendship links between the two locations, normalized by the total number of possible connections between them.[5]

I use the SCI for U.S. county-county pairs from the October 2021 snapshot.[6] The SCI is also available for U.S. zip code-zip code pairs. I focus on county-county pairs to facilitate matching to county-level outcomes data for vote counts and campaign contributions. However, I also construct district homophily using the zip code pairs, and I show that the results on voter information and self-reported turnout are robust to this alternative construction.

The SCI is an effective proxy for real-world social networks because it captures social ties that might not be revealed in geography-based proxies, like commuting flows, that rely on physical proximity. The SCI has been demonstrated to closely reflect offline networks (Bailey, Cao, et al. 2018; Bailey, Gupta, et al. 2021; Kuchler et al. 2022). Two features of the Facebook friendship graph aid this. First, it is very persistent over time because Facebook friendships accumulate throughout a lifetime.[7] Second, Facebook usage rates across counties are uncorrelated with demographics like income (Chetty et al. 2022).[8]

[5]The SCI is scaled from 1 to 1,000,000,000, areas with particularly small populations are removed, and noise is added to preserve privacy.

[6]There are 3,136 counties in the data, and each county appears in a pair with every other county (including itself). As such, there are 9,834,496 county-county pairs. The SCI only includes users who have interacted with (including simply logging into) any of Meta's apps (Facebook, Instagram, WhatsApp) in the 30 days prior to the snapshot. Locations are assigned based on users' provided information (such as their stated city) and their device connection information.

[7]For example, Enke et al. 2023, May report that according to their correspondence with the SCI authors, the correlation between years of the SCI is above 0.99. Additionally, Bailey, Gupta, et al. 2021 finds that countries with higher social connectedness trade more, and that this relationship is similar in every year back to 1980; Kuchler et al. 2022 find that institutional investors are more likely to invest in firms in regions with higher social connectedness to the investor's region, and that this relationship remains at least back to 2007 (the start of their data).

[8]Similarly, at the time of the snapshot, survey-reported Facebook usage rates were also relatively even across demographic groups nationwide (Auxier and Anderson 2021, April). In particular, in 2021, Facebook usage rates (defined as whether you ever use the platform) among American adults varied slightly between urban and suburban (70%) and rural areas (67%); when sliced by race, income, and education, usage rates varied between 61% and 74%. The largest gaps emerge by age, with the lowest usage rates among 65+ year-olds (50%) and the highest usage rates among 30-49 year-olds (almost 80%); use among 18-29 year-olds reflected the national average at 70%. There are also minimal differences in Facebook usage rates by political party (Vogels et al. 2021, April). Facebook usage rates rose until 2016, and remained stable at around 70% of U.S. adults from then until at least 2021. As of 2021, Facebook was the social media platform with the least heterogeneity in usage rates by

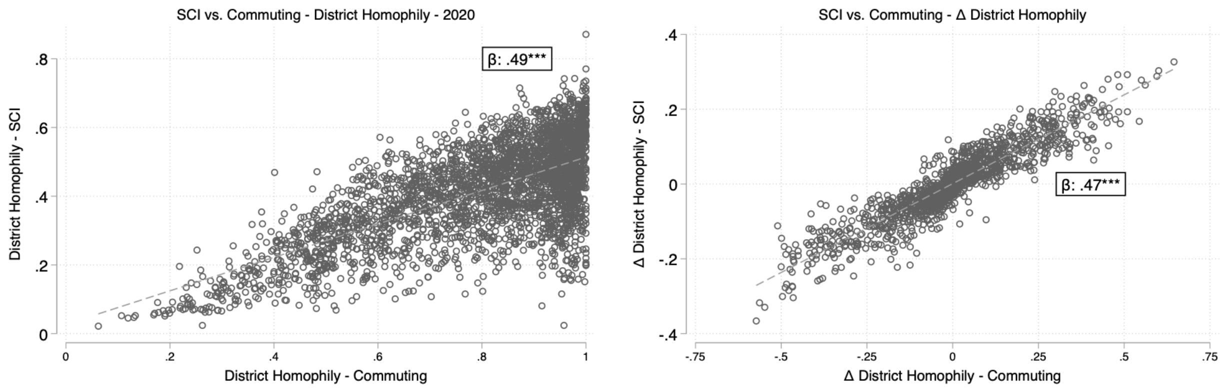


Figure 1: Comparison of District Homophily Constructed Using SCI vs. Commuting Flows. Left: District homophily constructed for 2020 district borders. Right: Changes in district homophily due to 2012 redistricting. β is the coefficient of the linear regression of SCI district homophily on commuting district homophily.

Nonetheless, in Section 5.2, I demonstrate that results are qualitatively similar if I use commuting flows as an alternative proxy for social networks. Indeed, the measures from each dataset are strongly correlated: The correlation coefficient between district homophily constructed using the SCI and district homophily constructed using commuting flows is 0.67, and the correlation coefficient between *changes* in each measure due to the 2012 redistricting (which I use for identification, discussed more below) is 0.92. Figure 1 plots district homophily constructed using the SCI against district homophily constructed using commuting flows (under 2020 borders), as well as the analogous changes in district homophily due to the 2012 redistricting constructed using each dataset. Evident in the figures, commuting flows are naturally concentrated in a smaller area than friendship networks, so commuting district homophily tends to be higher than SCI district homophily.

2.1.3 Construction of District Homophily from SCI

Whereas the SCI gives the relative probability of a friendship between two counties, district homophily is the share of a county's friends that live in the same congressional district as the people in the county. Accordingly, to construct district homophily I need to appropriately aggregate the SCI. I do this by using the SCI to construct the Π matrix of county-county friendship shares, and then for each county summing friendship shares across same-district counties (adjusting for counties that intersect multiple districts).

age. While 18-29 year-olds were the heaviest users of all other platforms, their Facebook use was only exceeded by their use of YouTube (95%) and Instagram (71%) (Auxier and Anderson 2021, April).

The share of county c's friends that live in county k, $\pi_{c,k}$, is constructed as

$$\pi_{c,k} = \frac{\text{Friendship Links}_{c,k}}{\sum_{j \in C} \text{Friendship Links}_{c,j}} \tag{5}$$

We can re-write the equation for $\text{SCI}_{c,k}$ as

$$\text{Links}_{c,k} = \text{SCI}_{c,k} \times \text{Facebook Users}_c \times \text{Facebook Users}_k \tag{6}$$

However, the number of Facebook users in each county is not made available, so this is not possible to directly construct. Bailey, Gupta, et al. 2020 argue that we can substitute the population of an area for the number of Facebook users. This requires the assumption that Facebook usage rates are the same across counties. This assumption is likely benign: as mentioned above, Chetty et al. 2022 demonstrate that while there is some variation in Facebook usage rates across counties, this variation cannot be predicted by demographics.

Replacing the number of Facebook users in a given county with the county's population and re-arranging, we get

$$\pi_{c,k} = \frac{\text{SCI}_{c,k} \times \text{Pop}_k}{\sum_{j \in C} \left(\text{SCI}_{c,j} \times \text{Pop}_j\right)} \tag{7}$$

which is feasible to calculate. I retrieved each county's population from the 2020 Decennial Census and then calculate $\bar{\pi}_c$ for each county.

Finally, the SCI is only available for one snapshot of the social network, in 2021. In order to derive district homophily in each year over the period, I hold the social network fixed, and I re-calculate district homophily with each congressional border change—consequently, all changes in district homophily are due solely to changes in the location of the district border.[9] As discussed above, social networks are slow-changing, so this assumption is not unreasonable. Further, while in 2021 Facebook was evenly used across demographic groups, around 2012 Facebook was expanding and still predominantly used by younger users—using the 2012 Facebook friendship network would be less reflective of offline networks. Lastly, holding the social network fixed at its 2021 structure and projecting it back in time will primarily introduce measurement error. I construct district homophily for each year from 2002 to 2022, i.e. the 107th–117th Congresses.

[9] Specifically, I assume that the share of a county's friends in each other county remains the same. To do this, I also assume that the populations are the same as Decennial Census 2020 populations (otherwise changes in population of one county would affect friendship shares for many other counties).

2.1.4 Examples of District Homophily

To illustrate the relationship between the SCI and district homophily, as well a how district homophily can change due to redistricting, consider Coosa County, Alabama, which experienced the biggest increase in district homophily of any county following the most recent redistricting, which occurred in 2022 based on population counts in the 2020 Decennial Census.

Figure 2 shows the value of the SCI between Coosa County and each other county in Alabama. Coosa County is highlighted with a blue border. The counties that Coosa has the strongest social connections with are in dark red, while the counties that Coosa is most weakly connected to are in light yellow; there is an equal number of counties in each color bin. The maps only reflect Coosa County's connections to the other counties—they do not reflect how any other two counties are connected to each other. The map on the left displays the congressional district borders in Alabama immediately prior to redistricting (the borders used in the 2020 election), while the map on the right displays the borders immediately following redistricting (the borders used in the 2022 election).

Coosa County is most strongly connected to other counties to its east, while the strength of its connections drops off more quickly going west. Under the 117th Congress borders, Coosa County lies in the southeastern corner of its congressional district, with the district border following the north, east, and south borders of the county; Coosa County only shares a border with another county in its district on its western side, and only one of the counties it is most strongly connected to (darkest red) lies in the same district. Under the 118th Congress borders, Coosa County is moved into the district that had been east of it. Coosa County is still in the corner of the district, but its district border is reflected to the opposite corner, and Coosa County is now grouped into a district with all but two of the counties it is most strongly connected to.

Thus, in the left map, Coosa County is cut off from much of its social network, while in the right map Coosa County is grouped in with much of its social network. This is reflected in Coosa County's district homophily before and after redistricting. In Figure 3, the left map represents the district homophily of each county of Alabama before the 2022 redistricting, while the right map represents district homophily after redistricting. Again, there is an equal number of counties in each bin, so district homophily levels should be interpreted as district homophily relative to other counties in Alabama. As we might predict from the SCI maps, Coosa County has among the lowest levels of district homophily in Alabama under the 117th Congress borders. However, under the 118th Congress borders, when it is grouped in with the counties with larger shares of its friends, Coosa County has one of the highest levels of district homophily in Alabama. In particular, Coosa County experiences a 39.3pp change in district homophily, going from 16.6% under the

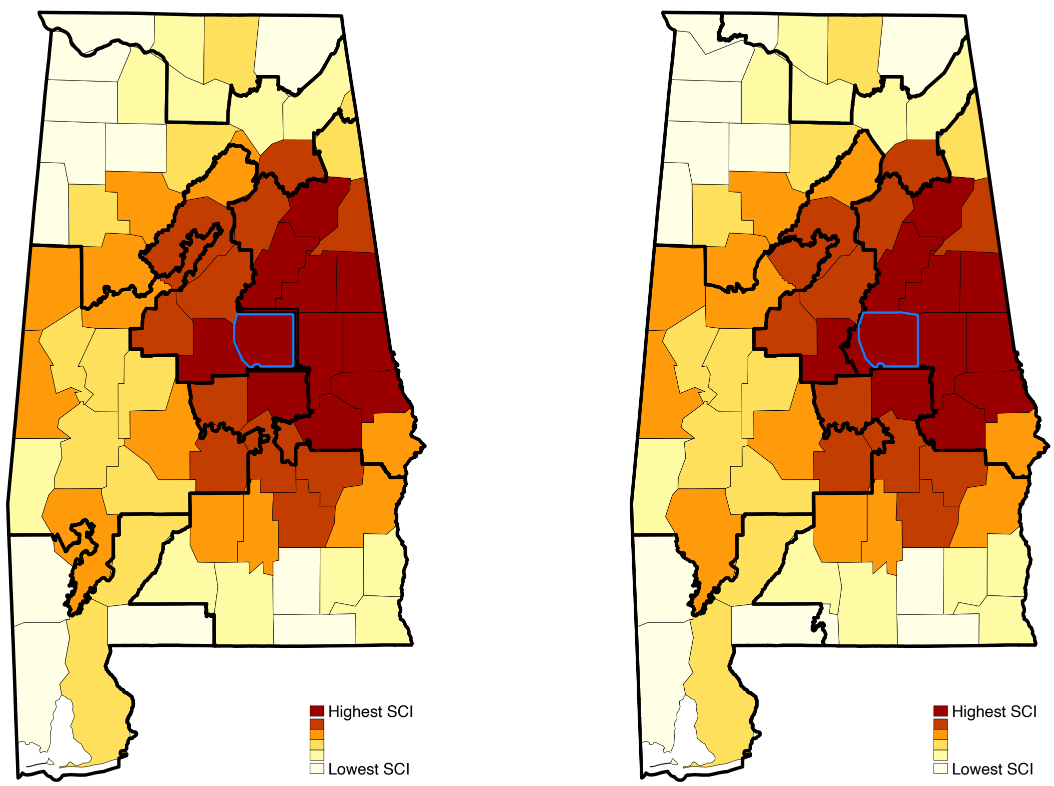


Figure 2: SCI of Coosa County, Alabama. Left: 117th Congress boundaries; Right: 118th Congress boundaries.

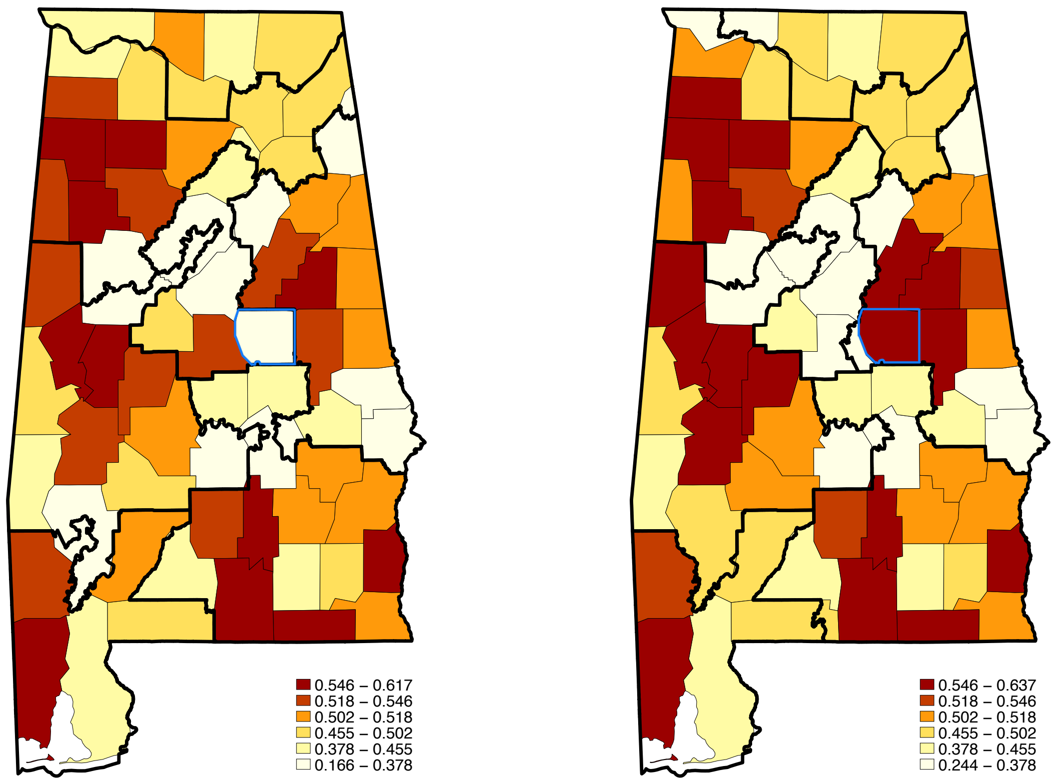


Figure 3: District Homophily of Alabama Counties. Left: 117th Congress boundaries; Right: 118th Congress boundaries.

old borders to 55.9% under the new borders.

2.1.5 Summary Statistics and Predictors of District Homophily

District homophily varies substantially across the continental U.S. District homophily is determined by both social networks and district borders; consequently, demographics and geographical features that are correlated with either of these determinants are highly correlated with district homophily.

Among the continental 48 states over the full period, mean district homophily is 41% with a standard deviation of 14pp; minimum district homophily is 2% and maximum is 87%, while the 1st percentile is 8% and the 99th percentile is 67%. The middle 50% of counties have district homophily between 32% and 51%, and the middle 80% of counties have district homophily between 22% and 58%.[10]

Appendix Figures C1–C2 summarize how various geographic and demographic features correlate with

[10]These statistics are roughly stable over the full period. Throughout this paper, I focus on the 48 contiguous states—i.e., in results I exclude Alaska, Hawaii, Washington, D.C., and territories. However, all counties as well as foreign friendships are included for calculating the scaled total number of friends (denominator) for each county.

district homophily, separately in 2012 and 2020.[11] Generally, patterns are very similar in 2012 and 2020, supporting the argument that at the county scale social networks remain very similar over this period. As might be expected given that social networks tend to follow state boundaries (Bailey, Cao, et al. 2018), counties in single district states[12] have higher district homophily on average (53%). In addition, due to the restriction that each congressional district within a state represent roughly the same population (across states averaging about 760,000 in 2020, see Eckman 2021, November; Whitaker 2017, March), counties with large populations (including most urban areas) are more likely to be split by a district boundary in order to accommodate this constraint; a one percent increase in county population is associated with a 0.05pp decrease in district homophily.

The other determinant of district homophily is the geography of social networks. The biggest predictor of social ties is distance (Bailey, Cao, et al. 2018), so counties that are further from a congressional district border will generally have higher district homophily. Naturally, this is more likely to occur in geographically large districts, which are necessarily in areas with lower population density (again because each district is meant to have roughly the same population). This leads to higher district homophily in rural areas. Simultaneously, urban areas have much more geographically dispersed social networks, because they have strong ties to other urban centers around the country (Bailey, Cao, et al. 2018); this further drives down district homophily in urban areas. Similarly, counties with more people who have moved in the past year have lower district homophily, with district homophily declining as movers come from further away.

2.1.6 Changes in District Homophily over Time

How is district homophily changing over time? I examine changes in district homophily due to redistricting following the 2010 and 2020 Decennial Censuses (that is, the changes due to boundaries first used in the 2012 and 2022 elections, respectively). Recall that I re-calculate district homophily for each year by holding the social network fixed. I then calculate the change in district homophily for a county following redistricting. In both years, the average change in district homophily is nearly zero—0.1pp in 2012, 0.3pp in 2022—with a standard deviation of 6.5pp and 5.9pp respectively. In 2012, the biggest drop in district homophily was by 36.6pp, while the biggest increase was by 32.7pp. In 2022, the biggest drop was by 31.0pp and the biggest increase was by 39.3pp. Following the 2010 Census, 500 counties (16%) experienced nearly zero change in

[11] Variables on education, income, and geographic mobility are from the 5-Year ACS; variables on partisanship are based on the presidential election in the given year and from MIT's Election Data Science Lab; and the remaining variables are form the Decennial Census. The figures show a regression of district homophily on each variable; that is, each coefficient should be interpreted as the change in district homophily associated with a zero to one change in the predictor variable.

[12] In 2022, these were Delaware, North Dakota, South Dakota, Vermont, and Wyoming (as well as Alaska, which is not included in my analysis). These five states contain 159 counties, or 5% of all counties in the data.

district homophily (specifically, an absolute change of less than 0.1pp), and following the 2020 Census, 467 counties (15%) experienced nearly no change in district homophily. Thus, most counties experience some change in district homophily, but very large changes are unusual.

In Appendix Figure C3, I show the predictors of changes in district homophily for 2012. Most county characteristics are not significantly correlated with changes in district homophily; the few significant correlations that do exist disappear once media market and congressional district fixed effects are included (which I will do in my preferred specifications, discussed below). When no fixed effects are included, the strongest pattern is that older populations are more likely to experience an increase in district homophily. When media market and congressional district fixed effects are included, the share of the population that is white and non-Hispanic is positively correlated with changes in district homophily and the share with income below the poverty line is negatively correlated. Accordingly, I control for these characteristics in my regressions.

2.2 Redistricting

District homophily measures variation in the match between social networks and congressional district boundaries, but it is not itself exogenous. District homophily is correlated with factors that determine district boundaries, factors that determine social networks, and sociodemographic characteristics. Accordingly, in order to have plausibly exogenous variation in district homophily, I need to control for these factors, especially when they are possibly correlated with outcomes of interest.

In order to capture plausibly exogenous variation in district homophily, I measure the impact on outcomes of a change in district homophily due to congressional redistricting. I use an event study design, focusing on the redistricting that followed the 2010 Census. Appendix Figure C4 provides a map of these changes. Focusing on a single redistricting event allows me to avoid concerns related to staggered treatment events, and also allows for a visual test of pre-trends in changes in district homophily. The Census was conducted in April 2010, and states needed to draw new congressional district borders in time for the November 2012 elections. Accordingly, the congressional representatives first elected under the new borders assumed office in January 2013. As such, the last year before the treatment (i.e., a change in district homophily) will consequently depend on the outcome. For outcomes that relate to the *current* representative, 2012 is the last year before treatment. For outcomes that relate to the upcoming election (therefore more related to the *next* representative), 2011 is the last year before treatment (or more commonly 2010, for outcomes only available in even years).

Assuming 2012 as the last year before treatment, the event studies accordingly take the following form:

$$y_{ict} = \lambda_t + \sum_{\tau=2006}^{\tau=2010} \beta_\tau \Delta\bar{\pi}_c \mathbb{I}(\tau = t) + \sum_{\tau=2014}^{\tau=2022} \beta_\tau \Delta\bar{\pi}_c \mathbb{I}(\tau = t) + X_{ct}\delta + Z_{ict}\gamma + \varepsilon_{ict} \tag{8}$$

where where y_{ict} is the outcome for a given individual i in county c in year t, $\Delta\bar{\pi}_c$ is the change in district homophily experienced by county c between 2012 and 2013, λ_t are year fixed effects, X_{ct} is a vector of county-by-year controls (to further adjust for things like changing demographics over time), and Z_{ict} is a vector of individual controls. Errors ε_{ict} are clustered at the county level.

I can additionally include district-by-year fixed effects. This can be thought of as controlling for House election-specific factors that impact outcomes for all counties in the district. These can include characteristics of each of the candidates, scandals, national attention, levels of fundraising and campaign spending, etc.

Another concern may be that social networks may be highly correlated with media markets, and consequently congruence actually just reflects the impacts of TV and radio news or political advertisements bought at the media market level. To address this concern, I use the boundaries of the Nielsen Designated Market Areas and include DMA-by-year fixed effects.

Lastly, I control for partisan biases in network connections by constructing each county's exposure to Democrats. For each county, I multiply the share of the county's friends in each other county by the Democratic vote share in the county in the most recent presidential election; I then sum this across all counties the given county is connected to. In essence, this forms a rough approximation of the share of a county's friends that voted Democratic.

3 Outcomes Data

I study the impact of district homophily on voters' knowledge and political behavior. I begin with survey data to study voters' knowledge of their representatives and their self-reported vote choices and candidate preferences. I then incorporate vote count data to reveal actual voting behavior, as well as data on campaign contributions to understand impacts on donation behavior.

3.1 Voters' Information

I test whether voters in counties with higher district homophily are more informed by using responses in the Cooperative Election Study (CES) (formerly the Cooperative Congressional Election Study, or CCES; see for example Schaffner, Ansolabehere, and Shih 2023) to measure voters' familiarity with their representatives.

The CES is a nationally representative survey that has run annually from 2006 to 2022 and ask about topics including demographics, political attitudes, political knowledge, and voting intentions and choices. In federal election years (i.e., all even years), a pre-election survey is conducted from late September to late October, and a post-election survey is conducted in November. In non-federal election years (i.e., all odd years) a single survey is conducted in the fall. I use the pre-election surveys (or single surveys in odd years) for 2006–2022. The CES sample consists of 50,000+ adults in every federal election year since 2010 (>30,000 in 2006 and 2008) and 10,000+ adults in every odd year. I use the CES's cumulative weights, which re-weight observations to make sample sizes comparable across years (see Kuriwaki 2018). The CES includes each respondent's county and congressional district, enabling me to link respondents to county-level district homophily measures and to observe responses to questions about each respondent's own representative.

I construct three binary variables to assess how familiar respondents are with their current representative. Detailed descriptions of these variables are in Appendix Table B1. Respondents are asked to "Please indicate whether you've heard of this person and if so which party he or she is affiliated with...". They are asked this about their current House representative, both of their senators, and their governor. Respondents can answer "Never Heard of Person", "Republican", "Democrat", "Other Party/Independent", or "Not Sure". The first dummy variable, "Heard of Incumbent", is coded as 0 if the respondent answered "Never Heard of Person" and 1 otherwise.[13] This variable captures whether the respondent claims to have any familiarity with their representative at all: do they even recognize the name? The second dummy variable, "Selected Party", is coded as 0 if the respondent answered "Never Heard of Person" *or* "Not Sure", and 1 otherwise. This variable indicates whether, beyond recognizing the representative, the respondent claims to have some knowledge about them: they claim to know the party the representative belongs to (though they may just be guessing). Lastly, the third dummy variable, "Selected Correct Party", is coded as 1 if the respondent selected the correct party for the incumbent and 0 otherwise. While lucky guesses cannot be ruled out, this variable generally indicates that the respondent at least knows enough about their representative to know what party their representative belongs to. Appendix Table B3 shows that, as expected, fewer people select their representative's party (68.6%) than claim to have heard of them (93.2%), and fewer still select the correct party (61.6%—though, among those who select a party, the overwhelming majority select the correct party).

[13]In 2006, 2007, and 2009, respondents do not have the option to say "Never Heard of Person" and instead can only say "Not Sure." Consequently, I drop these years in regressions using the "Heard of Incumbent" variable. Note that 2008, 2010, and 2011 still provide observations of this variable prior to the redistricting that follows the 2010 Decennial Census, because the new districts first apply in 2012.

As when constructing the district homophily measure, I only include respondents in the 48 contiguous states.[14] Additionally, not all counties are represented in every year; in even years, there is at least one respondent from 80–90% of counties, while in odd years about two-thirds of counties have at least one respondent. Because the weighted sample is representative of people living in the U.S. (rather than of U.S. counties) and more people live in urban areas (which tend to have lower congruence), the average respondent's county congruence is slightly lower at 37% (compared to 41% for the average county).[15]

3.2 Voter Turnout

I test impacts of district homophily on voter turnout and vote shares using both survey responses in the CES as well as county-level vote count data from Dave Leip's Election Atlas (David Leip 2024).

I begin with CES survey responses to study the impacts on voting within the same sample as the information outcomes. The pre-election surveys ask respondents questions about their voting intentions (e.g., who they prefer among candidates running), while the post-election survey asks respondents about who they ended up voting for. I use both the pre-election survey and the post-election survey: While the post-election survey asks about actual vote choices, outcomes from the pre-election survey utilize the same sample as the information outcomes (because there is some attrition between surveys).

Next, I use county-level vote counts from Dave Leip's Election Atlas to measure the impacts of district homophily on actual voting outcomes. I use the period spanning 2002–2020; 2002 is the first election under the district boundaries that are in place through the 2010 election, and 2020 is the last election under the district boundaries that are first used in the 2012 election.

To include district-by-year fixed effects, I take two approaches. First, when using this county-level vote count data, I include only counties that are in a single congressional district, or I only link the county to the congressional district that a majority of its population is in. Second, I use other data sources to construct the same voting outcomes at the county-by-district-level, using precinct-level vote count data (see Appendix Section B.1.1 for details). I do not use the precinct counts for my main results as it does not fully span my time period. However, I find qualitatively similar results using this data.

[14] In these 48 states and across all 17 years of the CES, there are 612,085 respondents (552,307 excluding 2006, 2007, and 2009). I exclude missing responses to the candidate party recognition question (<2% of respondents in each year; for most of these cases, the House candidate name is missing in the survey). When including individual demographic controls, I similarly exclude respondents who did not answer the relevant demographic questions. I also exclude a small number of respondents in 2006 and 2007 that are assigned to counties that are not in their state of residence. Lastly, in the 2020 survey, 925 respondents in North Carolina were assigned to incorrect congressional districts, and consequently were shown the candidate names for the wrong district. I exclude these respondents, since they were not asked about their familiarity with their own representative. See Schaffner, Ansolabehere, and Luks 2021.

[15] The distribution is otherwise similar to the county-level distribution, with a standard deviation of 11pp, a minimum of 9% and a maximum of 74%.

I focus on turnout in the House election relative to turnout in the top-of-ticket election, i.e., the election that is likely to receive the most attention and have the most force in driving voters to show up at the polling booth. I define the "top-of-ticket" election as the Presidential election when it occurs (every four years), and in midterm years as the Senate election (if occurring, which it does for about two-thirds of counties in midterm years), else the Governor election (if occurring, which it does for about a quarter of counties in midterm years).[16] Accounting for the top-of-ticket election helps to further control for factors unrelated to the House election that may drive differences in overall turnout, including differences in the cost of voting. In particular, I construct the difference between the number of votes cast in the top-of-ticket election and the number of votes cast in the House election, as a share of the top-of-ticket votes cast. This measure captures the share of voters who, despite having paid the cost of voting to vote in the top-of-ticket race, choose to abstain from the House election. (This is also referred to as "roll-off," the seemingly paradoxical phenomenon that Feddersen and Pesendorfer 1996 seek to explain with their model of the "Swing Voter's Curse," and which Miller 2022; Snyder and Strömberg 2010 also study empirically.) Details and descriptive statistics are provided in Appendix Section B.2.

3.3 Campaign Contributions

I test impacts of district homophily on donation behavior using data on campaign contributions to House candidates from Kuziemko et al. 2023, October, which is constructed from the Federal Election Commission campaign contribution data in Bonica 2014. For a given contribution, Kuziemko et al. 2023, October use geocoding to identify whether the contributor lives in the same congressional district as the House candidate they are donating to. Then, for each Census tract, they construct the aggregate amount of donations to in-district candidates and to out-of-district candidates (both the dollar amount, and the number of contributors). With this data in hand, for each county I construct the share of contributions to in-district candidates, for 2002-2016.

4 Results

Section 4 presents my main results that district homophily increases voters' knowledge about their representatives, and accordingly decreases abstention in House elections. I also find that district homophily shifts donations to same-district House candidates and away from out-of-district candidates.

[16] In each midterm year, there are a few states where neither a Senate election nor a Governor election occurs, but this only applies to about 10% of counties on average.

4.1 Voters

4.1.1 Voters' Knowledge about Representatives

I estimate positive and significant coefficients for the impact of district homophily on voters' knowledge about their representatives. Figure 4 presents the event studies for outcomes "Heard of Incumbent," "Selected Party," and "Selected Correct Party" constructed from the CES data. I find that an increase in district homophily has an immediate and persistent positive impact on voters' knowledge. I focus on even years of the CES survey until 2022 (the last year before the next national redistricting event).[17] The β and dashed line on the figures indicate the estimated aggregate effect, from the specification of the form:

$$y_{ict} = \alpha + \beta_1 \Delta\bar{\pi}_c + \beta_2 \mathbb{I}(t > 2012) + \beta_3 \Delta\bar{\pi}_c \times \mathbb{I}(t > 2012) + X_{ct}\delta + Z_{ict}\gamma + \varepsilon_{ict} \tag{9}$$

In the event studies shown, I include district-by-year fixed effects, DMA-by-year fixed effects, individual demographic controls from the CES, county-by-year demographic controls from the Decennial Census and 5-Year ACS, and the control for partisanship of the social network; however, adding the fixed effects and controls beyond the district-by-year fixed effects makes little difference. Further, results are similar when county fixed effects are included. District homophily is measured on a scale from 0 to 1, and outcome variables are binary. As such, reported estimates give the change in probability of the outcome (measured between 0 and 1) that would result from a 0 to 1 change in district homophily.

The event studies show that the change in voter knowledge due to changes in district homophily in redistricting most strongly takes effect in the first survey after redistricting (2014). Impacts are relatively stable over time. The stability of the estimates may be attributable to an attention story: only voters who have a high share of friends in their district are reminded by their friends about their representative often enough to actually remember their representative's name and political party when asked to fill out the survey.

Based on these estimates, if we assume linear impacts, an increase in a county's district homophily by 10pp would increase the probability that a respondent in that county has heard of their representative by 0.7pp (recall from Appendix Table B3 that the mean is 93.2%). The same change in a county's district homophily would increase the probability a respondent in that county selects a party by 3.2pp (from mean 68.6%) and selects the correct party by 3.3pp (from mean 61.7%).[18]

[17] The odd years have a sample about one-fifth the size of even years. As such, including odd years yields similar results with noisy estimates on the odd-year coefficients. Focusing on even years also gives consistency in interpretation: the event studies thus reflect voters' knowledge of their current representative shortly before the election that will replace or re-elect that representative. Further, voting outcomes (in the next section) are mostly only available in even years.

[18] Recall from Section 2.1.6 the a one standard deviation change in district homophily following redistricting is roughly 5pp; one standard deviation of district homophily itself is 11pp. The largest changes in district homophily following redistricting are

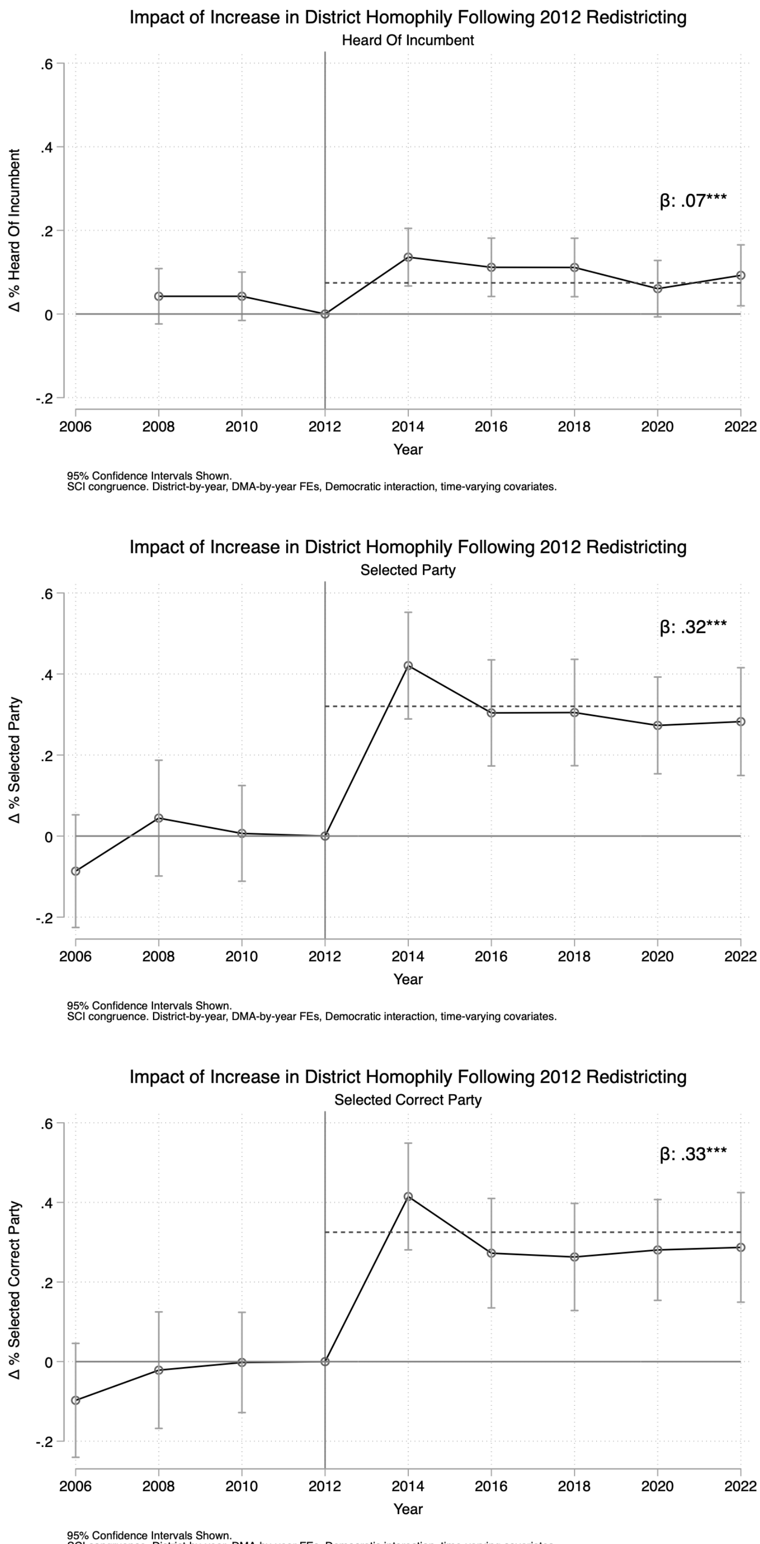


Figure 4: Effect of District Homophily on Voter Familiarity with Representative

As discussed in more detail in Section 5, I find similar results when using commuting flows as an alternative proxy of social networks. Additionally, I do not find evidence that district homophily increases voters' knowledge on placebo outcomes (i.e., the same three outcomes but for the respondent's governor and senators).

4.1.2 Voters' Choices

How does information translate into vote choices? I examine the impact of district homophily on voter turnout and on voters' candidate preferences.

Survey Responses I start with examining subjects self-reported voting preferences and choices in the CES survey. This allows me to look at voting outcomes using the same sample as the information outcomes. The CES asks voters about their voting intentions and preferred candidates (in the pre-survey, run in September or October) and later about their actual vote choices (in the post-survey, run following the November election).[19] I run event studies analogous to equation 8 to examine the impact of district homophily on voting-related outcomes; however, here I treat 2010 as the base year, as 2012 elections occur under the new district boundaries, and accordingly district homophily with the new district may begin to impact voter behavior in the 2012 election.

First, consider voters' House candidate preferences reported in the pre-survey. Subjects are asked "In the general election for U.S. House of Representatives in your area, who do you prefer?" and are shown a list of names of candidates running in the election for their district. Subjects can choose a name, or indicate no preference for any particular candidate with options like "No One" or "Not Sure." Accordingly, I construct indicators for whether the subject prefers the incumbent (i.e., the name of their preferred candidate matches the name of their current House representative), prefers an opponent (i.e., the subject chooses the name of a candidate that is not the incumbent), or prefers neither. In Figure 5, I show event studies for these three outcomes; I always restrict to cases in which an incumbent exists.[20]

Results are noisier, but an increase in district homophily is associated with an increase in preference for the incumbent. However, this increased preference for the incumbent does not come at the cost of preference

around 30pp.

[19] All subjects that complete the pre-survey are asked to participate in the post-survey, though there is some attrition. Weights do not account for this attrition.

[20] I define a candidate as an incumbent if they are an incumbent for *anyone* in the survey—i.e., a candidate is an incumbent if they are currently serving in the House. Consequently, in 2012 the definition of "incumbent" is somewhat spurious, as due to redistricting, there are many subjects for whom an incumbent exists, but that incumbent is not their current representative. Note, however, that we can exclude 2012 and the results are similar.

for the opponent (which remains unchanged) but rather comes from a reduction in subjects reporting that they prefer no candidate.

Second, I examine whether these preferences translate into actual changes in votes. Here, in order to disentangle effects on vote choice from effects on turnout (which I address below), I restrict to subjects who voted in the general election.[21] Here again, I find the same pattern: I find no impacts on votes for the opponent but an increase in reporting voting for the incumbent (see Appendix Figure C15), driven by a decrease in subjects who report *not* voting in the House election, as seen in Figure 6. Because I have restricted the sample to general election voters, this outcome is equivalent to roll-off: turning out for the general election, but choosing not to vote in the House election.

Together, conditional on already turning out to vote, district homophily may increase extensive margin participation in House elections. To test this, I turn next to actual vote count data.

Vote Count Data Figure 7 shows the impact of an increase in district homophily on turnout in the House election, relative to top-of-ticket turnout; specifically, the outcome is the share of top-of-ticket voters who abstain in the House election, or "roll-off." Consequently, a decrease in roll-off corresponds to an increase in turnout. The specification used includes district-by-year fixed effects (for the district the majority of a county's population is in—counties for which no district has a majority of the county's population are dropped), DMA-by-year fixed effects, and county-by-year demographic controls. Results are similar when restricting to counties fully within one congressional district.

The negative impact indicates that district homophily reduces roll-off: if a county has an increase in district homophily, its voters become more likely to vote in the House election *conditional* on turning out to vote in the top-of-ticket election. Recalling that mean rolloff is about 4pp, the estimate indicates that a 10pp increase in district homophily reduces rolloff by 0.4pp, or by 10%.

4.2 Campaign Contributions

Figure 8 shows the impact of an increase in district homophily on the share of dollars contributed to in-district candidates, as a share of all county donations to House candidates. In particular, a 10pp increase in district homophily is associated with a 7.4pp increase in the share of contributions to in-district candidates, from a mean of 51%.

[21] The CES links survey respondents to state voter rolls and constructs indicators of whether respondents are active registered voters and of which elections there is a record of the respondent voting in. I include both subjects who are in this manner validated as turning out in the general election, as well as subjects who self-reported turnout when asked whether they voted in the November election.

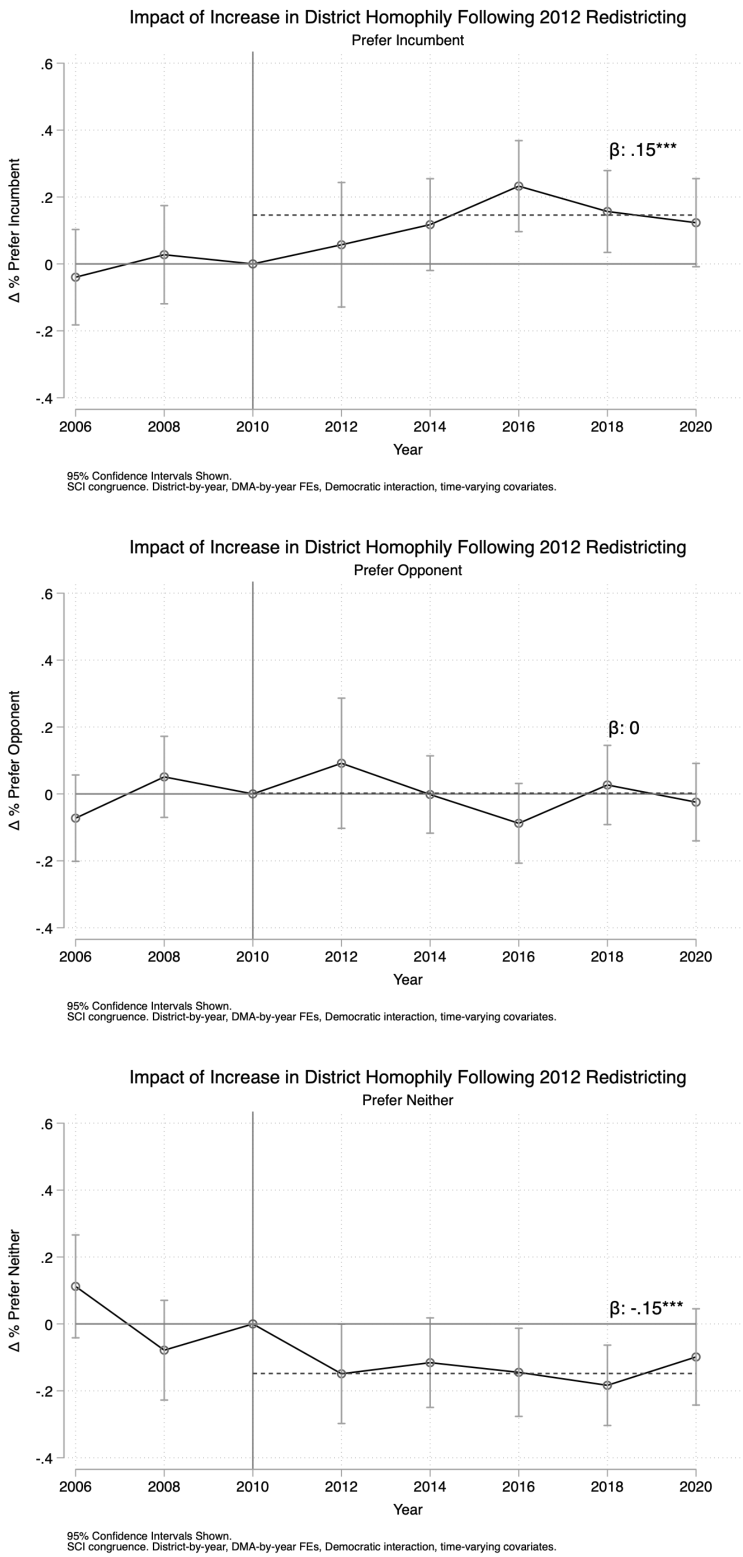


Figure 5: Effect of District Homophily on Voter Preferences

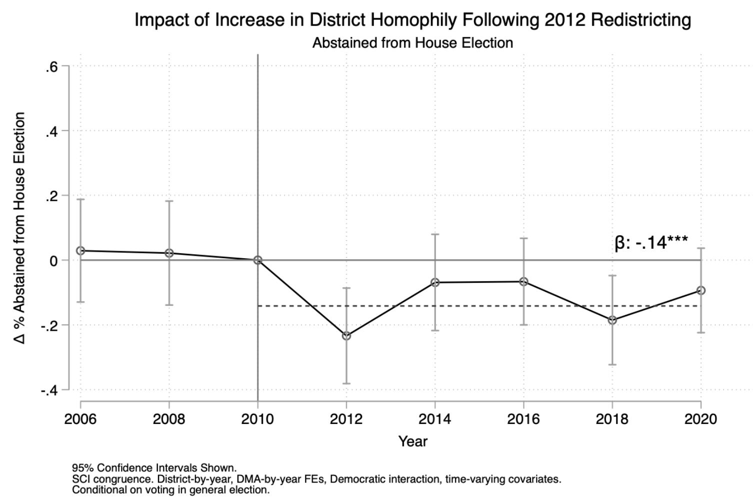


Figure 6: Effect of District Homophily on Roll-Off (Reported in CES)

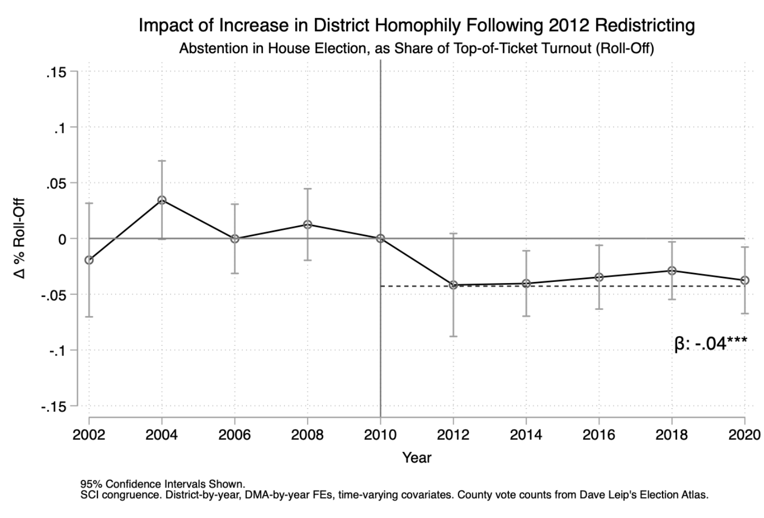


Figure 7: Effect of District Homophily on Roll-Off (Vote Counts)

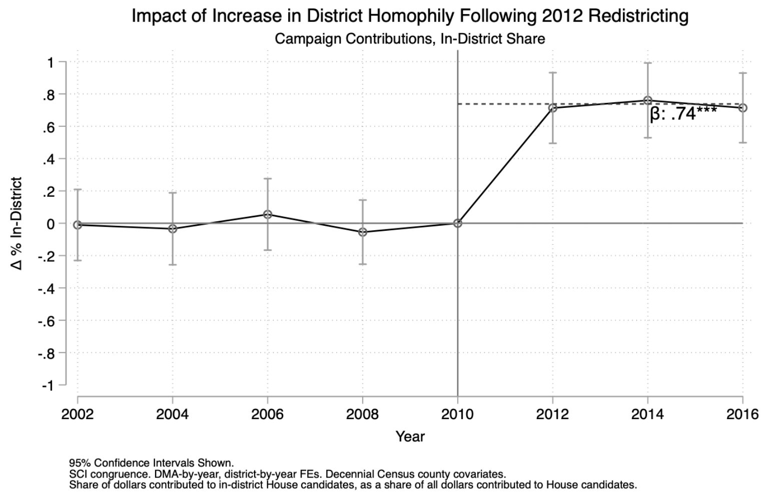


Figure 8: Effect of District Homophily on Share of Contributions to In-District Candidates

I do not find any impact on total donations to House candidates (see Appendix Section C.8), indicating that this increase is driven by shifting donations away from out-of-district candidates and towards in-district candidates, rather than changing the overall amount that contributors are allocating towards House races.

5 Robustness

I find in my main specification that district homophily has a positive effect on voters' knowledge of their representatives. I explore the robustness of this finding by testing whether district homophily impacts placebo outcomes, by constructing an alternative measure of district homophily using commuting flows, by using zip-code-level network data, and by using an alternative empirical strategy.

5.1 Placebo Outcomes

I test whether district homophily impacts voters' knowledge of their governor and senators because these offices are elected through statewide elections, and consequently congressional district borders are not relevant for them, so district homophily should not impact them.

CES respondents answer similar questions about whether they have heard of and can identify the party of their governor and each of their senators. From these responses, I construct outcome variables analogous to the ones in the main analysis, and which measure whether voters have heard of, select a party for, and select the correct party for their governor and their senators. These variables are summarized in Appendix

Table C1.

I find no significant impact of district homophily on these nine outcomes. Results are reported in Appendix Section C.2.

5.2 Commuting Flows as an Alternative Network Measure

Commuting flows can be used as an alternative measure of social networks: the number of people that commute between two counties reflects patterns of who is regularly physically proximate to each other. Replicating the analysis using commuting flows can shed light on the extent to which the SCI captures "real world" offline networks. I use the 2016 5-Year ACS County-County Commuting Flows, which report the average number of people that commute between two counties, and I construct commuting district homophily as the share of a county's commuters that stay within the county's district when commuting. For county i in district J (which contains counties j) and all US counties K (which contains counties k),

$$\text{Commuting District Homophily}_i = \frac{\sum_{j \in J} \text{Commuters}_{i,j}}{\sum_{k \in K} \text{Commuters}_{i,k}} \tag{10}$$

In Appendix Section C.3, I report results for the effect of commuting district homophily on voters' familiarity with their representatives. Estimates are of smaller magnitudes (about half as large) but otherwise are similar: Commuting district homophily has a positive effect on measures of voters' familiarity with their representatives, as well as voters' likelihood of reporting that they plan to vote for or did vote for the incumbent—with no decrease in support for the opponent, but rather a decrease in having no preference or not voting in the House election. I interpret the smaller estimates as reflecting the fact that commuting flows are a rougher approximation of social networks than the SCI. Additionally, the larger effects when using the SCI to construct district homophily likely also reflect the use of Facebook to share news about representatives. Furthermore, while not reported for brevity, there is again no significant impact of district homophily constructed using commuting flows on the placebo outcomes in Section 5.1. (Effects of commuting district homophily on actual vote counts and campaign contributions to be reported in a future draft.)

5.3 Zip-Code-Level Social Network Data

I construct district homophily using the SCI for zip code-zip code pairs, analogously to how I constructed district homophily using the SCI for county pairs. (To be precise, by "zip code" I refer to Zip Code Tabulation Areas, which are Census statistical areas, rather than the U.S. Postal Service Zip Codes used for mail delivery.) Zip code pairs in most cases provide a much more fine grained measure of social networks (there are an order of magnitude more zip codes in the U.S. than counties), which helps to reduce measurement error. Additionally, because zip codes are usually considerably smaller than counties, a larger share of zip codes intersect only one congressional district, thereby reducing the reliance on population-weighted averages when constructing district homophily. This especially matters in large urban areas where counties are much larger than congressional districts, such as New York City and Los Angeles.

However, they less readily map to many data sources—because zip codes are designed for mail delivery, for most purposes there is no need to align boundaries with zip codes. For example, vote count data is readily available at the county level; voting precincts usually sit within a county, but may intersect multiple zip codes. Additionally, due to privacy concerns, zip codes with small populations do not appear in the SCI data. Accordingly, I use the county-level social networks in the main analysis. Nonetheless, respondents' zip codes are reported in the CES, so I can check whether voter knowledge and self-reported voting outcomes are different when district homophily is constructed using the zip-code-level network data.

Appendix Section C.4 reports the results. The patterns are similar, and estimated effects are slightly higher, consistent with zip-code-level network data reducing measurement error compared to county-level network data. And again, while not reported for brevity, there is again no significant impact of district homophily constructed using zip-code-level networks on the placebo outcomes in Section 5.1.

5.4 Border Pairs Specification

An alternative identification strategy that does not rely so heavily on the 2012 redistricting event is to compare pairs of counties that lie across a district border from each other (Snyder and Strömberg 2010, Spenkuch and Toniatti 2018). The two counties in a pair should be largely similar, except for which district they are assigned to. In particular, because they are in different districts, they will likely have different district homophily levels. Accordingly, we can identify the impact of district homophily by comparing deviations from the county-pair's mean in one county to deviations from the county-pair's mean in the neighboring county. The specification for this design is

$$y_{ct} = \alpha_c + \mu_{pt} + \beta \bar{\pi}_{c,t} + X'_{ct}\delta + \varepsilon_{ct} \tag{11}$$

where y_{ct} is the outcome of interest for county c in year t, μ_{pt} is the pair-by-year fixed effect, β is the coefficient of interest, and X'_{ct} is a vector of time-varying county-level controls. I restrict to counties fully within one district. Because counties can border multiple counties across a district border, I follow Spenkuch and Toniatti 2018 and collapse all outcomes to the county level then include one observation for every pair that a given county is in.

Because the sample becomes quite restricted when we focus only on border counties within one district, precision decreases substantially. I do not include district-by-year fixed effects because there is not enough data to accommodate them, so I instead include state-by-year fixed effects. I also restrict to only comparing pairs within the same state (though results are qualitatively similar when I include all county pairs).

Appendix Table C2 reports the results of the border pairs specification. With the border pairs design, I find qualitatively similar results as in the redistricting design, except estimates on "Selected Party" become insignificant after adding DMA-by-year fixed effects.

6 Conceptual Framework: Information Diffusion within Districts

The empirical results demonstrate a plausibly causal link between district homophily and voter information. To guide the interpretation of the event study estimates, I develop a theoretical framework of information diffusion within congressional districts. This model formalizes the diffusion mechanism through which the network structure determines the equilibrium share of informed voters. I clarify how district homophily is a summary statistic of the network structure.

Consider pieces of news about elected officials arising and spreading in a population. In a given area, what is the steady state share of people who have learned some relevant (i.e., sufficiently recent) news about their elected official? I represent this process using a mean-field approximation, applying Jackson and López-Pintado 2013's model of diffusion with homophily and heterogenous types. A mean-field approximation allows for tractable analysis when the network structure is only known at the group level, such as in the SCI data I use here.

6.1 Types of Individuals

The society consists of a continuum of agents $N = [0, 1]$. Within the society, each agent is assigned a type based on where they live.

In particular, let $\mathcal{D} = [1, ..., D]$ be the set of all congressional districts and $\mathcal{C} = [1, ..., C]$ be the set of counties. There is no ranking between districts and counties: counties can be fully within districts, districts can be fully within counties, or neither. Agents are characterized by the congressional district $d \in \mathcal{D}$ and county $c \in \mathcal{C}$ in which they reside. An agent is of type (c, d) if they live in the intersection of county c and district d. Accordingly, there $C \times D$ possible types. The society is partitioned by type, such that $n(c, d) \in [0, 1]$ is the fraction of agents of type (c, d).

Let $\mathcal{S} = [1, ..., S]$ be the set of states. The set of districts and the set of counties are each partitioned by state, such that each district is fully within one state, as is each county. In particular, each type (c, d) intersects one and only one state s. As such, let $s(d)$ represent the state district d lies in and $s(c)$ represent the state county c lies in. Each state elects one governor and two senators.[22] Thus, I consider the set of offices $\mathcal{O} = [H, G, S_1, S_2]$, with office $o \in \mathcal{O}$. I use "constituency" to refer to a district when discussing those represented by a congressional representative, and to refer to a state when discussing those represented by a governor or senator.

6.2 Friendship Shares between Types

The share of friends each type has of each other type can be described by the matrix

$$\Pi' = \begin{pmatrix} \pi_{(1,1),(1,1)} & \cdots & \pi_{(1,1),(C,D)} \\ \vdots & \ddots & \vdots \\ \pi_{(C,D),(1,1)} & \cdots & \pi_{(C,D),(C,D)} \end{pmatrix} \tag{12}$$

where $\pi_{(c,d)(c',d')} \geq 0$ is the share of type (c, d)'s friends that live in (c', d'). Equivalently, in any given encounter, this is the probability that an agent from county c and district d meets an agent from county c' and district d'. Accordingly, $\sum_{c'=1}^{C} \sum_{d'=1}^{D} \pi_{(c,d)(c',d')} = 1$. Assume that if $\pi_{(c,d)(c',d')} = 0$ then $\pi_{(c',d')(c,d)} = 0$, because friendships are mutual (but observe that otherwise $\pi_{(c,d)(c',d')}$ need not equal $\pi_{(c',d')(c,d)}$). If $c \cap d = \emptyset$, for any c' and d' $\pi_{(c,d)(c',d')} = 0$.

6.3 Information Sharing Process

Any individual can be informed (state 1) or uninformed (state 0) about their own congressional representative, governor, or senators at any given point in time. Individuals only care about news about elected officials that represent them—i.e., officials representing their own district or their own state. Consequently, they only become informed if they receive a piece of news about their *own* elected official.[23] Agents can

[22] Knowledge of one office does not depend on knowledge of any other office, so the model can be applied with one office, or any number of offices. I include these four offices here to map to the offices asked about in the CES.

[23] An agent who is informed about one office need not be informed about other offices.

learn news exogenously (such as from the media) at a rate $\mu_{(c,d)}$, which is allowed to vary by type. Once informed, agents forget the news and become uninformed at rate $\delta > 0$.

Uninformed agents can become informed about an office o if they receive news from a friend from the same constituency who is informed about o. In particular, each period, every agent meets with one friend to receive news. The meeting does not need to be reciprocal: One agent can receive news from the other without the reverse being true.[24] For brevity, I will say that an agent "meets" a friend to mean that an agent "receives news from" a friend, using the two terms interchangeably. While an agent may meet any friend, they are only interested in the news about their own elected officials, as this is the only information that matters for their decision at the ballot box. Accordingly, an uninformed individual becomes informed about their representative (governor and senators) if (i) the friend they meet is from their same district (state), (ii) that friend is informed.

In particular, let $I_o((c,d),(c',d'))$ be an office-specific relevance filter, such that

$$I_o((c,d),(c',d')) = \begin{cases} 1 \text{ if news about office } o \text{ is relevant to both } (c,d) \text{ and } (c',d') \\ 0 \text{ otherwise} \end{cases}$$

Thus, meetings can only transmit information about office o if $I_o((c,d),(c',d')) = 1$.[25]

However, assume that there are some frictions to communicating information, such that when an uninformed agent meets an informed same-constituency friend, the informed friend's news is communicated with probability $\alpha \in (0,1]$.[26]

[24] As Jackson and López-Pintado 2013 explain, assuming that the meetings are reciprocal requires adding the constraint that $n(i,j)\pi_{(i,j)(k,l)} = n(k,l)\pi_{(k,l)(i,j)}$—that is, that the number of interactions from type (i,j) to type (k,l) in a period is the same as the number of interactions from type (k,l) to type (i,j) in a period.

[25] Specifically, for House representatives the filter is defined as

$$I_H((c,d),(c',d')) = \mathbb{1}\{d = d'\}$$

and for governors and senators it is defined as

$$\begin{aligned} I_G((c,d),(c',d')) &= \\ I_{S_1}((c,d),(c',d')) &= \\ I_{S_2}((c,d),(c',d')) &= \mathbb{1}\{s(d) = s(d')\} \\ &= \mathbb{1}\{s(c) = s(c')\} \end{aligned}$$

[26] α captures frictions in communication from both ends of the interaction: both the probability that the recipient of the news does not pay attention to it, as well as the probability that the conveyor of the news fails to pass it on. α is the probability of transmission, conditional on an uninformed agent receiving news from an informed, same-constituency agent.

6.4 Timing

To summarize, consider the process as occurring in discrete periods that each proceed as follows:

1. Begin each period with some agents in state 1 (informed) and some agents in state 0 (uninformed).
2. News is shared. Each agent meets one friend.
3. Uninformed agents who meet an informed friend such that $I_o = 1$ become informed about office o with probability α. (No change occurs if an informed agent meets another informed agent, or meets an agent such that $I_o = 0$, i.e., from a different constituency.)
4. At the end of each period, a share δ of informed individuals become uninformed, and a share $\mu_{(c,d)}$ of uniformed individuals of each type become exogenously informed.

6.5 Individual Transition Probabilities

Per unit of time, what is the probability that an uninformed individual becomes informed about a given office? This will depend on the share of an individual's same-constituency friends that are informed. For clarity, I will first focus on congressional representatives.

Let $\rho^H_{(c,d)}(t)$ denote the probability a type (c,d) agent is informed about their congressional representative at time t. Let $P^H(t) \in [0,1]^C \times D$ represent the matrix with entries $\rho^H_{(c,d)}(t)$. Let $\tilde{\rho}^H_{(c,d)}(t)$ represent the probability that a type (c,d) agent meets a same-district informed friend at time t. That is, $\tilde{\rho}^H_{(c,d)}(t)$ is the share of type (c,d)'s friends that (i) live in the same district, and (ii) are informed about H at time t. $\tilde{\rho}^H_{(c,d)}(t)$ is constructed as the weighted average share of informed friends, with weights given by the friendship shares from Π, and with friends from other districts treated as if they are all uninformed:

$$\tilde{\rho}^H_{(c,d)}(t) = \sum_{c' \in \mathcal{C}} \sum_{d' \in \mathcal{D}} \left(\pi_{(c,d)(c',d')} \times \rho^H_{(c',d')}(t) \times \mathbb{1}\{d' = d\} \right) \tag{13}$$

$$= \sum_{c' \in \mathcal{C}} \left(\pi_{(c,d)(c',d)} \times \rho^H_{(c',d)}(t) \right) \tag{14}$$

Therefore, at time t, the rate that an uninformed type-(c,d) agent transitions to informed is $\alpha \tilde{\rho}^H_{(c,d)}(t) + \mu_{(c,d)}$: the frictions in sharing information multiplied by the probability of meeting a same-district informed friend, plus the probability of becoming exogenously informed. Representing time $t \in \mathbb{R}_+$ as continuous, the dynamics are

$$\frac{d\rho^H_{(c,d)}(t)}{dt} = \underbrace{(1-\rho^H_{(c,d)}(t))}_{\text{Share in state 0}}\underbrace{(\alpha\tilde{\rho}^H_{(c,d)}(t)+\mu_{(c,d)})}_{\text{Rate 0}\rightarrow 1} - \underbrace{\delta\rho^H_{(c,d)}(t)}_{\text{Rate 1}\rightarrow 0\times\text{Share in state 1}} \tag{15}$$

That is, the change in the probability a type (c,d) person is informed is given by the difference between the share of (c,d) people who become newly informed and the share of (c,d) people who become newly uninformed.

We can write analogous terms for the other offices. Consider the governor:[27] We track $\rho^G_{(c,d)}(t)$, the probability that a type (c,d) agent is informed about their governor at time t. $\tilde{\rho}^G_{(c,d)}(t)$ represents the probability that a type (c,d) agent meets a *same-state* friend informed about G at time t:

$$\tilde{\rho}^G_{(c,d)}(t) = \sum_{c'\in\mathcal{C}}\sum_{d'\in\mathcal{D}}\Big(\pi_{(c,d)(c',d')}\times\rho^G_{(c',d')}(t)\times\mathbb{1}\{s(c)=s(c')\}\Big) \tag{16}$$

Accordingly, the dynamics for the state-wide offices can be written analogously to Equation 15.

6.6 Steady State

In the steady state, $\frac{d\rho^o_{(c,d)}(t)}{dt}=0$ for all (c,d) and all o. Consequently, at the steady state, the probability that an individual of type (c,d) is informed about o is given by

$$\rho^o_{(c,d)} = \frac{\alpha\tilde{\rho}^o_{(c,d)}+\mu_{(c,d)}}{\alpha\tilde{\rho}^o_{(c,d)}+\mu_{(c,d)}+\delta} \tag{17}$$

6.7 Aggregating to County-Level

The model above describes information diffusion at the level of county-district types (c,d). In the data, however, I can observe the fixed county-level friendship shares and population shares across districts, but not the full type-level adjacency matrix. To connect the model to the data, I construct the aggregated county steady state probabilities by taking population-weighted averages.[28]

[27] The terms for the senators are written in the corresponding way.

[28] Note that such an aggregation is not strictly necessary for the statewide offices, but I apply it for all offices here to form the connection to the event study estimates of the impact of district homophily. Indeed, for the statewide offices the county-county friendship matrix Π can immediately be used instead of Π' because the way districts divide counties does not affect whether two individuals are in the same state. As such, for the statewide offices, steady states at the county level can be written analogously to the county-district steady states. By replacing types (c,d) everywhere with types c, for $o'\in[G,S_1,S_2]$ we arrive at a steady state share of voters in each county informed about office o':

$$\rho^{o'}_c = \frac{\alpha\tilde{\rho}^{o'}_c+\mu_c}{\alpha\tilde{\rho}^{o'}_c+\mu_c+\delta}$$

Let $\pi_{c,c'}$ represent the probability that an agent from county c meets an agent from county c' (regardless of district) in any given meeting. The county-county friendship matrix (as introduced in Equation 1) summarizes these probabilities:

$$\Pi = \begin{pmatrix} \pi_{1,1} & \cdots & \pi_{1,C} \\ \vdots & \ddots & \vdots \\ \pi_{C,1} & \cdots & \pi_{C,C} \end{pmatrix} \tag{18}$$

Let $D(c)$ be the set of districts that county c intersects with. Recall that $n(c,d)$ is the fraction of agents of type (c,d). Then, the share of county c's population in each district d it intersects is

$$q_{(c,d)} = \frac{n(c,d)}{\sum_{d' \in D(c)} n(c,d')} \tag{19}$$

As such, $\sum_{d \in D(c)} q_{(c,d)} = 1$. Let $Q \in [0,1]^{C \times D}$ represent the matrix of the population shares $q_{(c,d)}$, and observe that $q_{(k,d)} = 0$ whenever $k \cap d = \emptyset$.

Assume that friendships are uniformly distributed within a county. Then, $\pi_{(c,d)(c',d')} = \pi_{c,c'} \times q_{(c',d')}$: the probability a type (c,d) agent meets a type (c',d') agent is approximated by the probability an agent from county c meets an agent from county c', multiplied by the probability that an agent living in c' also lives in d'.

The share of people in a given county that are informed about office o at time t can then be constructed as $\rho_c^o(t) = \sum_{d \in D(c)} \left(q_{(c,d)} \times \rho_{(c,d)}^o(t)\right)$: The population-weighted average share of informed people, summing across each district the county intersects. Equivalently, we can write

$$\begin{aligned} \rho_c^o(t) &= \sum_{c' \in \mathcal{C}} \sum_{d \in D(c)} \left(\pi_{(c,d)(c',d')} \times q_{(c,d)} \times \rho_{(c',d)}^o(t)\right) && (20) \\ &= \sum_{c' \in \mathcal{C}} \sum_{d \in D(c)} \left(\pi_{c,c'} \times q_{(c,d)} \times q_{(c',d)} \times \rho_{(c',d)}^o(t)\right) && (21) \\ &= \sum_{c' \in \mathcal{C}} \left[\Pi \circ \left(Q \left(Q \circ P^o(t)\right)^T\right)\right]_{c,c'} && (22) \end{aligned}$$

Recall that a county's district homophily is the probability that a person from that county meets a same-district friend: district homophily for county c is represented by $\bar{\pi}_c = \sum_{c' \in \mathcal{C}} \sum_{d \in D(c)} \left(\pi_{c,c'} \times q_{(c,d)} \times q_{(c',d)}\right)$, or equivalently

$$\bar{\pi}_c = \sum_{c' \in \mathcal{C}} \left[\Pi \circ (QQ^T)\right]_{c,c'} \tag{23}$$

The framework developed here clarifies that district homophily is mechanically induced by the intersection between social networks and district boundaries. Future work will focus on estimating the parameters of the diffusion process, which would allow for the simulation of counterfactual district maps. In particular, the parameters governing information transmission (α and δ) could be pinned down using the event study estimates of the effect of changes in district homophily on the county-level shares of voters informed about their House representative, governor, and both senators—leveraging the fact that for the latter three, empirically there is no effect. The exogenous information rates ($\mu_{(c,d)}$) could be modeled as functions of observable characteristics to capture systematic differences in the baseline likelihood of being informed. These parameter estimates could be accomplished via indirect inference (Gourieroux et al., 1993): the diffusion process would be simulated at the (c,d)-level to steady state (before and after redistricting), aggregated to the county level, and the implied coefficients compared to their empirical counterparts.

7 Comparisons Across Many Simulated Maps

McCartan, Kenny, Simko, Kuriwaki, et al. 2021 simulate 5,000 congressional district maps for each of the 50 states. They use Monte Carlo simulation. Their maps are constrained to follow the given state's redistricting laws. The goal of these simulations is to construct a benchmark that accounts for a state's political geography—the simulated maps allow comparing a state against counterfactual versions of itself, with all structural aspects of the state's population distribution and physical geography accounted for. While drawing an optimal map is known to be very difficult, their dataset instead enables looking across a distribution of feasible alternative maps, and comparing a feature of interest of a given map against the distribution of the feature across the sampled maps. For example, the authors calculate commonly used measures of gerrymandering for each map, and identify gerrymandered maps as those that fall outside of the distribution of the measure calculated on the simulated maps. I use maps from their database and calculate district homophily under counterfactual maps.

With the simulated maps in hand, I study how the levels of and inequality in district homophily under current congressional district maps compare to the distribution of simulated maps—how much are we leaving on the table, compared to reasonable alternatives?[29] In Appendix Section C.9, I explore the features of the distribution of district homophily in the simulated maps.

[29] Note that the simulated maps do not contain *all* possible maps. For example, the distributions need not contain the district homophily-maximizing map.

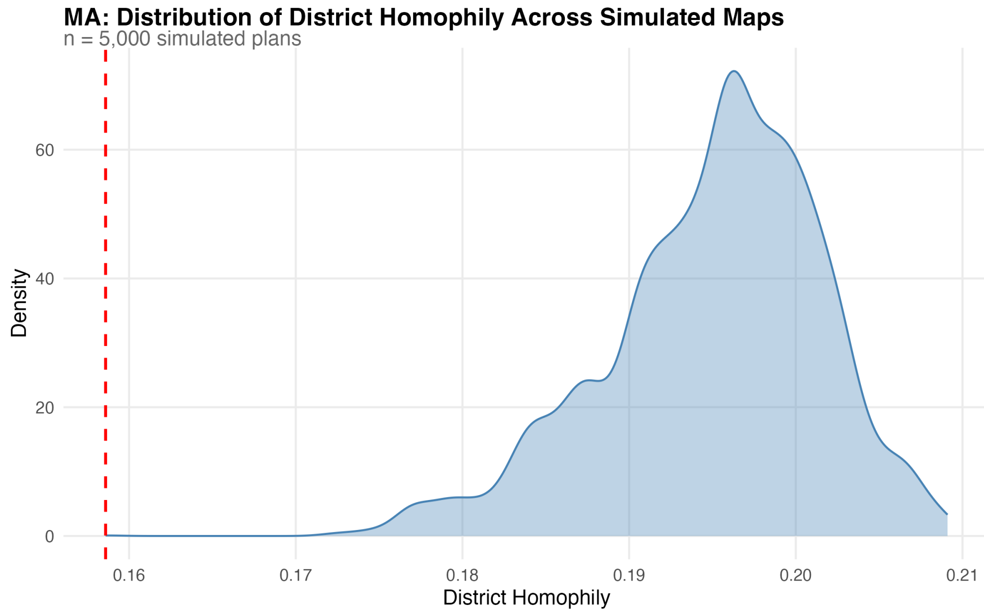


Figure 9: Massachusetts District Homophily: Enacted Map vs. Simulated Maps

7.1 District Homophily of Enacted Maps Compared to Simulated Maps

I explore how the congressional district maps enacted following the 2020 Census (which I refer to as the "enacted maps", though some states have since re-drawn their boundaries at least once) compare to the distribution of simulated maps in terms of district homophily.

For each state, I compare the state-wide average district homophily of the enacted map to the distribution of state-wide average district homophily across the 5,000 simulated maps. Figure 9 shows an example of this for Massachusetts. (Appendix Section C.11 shows the distribution for each state.)

Average district homophily under the enacted map in Massachusetts is just shy of 16% (represented by the vertical dashed line), but the simulated maps have average district homophily ranging from 17% to 21%. As such, Massachusetts' enacted map lies entirely below the distribution.

Accordingly, for each state I can measure the enacted map's percentile in the distribution. Figure 10 plots a histogram of these percentiles. About half of states (21 out of 43) rank in the 1st percentile—the district homophily of their enacted map lies entirely below the distribution of simulated maps. Only a handful of states have an enacted maps that lie above the median.

Figure 11 shows the percentile of each state on a map. Indiana, Minnesota, and West Virginia have the highest percentile maps, meaning that their enacted maps have higher district homophily than the vast majority of the simulated maps.

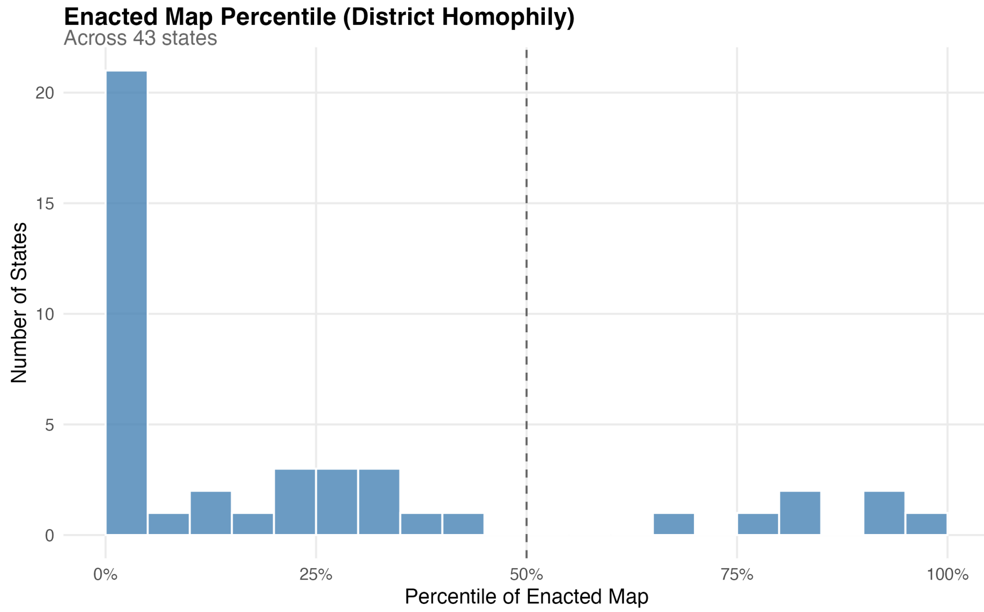


Figure 10: Percentile of Enacted Map in terms of District Homophily

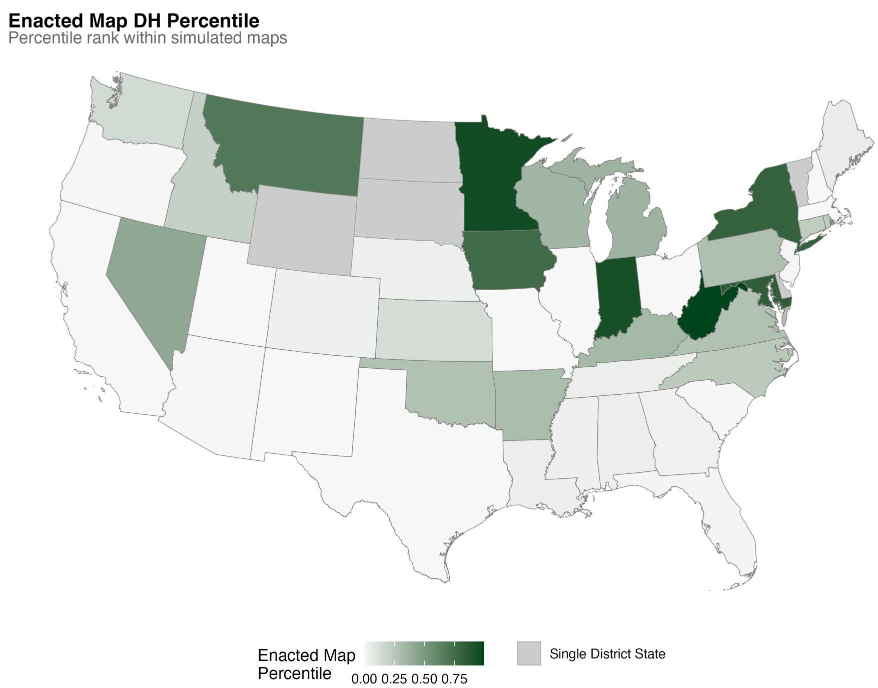


Figure 11: Percentile of Enacted Map in terms of District Homophily

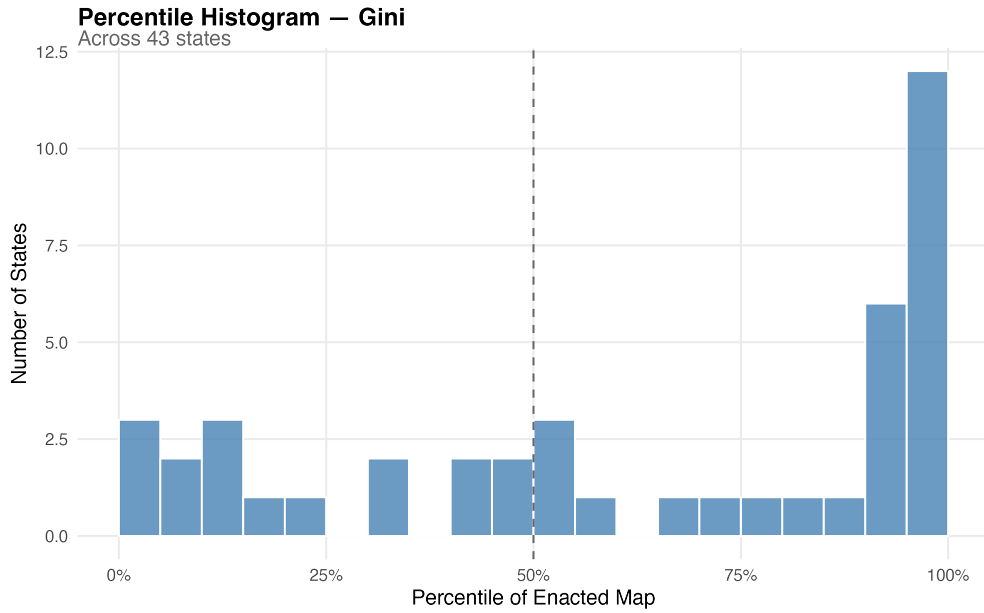


Figure 12: Percentile of Enacted Map in terms of Gini Coefficient of County District Homophily

A perhaps even more pressing concern is how enacted maps compare to the simulated maps in terms of within state inequality in district homophily across counties. Figure 12 shows a histogram of the percentile of each state in terms of the state's Gini coefficient for county district homophily. More states lie within the simulated maps' distributions of Gini coefficients, though still over a quarter of states (12) are more unequal than all of the simulated maps.

Figure 13 plots the percentile of the Gini coefficient for each state. In general, there is considerable overlap between states with very low percentiles in terms of district homophily and states with very high percentiles in terms of the Gini coefficient. This correlation is confirmed in Figure 14, which plots a scatter plot of the district homophily percentile vs. the Gini coefficient percentile.

For a given state, we can also consider where the enacted map falls as compared to the simulated maps in terms of the trade-off between inequality and average district homophily. Figure 15 shows an example of this for Massachusetts. (Appendix Section C.11 shows the figure for each state.) The red dot represents the enacted map, and each blue dot represents a simulated map. The x-axis plots the Gini coefficient of county district homophily, and the y-axis plots the statewide average district homophily. Assuming a preference for a higher average district homophily and more equality (a lower Gini), the top-left corner is better on these dimensions.

In sum, the enacted maps largely under perform the feasible, neutrally-drawn simulated maps in terms

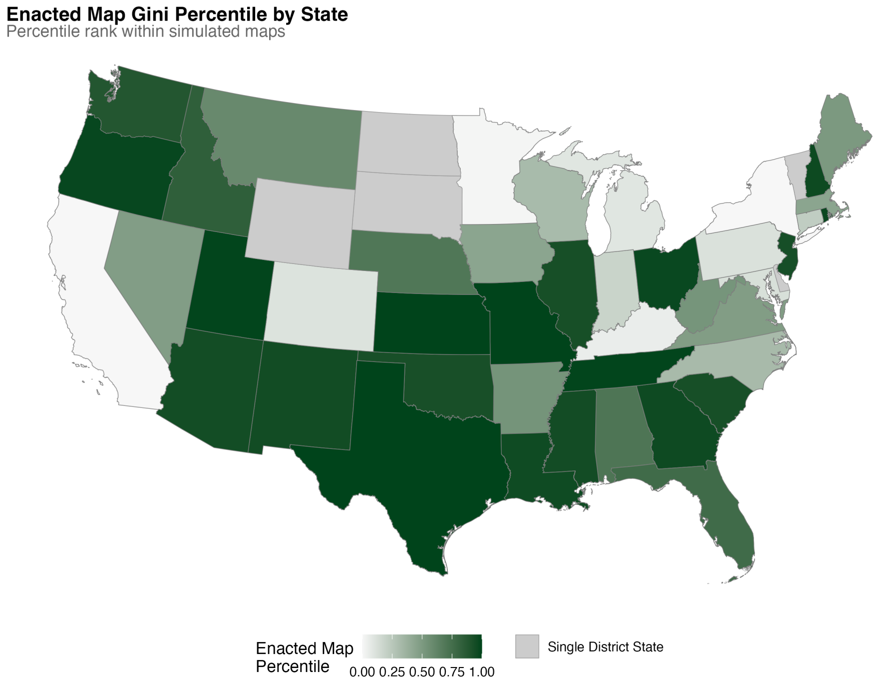


Figure 13: Percentile of Enacted Map in terms of Gini Coefficient of County District Homophily

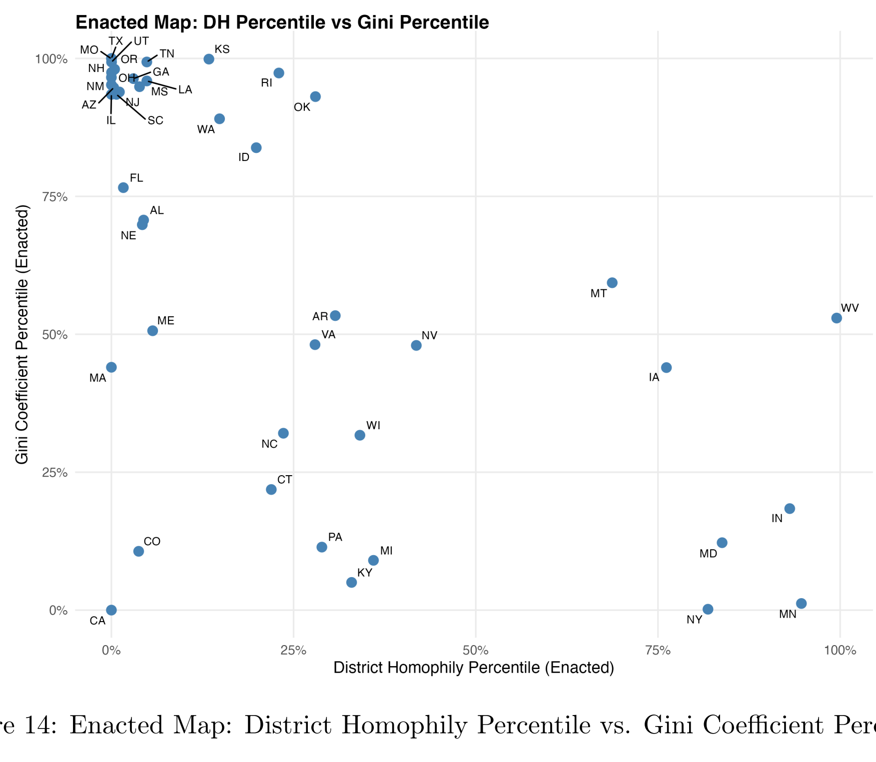


Figure 14: Enacted Map: District Homophily Percentile vs. Gini Coefficient Percentile

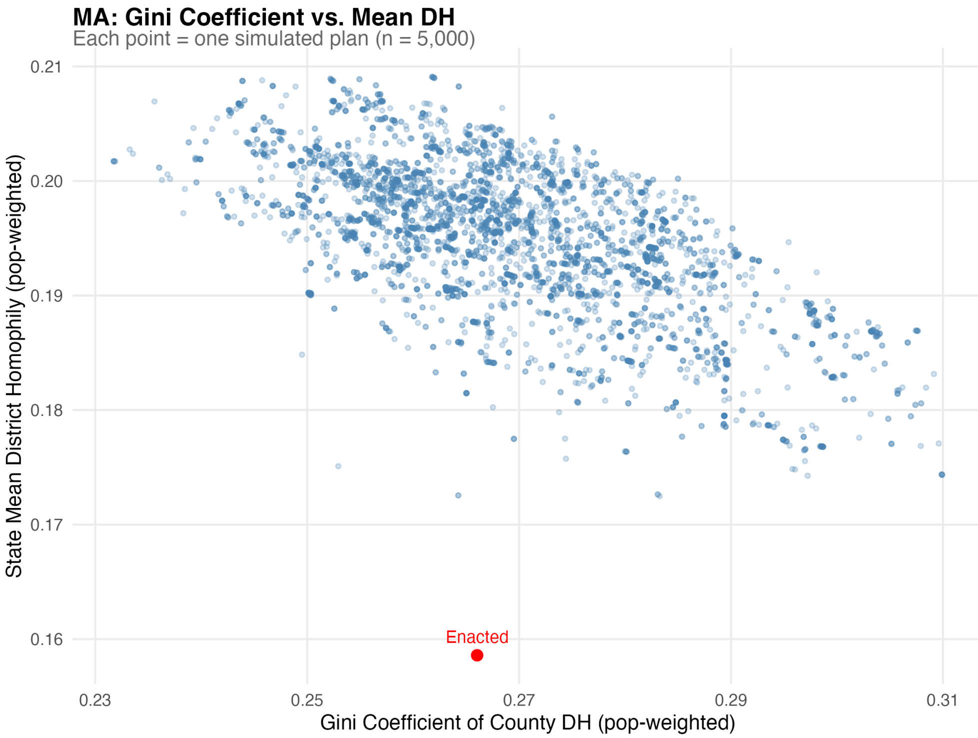


Figure 15: Massachusetts District Homophily: Inequality vs. Average

of both the average level of district homophily and the within-state inequality in district homophily.

8 Conclusion

Communities across the U.S. vary substantially in their social cohesiveness with their congressional district, i.e., their district homophily. While people living in the average county live in the same district as about half of their friends, this varies from 2% to 87%. I show that district homophily increases voters' familiarity with their representative: when a county's district homophily increases due to redistricting, voters are more likely to recognize the name and know the party of their representative. I find similar results for voter information regardless of whether I construct the network using the SCI or commuting flows, which strengthens the case that these impacts are not unique to Facebook users. I also find that district homophily decreases rates of abstaining in House elections. Additionally, when district homophily increases, donors shift contributions towards House candidates in their own districts, away from House candidates running in other districts. These results prompt future research into how candidates respond to district homophily in their campaign strategies, to build a general equilibrium understanding of the social learning consequences of how district boundaries are drawn. This evidence is especially important as detailed social network data, like the SCI, has become publicly available for the first time in recent years—enabling its use by policymakers to draw fairer districts, but also by partisan gerrymanderers who may seek to exploit it.

References

Alt, J. E., Jensen, A., Larreguy, H., Lassen, D. D., & Marshall, J. (2022). Diffusing Political Concerns: How Unemployment Information Passed between Social Ties Influences Danish Voters. *The Journal of Politics*, *84*(1), 383–404. https://doi.org/10.1086/714925

Arias, E., Balán, P., Larreguy, H., Marshall, J., & Querubín, P. (2019). Information Provision, Voter Coordination, and Electoral Accountability: Evidence from Mexican Social Networks. *American Political Science Review*, *113*(2), 475–498. https://doi.org/10.1017/S0003055419000091

Auxier, B., & Anderson, M. (2021, April). Social Media Use in 2021.

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., & Wong, A. (2018). Social Connectedness: Measurement, Determinants, and Effects. *Journal of Economic Perspectives*, *32*(3), 259–280. https://doi.org/10.1257/jep.32.3.259

Bailey, M., Gupta, A., Hillenbrand, S., Kuchler, T., Richmond, R., & Stroebel, J. (2020). International Trade and Social Connectedness, 55.

Bailey, M., Gupta, A., Hillenbrand, S., Kuchler, T., Richmond, R., & Stroebel, J. (2021). International trade and social connectedness. *Journal of International Economics*, *129*, 103418. https://doi.org/10.1016/j.jinteco.2020.103418

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). The Diffusion of Microfinance. *Science*, *341*(6144). https://doi.org/10.1126/science.1236498

Baybeck, B., & Huckfeldt, R. (2002). Urban contexts, spatially dispersed networks, and the diffusion of political information. *Political Geography*, *21*(2), 195–220. https://doi.org/10.1016/S0962-6298(01)00056-7

Beaman, L., BenYishay, A., Magruder, J., & Mobarak, A. M. (2021). Can Network Theory-based Targeting Increase Technology Adoption?, 60.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*(7415), 295–298. https://doi.org/10.1038/nature11421

Bonica, A. (2014). Mapping the Ideological Marketplace. *American Journal of Political Science*, *58*(2), 367–386. https://doi.org/10.1111/ajps.12062

Bouton, L., Genicot, G., Castanheira, M., & Stashko, A. (2023). Pack-Crack-Pack: Gerrymandering with Differential Turnout.

Brown, J. R., Cantoni, E., Chinoy, S., Koenen, M., & Pons, V. (2023). The Effect of Childhood Environment on Political Behavior: Evidence from Young U.S. Movers, 1992–2021. *NBER Working Paper*. https://doi.org/http://www.nber.org/papers/w31759

Campbell, D. E. (2013). Social Networks and Political Participation. *Annual Review of Political Science*, *16*(1), 33–48. https://doi.org/10.1146/annurev-polisci-033011-201728

Cantoni, E., & Pons, V. (2022). Does Context Outweigh Individual Characteristics in Driving Voting Behavior? Evidence from Relocations within the United States. *American Economic Review*, *112*(4), 1226–1272. https://doi.org/10.1257/aer.20201660

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., Johnston, D., Koenen, M., Laguna-Muggenburg, E., Mudekereza, F., Rutter, T., Thor, N., Townsend, W., Zhang, R., Bailey, M., . . . Wernerfelt, N. (2022). Social capital I: Measurement and associations with economic mobility. *Nature*, *608*(7921), 108–121. https://doi.org/10.1038/s41586-022-04996-4

Conley, T. G., & Udry, C. R. (2010). Learning about a New Technology: Pineapple in Ghana. *American Economic Review*, *100*(1), 35–69. https://doi.org/10.1257/aer.100.1.35

David Leip. (2024). Dave Leip's Atlas of U.S. Presidential Elections.

Delli Carpini, M. X., & Keeter, S. (1996). *What Americans know about politics and why it matters* (3. [pr.]). Yale Univ. Press.

DiPrete, T. A., Gelman, A., McCormick, T., Teitler, J., & Zheng, T. (2011). Segregation in Social Networks Based on Acquaintanceship and Trust. *American Journal of Sociology*, *116*(4), 1234–1283. https://doi.org/10.1086/659100

Druckman, J. N., Levendusky, M. S., & McLain, A. (2018). No Need to Watch: How the Effects of Partisan Media Can Spread via Interpersonal Discussions. *American Journal of Political Science*, *62*(1), 99–112. https://doi.org/10.1111/ajps.12325

Echenique, F., & Fryer, R. G. (2007). A Measure of Segregation Based on Social Interactions. *The Quarterly Journal of Economics*, *122*(2), 441–485. https://doi.org/10.1162/qjec.122.2.441

Eckman, S. J. (2021, November). *Apportionment and Redistricting Process for the U.S. House of Representatives* (tech. rep. No. R45951). Congressional Research Service.

Egorov, G., Guriev, S., Mironov, M., & Zhuravskaya, E. (2026). Political Information and Network Effects. *Journal of the European Economic Association*, *24*(1), 82–121. https://doi.org/10.1093/jeea/jvaf052

Eisensee, T., & Strömberg, D. (2007). News Droughts, News Floods, and U. S. Disaster Relief. *The Quarterly Journal of Economics*, *122*(2), 693–728. https://doi.org/10.1162/qjec.122.2.693

Enke, B., Fisman, R., Freitas, L. M., & Sun, S. (2023, May). *Universalism and Political Representation: Evidence from the Field* (tech. rep. No. w31265). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w31265

Fafchamps, M., Vaz, A., & Vicente, P. C. (2019). Voting and Peer Effects: Experimental Evidence from Mozambique. *Economic Development and Cultural Change*.

Falck, O., Gold, R., & Heblich, S. (2014). E-lections: Voting Behavior and the Internet. *American Economic Review*, *104*(7), 2238–2265. https://doi.org/10.1257/aer.104.7.2238

Feddersen, T. J., & Pesendorfer, W. (1996). The Swing Voter's Curse. *The American Economic Review*, *86*(3), 408–424.

Ferraz, C., & Finan, F. (2008). Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes. *The Quarterly Journal of Economics*, *123*(2), 703–745. https://doi.org/10.1162/qjec.2008.123.2.703

Fowler, J. H., Heaney, M. T., Nickerson, D. W., Padgett, J. F., & Sinclair, B. (2011). Causality in Political Networks. *American Politics Research*, *39*(2), 437–480. https://doi.org/10.1177/1532673X10396310

Friedman, J. N., & Holden, R. (2020). Optimal Gerrymandering in a competitive environment. *Economic Theory Bulletin*, *8*(2), 347–367. https://doi.org/10.1007/s40505-020-00188-3

Friedman, J. N., & Holden, R. T. (2008). Optimal Gerrymandering: Sometimes Pack, But Never Crack. *American Economic Review*, *98*(1), 113–144. https://doi.org/10.1257/aer.98.1.113

Gerber, A. S., Green, D. P., & Larimer, C. W. (2008). Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment. *American Political Science Review*, *102*(1), 33–48. https://doi.org/10.1017/S000305540808009X

Gourieroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, *8*(S1), S85–S118. https://doi.org/10.1002/jae.3950080507

Gul, F., & Pesendorfer, W. (2010). Strategic Redistricting. *American Economic Review*, *100*(4), 1616–1641. https://doi.org/10.1257/aer.100.4.1616

Guriev, S., Melnikov, N., Silva, J. E., & Zhuravskaya, E. (2026). Mobile Broadband and the Decline of Incumbency Advantage.

Guriev, S., Melnikov, N., & Zhuravskaya, E. (2021). 3G Internet and Confidence in Government. *The Quarterly Journal of Economics*, *136*(4), 2533–2613. https://doi.org/10.1093/qje/qjaa040

Hampton, K. N., Goulet, L. S., & Marlow, C. (2012, February). *Why most Facebook users get more than they give* (tech. rep.). Pew Research Center.

Jackson, M. O., & López-Pintado, D. (2013). Diffusion and contagion in networks with heterogeneous agents and homophily. *Network Science*, *1*(1), 49–67. https://doi.org/10.1017/nws.2012.7

Klar, S., & Shmargad, Y. (2017). The Effect of Network Structure on Preference Formation. *The Journal of Politics*, *79*(2), 717–721. https://doi.org/10.1086/689972

Klofstad, C. A. (2007). Talk Leads to Recruitment: How Discussions about Politics and Current Events Increase Civic Participation. *Political Research Quarterly*, *60*(2), 180–191. https://doi.org/10.1177/1065912907301708

Kolotilin, A., & Wolitzky, A. (2020). The Economics of Partisan Gerrymandering. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3698690

Kuchler, T., Li, Y., Peng, L., Stroebel, J., & Zhou, D. (2022). Social Proximity to Capital: Implications for Investors and Firms (I. Goldstein, Ed.). *The Review of Financial Studies*, *35*(6), 2743–2789. https://doi.org/10.1093/rfs/hhab111

Kuriwaki, S. (2018). Cumulative CES Common Content. https://doi.org/10.7910/DVN/II2DB6

Kuziemko, I., Longuet-Marx, N., & Naidu, S. (2023, October). *"Compensate the Losers?" Economic Policy and Partisan Realignment in the US* (tech. rep. No. w31794). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w31794

Lazarsfeld, P., Berelson, B., & Gaudet, H. (1944). *The people's choice: How the voter makes up his mind in a presidential campaign*. Duell, Sloan and Pearce.

Manacorda, M., & Tesei, A. (2020). Liberation Technology: Mobile Phones and Political Mobilization in Africa. *Econometrica*, *88*(2), 533–567. https://doi.org/10.3982/ECTA14392

McCartan, C., & Imai, K. (2023). Sequential Monte Carlo for sampling balanced and compact redistricting plans. *The Annals of Applied Statistics*, *17*(4). https://doi.org/10.1214/23-AOAS1763

McCartan, C., Kenny, C. T., Simko, T., Ebowe, E., Zhao, M. Y., & Imai, K. (2024). Redistricting Reforms Reduce Gerrymandering by Constraining Partisan Actors. https://doi.org/10.48550/ARXIV.2407.11336

McCartan, C., Kenny, C. T., Simko, T., Kuriwaki, S., Garcia, G., Wang, K., Wu, M., & Imai, K. (2021). 50-State Redistricting Simulations. https://doi.org/10.7910/DVN/SLCD3E

McCormick, T. H., Salganik, M. J., & Zheng, T. (2010). How Many People Do You Know?: Efficiently Estimating Personal Network Size. *Journal of the American Statistical Association*, *105*(489), 59–70. https://doi.org/10.1198/jasa.2009.ap08518

Miller, M. G. (2022). Candidate Extremism and Voter Roll-Off in US House Elections. *Legislative Studies Quarterly*, *47*(4), 791–821. https://doi.org/10.1111/lsq.12356

Nickerson, D. W. (2008). Is Voting Contagious? Evidence from Two Field Experiments. *American Political Science Review*, *102*(1), 49–57. https://doi.org/10.1017/S0003055408080039

Owen, G., & Grofman, B. (1988). Optimal partisan gerrymandering. *Political Geography Quarterly*, *7*(1), 5–22. https://doi.org/10.1016/0260-9827(88)90032-8

Pons, V. (2018). Will a Five-Minute Discussion Change Your Mind? A Countrywide Experiment on Voter Choice in France. *American Economic Review*, *108*(6), 1322–1363. https://doi.org/10.1257/aer.20160524

Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press.
OCLC: ocm71173667.

Prior, M. (2019). *Hooked: How politics captures people's interest*. Cambridge University Press.

Schaffner, B., Ansolabehere, S., & Luks, S. (2021). Cooperative Election Study Common Content, 2020. https://doi.org/10.7910/DVN/E9N6PH

Schaffner, B., Ansolabehere, S., & Shih, M. (2023). Cooperative Election Study Common Content, 2022. https://doi.org/10.7910/DVN/PR4L8P

Sinclair, B. (2012). *The social citizen: Peer networks and political behavior*. The University of Chicago Press.

Sinclair, B., McConnell, M., & Green, D. P. (2012). Detecting Spillover Effects: Design and Analysis of Multilevel Experiments. *American Journal of Political Science*, *56*(4), 1055–1069.

Snyder, J. M., & Strömberg, D. (2010). Press Coverage and Political Accountability. *Journal of Political Economy*, *118*(2), 355–408. https://doi.org/10.1086/652903

Sokhey, A. E., & Djupe, P. A. (2011). Interpersonal Networks and Democratic Politics. *PS: Political Science & Politics*, *44*(1), 55–59. https://doi.org/10.1017/S1049096510001861

Sokhey, A. E., & McClurg, S. D. (2012). Social Networks and Correct Voting. *The Journal of Politics*, *74*(3), 751–764. https://doi.org/10.1017/S0022381612000461

Spenkuch, J. L., & Toniatti, D. (2018). Political Advertising and Election Results*. *The Quarterly Journal of Economics*, *133*(4), 1981–2036. https://doi.org/10.1093/qje/qjy010

Stephanopoulos, N. O., & McGhee, E. M. (2015). Partisan Gerrymandering and the Efficiency Gap. *University of Chicago Law Review*.

Strömberg, D. (2004). Radio's Impact on Public Spending. *The Quarterly Journal of Economics*, *119*(1), 189–221. https://doi.org/10.1162/003355304772839560

Vogels, E., Auxier, B., & Anderson, M. (2021, April). Partisan differences in social media use show up for some platforms, but not Facebook.

Whitaker, L. P. (2017, March). *Congressional Redistricting Law: Background and Recent Court Rulings* (tech. rep. No. R44798). Congressional Research Service.

Appendix

A Model Details and Derivations

A.1 Constructing Weighted Average District Homophily

For any person in district d, there are two general types of people they befriend: people who are also in district d, and people who are outside of the district (call them members of $nd \equiv \mathcal{D} \setminus d$).[30] To take a shortcut with notation, say that any person who is in county c and in district d has (on average) a share $\pi^{d}_{(c,d)}$ of friends in district d, and they have a share $\pi^{nd}_{(c,d)} = 1 - \pi^{d}_{(c,d)}$ of friends outside of their district, where $\pi^{d}_{(c,d)} = \sum_{c' \in \mathcal{C}} \pi_{(c,d)(c',d)}$, and $\pi^{nd}_{(c,d)} = \sum_{c' \in \mathcal{C}} \sum_{f \in \mathcal{D} \setminus d} \pi_{(c,d)(c',f)}$.

To construct $\pi^{d}_{(c,d)}$ (the share of type (c,d)'s friends that also live in d), I sum as follows:

$$\pi^{d}_{(c,d)} = \sum_{c' \in \mathcal{C}} \pi_{(c,d)(c',d)} \tag{24}$$

$$= \sum_{c' \in \mathcal{C}} \sum_{d' \in D(c)} \left(\pi_{(c,d)(c',d')} \times \mathbb{I}\{d' = d\} \right) \tag{25}$$

$$= \sum_{c' \in \mathcal{C}} \sum_{d' \in D(c)} \left(\pi_{c,c'} \times q_{(c',d')} \times \mathbb{I}\{d' = d\} \right) \tag{26}$$

$$= \sum_{c' \in \mathcal{C}} \left(\pi_{c,c'} \times q_{(c',d)} \right) \tag{27}$$

Next, define $\bar{\pi}_c$ as the share of county c's friendships that are between people in the same district. Put differently, $\bar{\pi}_c$ represents the probability that a randomly chosen person from county c interacts with a person from their own district (without conditioning on which district $d \in D(c)$ the county c person is from). We can construct $\bar{\pi}_c$ as follows:

[30] Both people in or outside the district may be in the same or different counties: for example, my county may be split across two districts (such that there are other people in my same county but in a different district), or my district may contain multiple counties (such that there are people in my same district but in a different county).

$$\bar{\pi}_c = \sum_{d \in D(c)} \left(\pi^d_{(c,d)} \times q_{(c,d)} \right) \tag{28}$$

$$= \sum_{d \in D(c)} \sum_{c' \in \mathcal{C}} \sum_{d' \in D(c')} \left(\pi_{c,c'} \times q_{(c,d)} \times q_{(c',d')} \times \mathbb{I}\left\{ d' = d \right\} \right) \tag{29}$$

$$= \sum_{d \in D(c)} \sum_{c' \in \mathcal{C}} \left(\pi_{c,c'} \times q_{(c,d)} \times q_{(c',d)} \right) \tag{30}$$

$$= \sum_{c' \in \mathcal{C}} \left[(QQ^T) \circ \Pi' \right]_{c,c'} \tag{31}$$

where Q represents the matrix of the population shares $q_{(c,d)}$, and observing that $q_{(k,d)} = 0$ whenever $k \cap d = \emptyset$.

$\bar{\pi}_c$ is district homophily, or the share of friends that live in the same district. $\pi^d_{(c,d)}$ is analogous to district homophily, but specifically for people living in county c and district d (i.e., for people in county c and district d, the share of their friends that live in district d).

A.2 Local Uniqueness of District Homophily

This section establishes a local mapping result that clarifies the relationship between district homophily and the underlying county-district assignment. While the result is limited in scope, it illustrates the conditions under which district homophily is informative about the structure of the underlying network. This section builds towards a foundation for future estimation of the parameters of the information diffusion process.

The diffusion process in the conceptual framework operates at the (c, d) level, while the empirical analysis aggregates to the county level. Because counties may intersect multiple districts, county-level aggregation generally obscures the same-district structure of information transmission, making the evolution of county-level shares non-Markov. To exploit variation from redistricting, I require units of analysis that do not change across maps. Accordingly, I focus on county-level outcomes ρ^o_c, because counties are stable across redistricting and align with the available SCI data.

While Equations 22 and 23 show that district homophily and county-level information shares are closely related, district homophily is not, in general, a sufficient statistic for the full assignment matrix. Distinct countyâdistrict assignment matrices Q can generate the same homophily vector $\bar{\pi}$. Nonetheless, institutional features of U.S. congressional districts substantially restrict the set of feasible assignment matrices. Under these restrictions, I derive conditions under which district homophily is locally informative about the underlying assignment matrix.

First, because districts and counties are contained within states, the assignment matrix Q is block-diagonal by state. All of a county's population must be assigned to some district, such that for each county c, $\sum_{d:s(d)=s(c)} q_{(c,d)} = 1$.

Second, the finite number of states, districts, and counties further constrains the set of feasible values of Q. While some states have many districts (California has the maximum, with 52 following the 2020 Census), many have very few (13 states have 3 or fewer, and 13 have 10 or more, following the 2020 Census). Combined with the constraints that districts be geographically contiguous and have roughly equally populations, the partition of counties into relatively few districts induces several regularities in the empirically observed distributions of the values of Q. Empirically, many counties lie entirely within a single district. Let $\mathcal{C}1 = c : \exists d$ such that $q(c,d) = 1$. This is true for nearly 90% of U.S. counties in the contiguous 48 states. For such counties, $\rho_{(c,d)} = \rho_c$ exactly, and $\bar{\pi}_c = \sum_{c'} \left(\pi_{c,c'} \times q_{(c',d)}\right)$.

Under these institutional restrictions, I derive a local mapping result relating district homophily to the underlying assignment matrix. The result establishes that, to first order, small changes in the countyâdistrict assignment that are consistent with institutional and topological constraints produce distinct changes in district homophily.[31] Intuitively, when counties are not heavily fragmented across districts, district homophily captures the variation in feasible assignments that will be relevant for information flows.

Proposition 1 (State-conditional local injectivity of district homophily)**.** *Fix a state s. Let C_s be the set of counties in state s and D_s be the set of districts in state s. Let $Q_s^0 = \left(q^0_{(c,d)}\right)_{c\in C_s, d\in D_s}$ be a feasible county-district assignment matrix satisfying (a) $q^0_{(c,d)} \geq 0$ and (b) $\sum_{d\in D_s} q^0_{(c,d)} = 1$.*

Write district homophily for county c as $\bar{\pi}_c(Q_s) = \sum_{c'\in C_s} \sum_{d\in D_s} \pi_{c,c'} q_{(c,d)} q_{(c',d)}$, and let $\bar{\pi}^s(Q_s) \in [0,1]^{|C_s|}$ be the vector of these values.

Suppose the following assumptions hold:

1. ***Limited fragmentation:*** *There exists $\varepsilon_s \in (0,1)$ such that for all $c \in C_s$, $\max_d q^0_{(c,d)} \geq 1 - \varepsilon_s$. Refer to the district satisfying this condition as the "dominant district" $d(c)$.*

2. ***Simple boundaries:*** *The map configuration is locally simple. For the observed map Q_s^0, every county intersects at most two districts ($|\{d : q^0_{(c,d)} > 0\}| \leq 2$). Furthermore, we restrict attention to alternative feasible maps Q_s in the neighborhood of Q_s^0 that preserve this topological constraint (every county maintains support in at most two districts).*

[31] One condition is that no county is split across more than two districts. While this sounds restrictive, it is true for over 95% of U.S. counties. Only half-a-percent of counties are split across more than four districts.

3. ***Non-trivial out-of-state connections:*** *The state social network is an open system. For every connected component of the within-state social network* Π_s*, there is strictly positive leakage to out-of-state counties. Formally, this implies the spectral radius satisfies* $\rho(\Pi_s) < 1$.

4. ***Network non-degeneracy:*** *The within-state friendship matrix* Π_s *is non-degenerate. Specifically, the rows* $\{\pi_{c,\cdot}\}_{c \in C_s}$ *are linearly independent.*

Then there exists $\bar{\varepsilon}_s > 0$ *such that if* $\varepsilon_s < \bar{\varepsilon}_s$*, the mapping* $Q_s \mapsto \bar{\pi}^s$ *is locally injective at* Q_s^0 *within the set of maps satisfying Simple Boundaries. That is, there is no other feasible map arbitrarily close to* Q_s^0 *that reproduces the same district homophily vector.*

Proof. Fix a state s. We restrict attention to counties and districts within s, as out-of-state terms vanish in the derivative with respect to Q_s. We study the mapping $F : \mathcal{Q}_s \to [0,1]^{|C_s|}$ defined by $F_c(Q_s) = \bar{\pi}_c(Q_s)$, where the domain is the simplex-constrained set:

$$\mathcal{Q}_s = \left\{ Q_s \in [0,1]^{|C_s| \times |D_s|} : \sum_{d \in D_s} q_{(c,d)} = 1 \ \forall c \right\}$$

We analyze local identification by checking the invertibility of the Jacobian. Let $Q_s = Q_s^0 + \Delta Q_s$. By Assumption 1 (Limited Fragmentation), we treat Q_s^0 as a perturbation of a "clean" map where $q_{(c,d(c))} = 1$ for all c. By Assumption 2 (Simple Boundaries), any feasible perturbation ΔQ_s acts as a flow between at most two districts for each county. For a county c, let $d(c)$ be the dominant district and $d_{\text{alt}}(c)$ be the unique alternative district involved in the perturbation.

We can thus parameterize the perturbation by a single vector $v \in \mathbb{R}^{|C_s|}$, where v_c represents the mass moving from $d(c)$ to $d_{\text{alt}}(c)$.

$$\Delta q_{(c,d_{\text{alt}}(c))} = v_c, \quad \Delta q_{(c,d(c))} = -v_c, \quad \Delta q_{(c,k)} = 0 \text{ otherwise.}$$

A first-order Taylor expansion gives:

$$F(Q_s^0 + \Delta Q_s) - F(Q_s^0) = DF(Q_s^0)[\Delta Q_s] + R(\Delta Q_s)$$

By the Inverse Function Theorem, local injectivity is determined by the invertibility of the Jacobian $DF(Q_s^0)$. If the linear map $DF(Q_s^0)$ is invertible (has a trivial kernel), the higher-order terms $R(\Delta Q_s)$ do not affect the local uniqueness result.

Consider the limit as $\varepsilon_s \to 0$ (the clean map limit). Let A denote the Jacobian matrix $DF(Q_s^0)$ in this limit. The action of A on the perturbation vector v for county c simplifies to:

$$\begin{aligned}
(Av)_c &= \sum_{c'} \pi_{c,c'} \left(\Delta q_{(c',d(c))} + \Delta q_{(c,d(c'))}\right) \\
&= \sum_{c'} \pi_{c,c'} \Bigg[\underbrace{(\mathbb{I}(d(c') = d(c))(-v_{c'}) + \mathbb{I}(d(c) = d_{\text{alt}}(c'))(v_{c'}))}_{\text{Effect of } c' \text{ moving}} \\
&\qquad + \underbrace{(\mathbb{I}(d(c') = d(c))(-v_c) + \mathbb{I}(d(c') = d_{\text{alt}}(c))(v_c))}_{\text{Effect of } c \text{ moving}} \Bigg]
\end{aligned}$$

We can explicitly define the entries of the matrix A from the terms above. The diagonal elements A_{cc} capture the "self effect":

$$A_{cc} = \sum_k \pi_{c,k} \left(\mathbb{I}[d(k) = d_{\text{alt}}(c)] - \mathbb{I}[d(k) = d(c)]\right)$$

The off-diagonal elements $A_{cc'}$ (for $c \neq c'$) capture the effects from how friend counties move:

$$A_{cc'} = \pi_{c,c'} \left(\mathbb{I}[d(c) = d_{\text{alt}}(c')] - \mathbb{I}[d(c') = d(c)]\right)$$

To prove identification, we must show $\ker(A) = \{0\}$. We can decompose A into a diagonal shift and a signed interaction matrix: $A = D_{\text{net}} + (\Pi_s \circ S)$.

By Assumption 3 (Out-of-State Connections), we have $\sum_{c' \in C_s} \pi_{c,c'} < 1$, which implies the spectral radius $\rho(\Pi_s) < 1$. This condition ensures the system has adequate "leakage": the network feedback (off-diagonal terms) is structurally strictly weaker than the aggregate sensitivity to one's own perturbation (the diagonal terms A_{cc}, which sum over the entire row of Π_s).

By Assumption 4 (Network Non-Degeneracy), the rows of Π_s are linearly independent. The matrix A is effectively a projection of $(I - \Pi_s)$ weighted by the specific swap directions. Since $\rho(\Pi_s) < 1$, the matrix $(I - \Pi_s)$ is strictly invertible. The structure of A preserves this property because the leakage to out-of-state nodes prevents the existence of a perfect cycle that could sum to zero. Thus, $Av = 0 \implies v = 0$.

Finally, since the determinant $\det(DF(Q))$ is a continuous function of Q, and we have shown $\det(DF(Q^{\text{clean}})) \neq 0$, there exists an open neighborhood around the clean map where the Jacobian remains invertible. Thus, for sufficiently small ε_s, the map is locally identified. $\square$

B Data Descriptions

B.1 Construction of Vote Count Measures

B.1.1 County-by-congressional district measures

I construct the voting outcomes at the county-by-CD-level by using precinct-level vote count data from the Harvard Election Data Archive (for 2000-2010) and the MIT Election Data and Science Lab (for 2016-2020), combined with county-by-congressional-district vote count data from Dave Leip's Election Atlas (for House elections) and Daily Kos (for President, Senator, and Governor elections).

B.2 Variable Descriptions

Variable	**Description**
Heard of Representative	When shown the name of their current House representative and asked to indicate the party their representative is affiliated with, respondent did not indicate they had "Never Heard of Person" , and they instead chose "Republican", "Democrat", "Other Party/Independent", or "Not Sure". Binary. From pre-survey.
Selected Party	When shown the name of their current House representative and asked to indicate the party their representative is affiliated with, respondent did not indicate they had "Never Heard of Person" or "Not Sure", and they instead chose "Republican", "Democrat", or "Other Party/Independent". Binary. From pre-survey.
Selected Correct Party	When shown the name of their current House representative and asked to indicate the party their representative is affiliated with, respondent chose the correct party. Binary. From pre-survey.

Table B1: Descriptions for CES Voter Knowledge Outcomes

Variable	**Description**
Prefer Incumbent	When asked “In the general election for U.S. House of Representatives in your area, who do you prefer?”, respondent chose the name of their current House representative. Binary. From pre-survey. Missing if there is no incumbent running.
Prefer Opponent	When asked “In the general election for U.S. House of Representatives in your area, who do you prefer?”, respondent chose the name of someone other than their current House representative. Binary. From pre-survey. Missing if there is no incumbent running.
Prefer Neither	When asked “In the general election for U.S. House of Representatives in your area, who do you prefer?”, respondent did not choose the name of any candidate. Binary. From pre-survey. Missing if there is no incumbent running.
Voted for Incumbent	When asked “For whom did you vote for U.S. House?”, respondent chose the name of their current House representative. Binary. From post-survey. Missing if there is no incumbent running. Missing if both “Voted in General Election” variables are missing.
Voted for Opponent	When asked “For whom did you vote for U.S. House?”, respondent chose the name of someone other than their current House representative. Binary. From post-survey. Missing if there is no incumbent running. Missing if both “Voted in General Election” variables are missing.
Voted for Neither	When asked “For whom did you vote for U.S. House?”, respondent did not choose the name of any candidate. Binary. From post-survey. Missing if there is no incumbent running. Missing if both “Voted in General Election” variables are missing.
Voted in General Election (Validated)	Respondent can be linked to state voter rolls, and there is a record of the respondent voting in the general election. Binary. From post-survey.
Voted in Primary Election (Validated)	Respondent can be linked to state voter rolls, and there is a record of the respondent voting in the primary election. Binary. From post-survey.
Voted in General Election (Self-Report)	Respondent answered that they voted in the general election. Binary. From post-survey.

Table B2: Descriptions for CES Voting Outcomes

Variable	Observations	Mean (%)	SD (pp)
Heard of Representative	545,185	93.2	25.2
Selected Party	604,254	68.6	46.4
Selected Correct Party	604,254	61.7	48.6
Prefer Incumbent	419,545	40.14	49.0
Prefer Opponent	419,545	26.7	44.3
Prefer Neither	419,545	33.1	47.1
Voted for Incumbent	385,212	41.0	49.2
Voted for Opponent	385,212	29.1	45.4
Voted for Neither	385,212	29.9	45.8
Voted in General Election (Validated)	417,421	57.5	49.4
Voted in Primary Election (Validated)	381,277	31.8	46.6
Voted in General Election (Self-Report)	388,262	87.8	32.8

Table B3: CES Data: Summary Statistics

Variable	Description
House Turnout, Relative to Top-of-Ticket ("Roll-Off")	$\frac{\text{\# Votes in Top-of-Ticket Race} - \text{\# Votes in House Race}}{\text{\# Votes in Top-of-Ticket Race}}$ For main analysis, from Dave Leip's Election Atlas (county-level). For robustness, from Harvard Election Data Archive, Daily Kos, Dave Leip's Election Atlas, and MIT Election Data and Science Lab (for county-by-congressional district-level). Elections where there is no top-of-ticket race are excluded.
Turnout in Top-of-Ticket Election	Turnout in the top-of-ticket election, as a share of the Voting Age Population (VAP), i.e., the population 18 years old or older. Vote counts from Dave Leip's Election Atlas, VAP from Census. Elections where turnout exceeds the VAP are excluded; identical to House turnout when the House election is top-of-ticket.
Turnout in House Election	Turnout in the House election, as a share of the Voting Age Population (VAP). Vote counts from Dave Leip's Election Atlas, VAP from Census. Elections where turnout exceeds the VAP are excluded.

Table B4: Descriptions for Voting Outcome Variables

Variable	Observations	Mean (%)	SD (pp)
Roll-Off	29,133	4.42	12.22
Turnout in Top-of-Ticket Election	30,206	51.34	13.44
Turnout in House Election	30,308	49.07	13.49

Table B5: Voting Outcomes: Summary Statistics

C Additional Empirical Results

C.1 Correlates of District Homophily

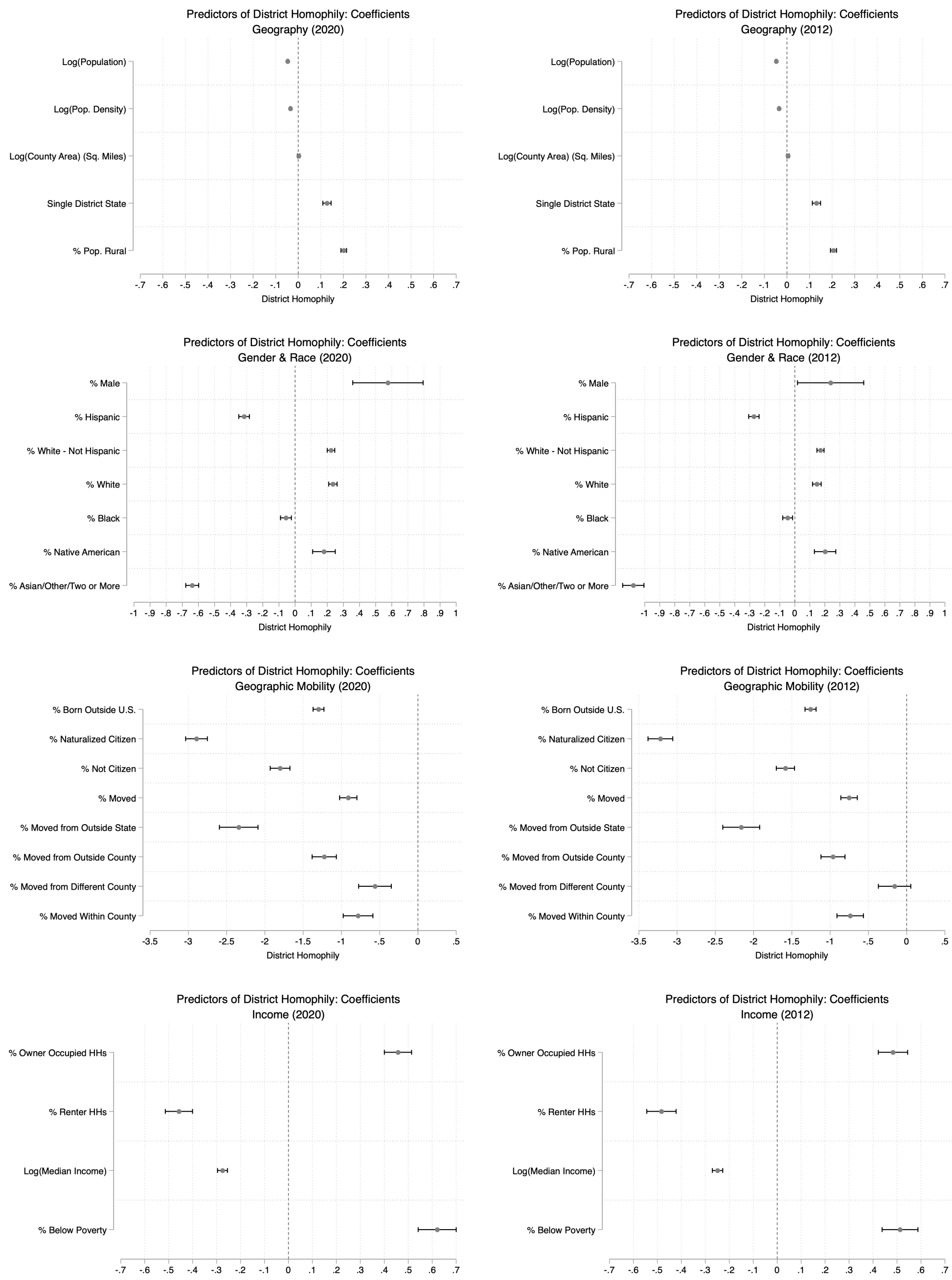


Figure C1: Correlates of District Homophily

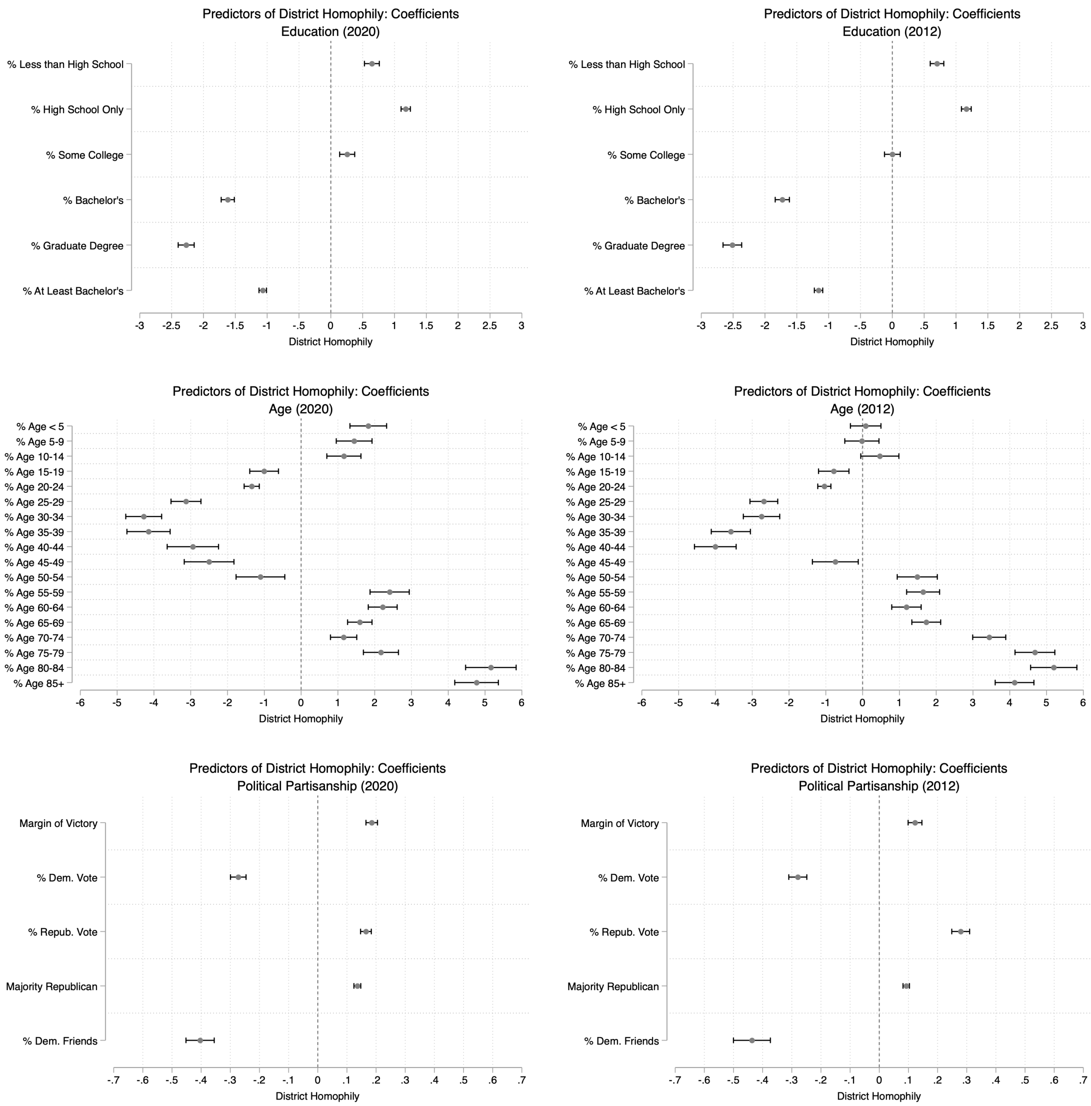


Figure C2: Correlates of District Homophily (cont.)

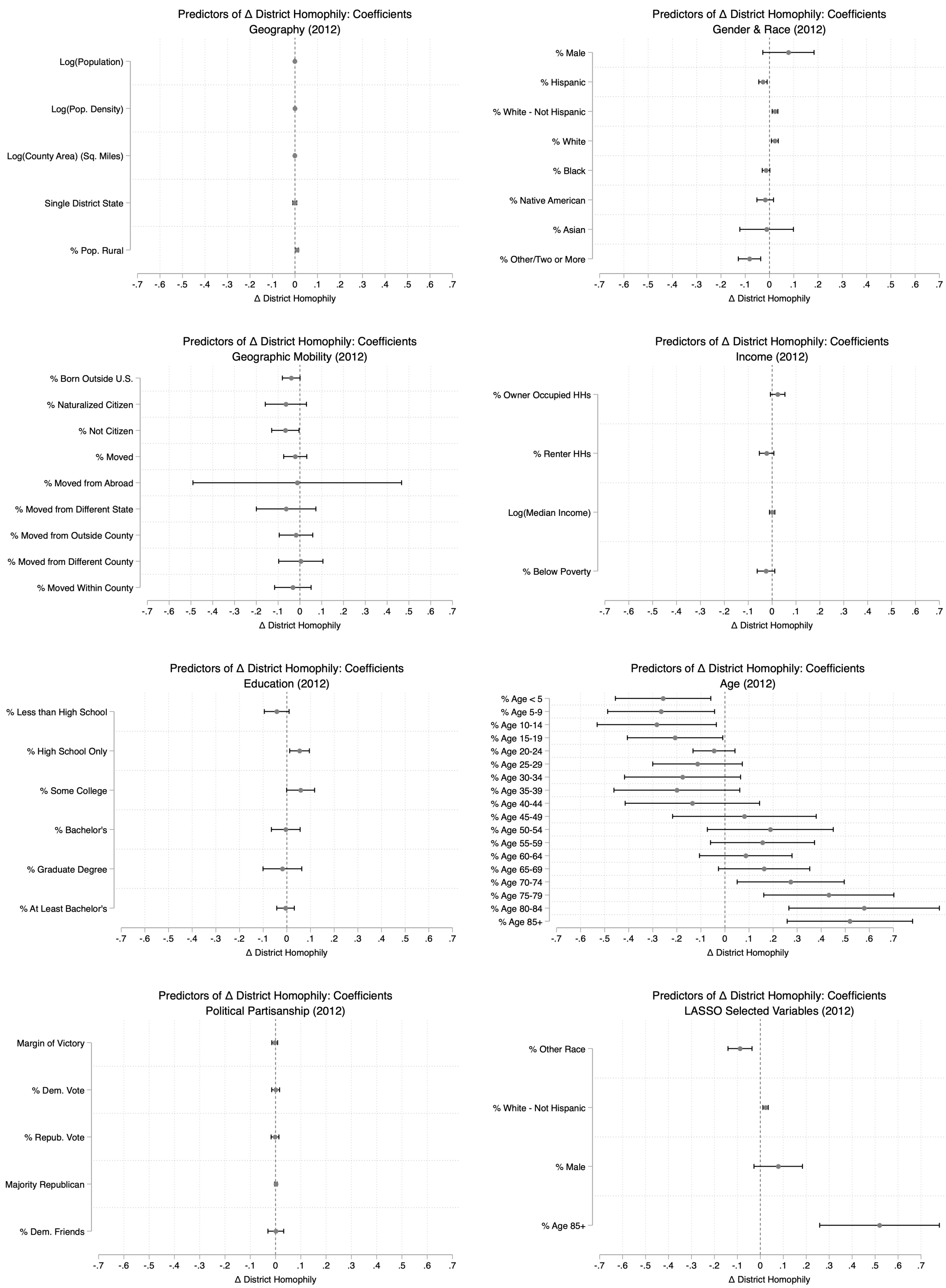


Figure C3: Correlates of Changes in District Homophily

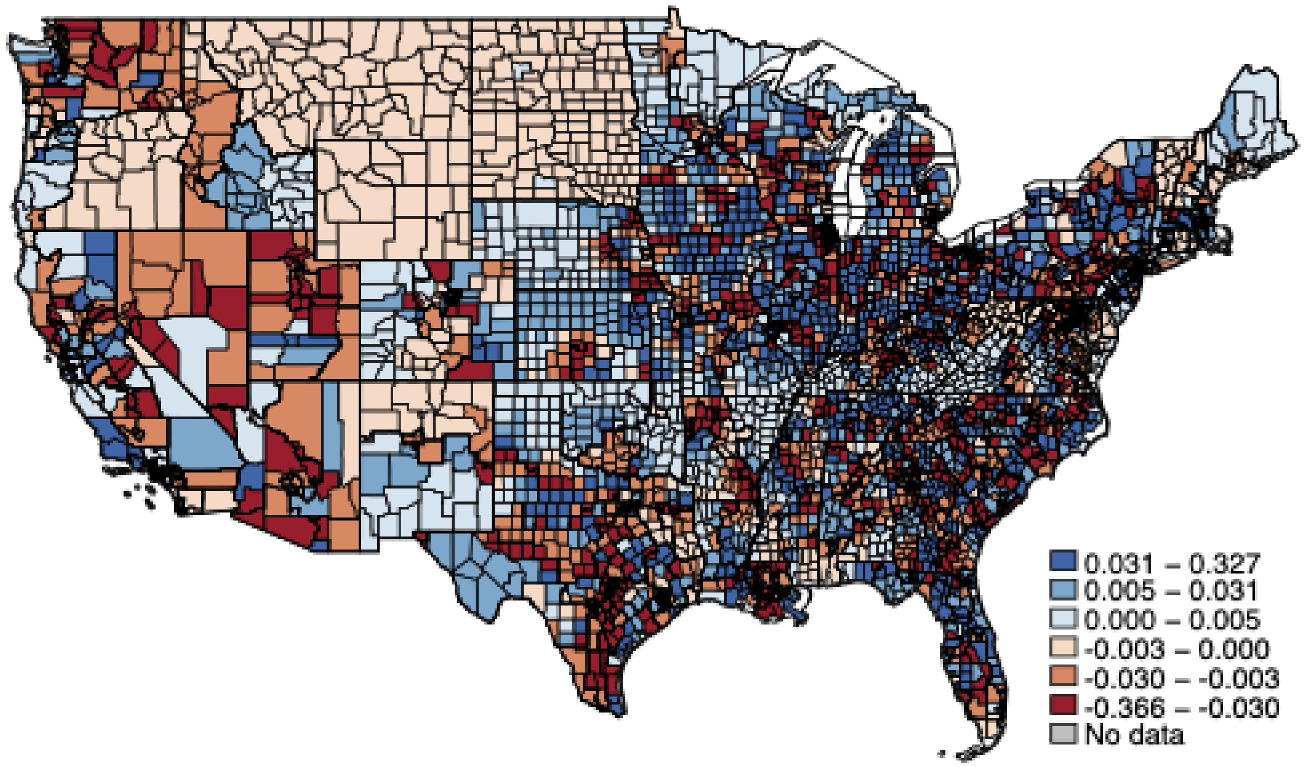


Figure C4: Changes in District Homophily: Equal Number of Counties Per Bin

The maps in Figures C4-C5 illustrate how the changes in district homophily used as identifying variation (that is, occurring following the 2012 redistricting) are distributed around the country.

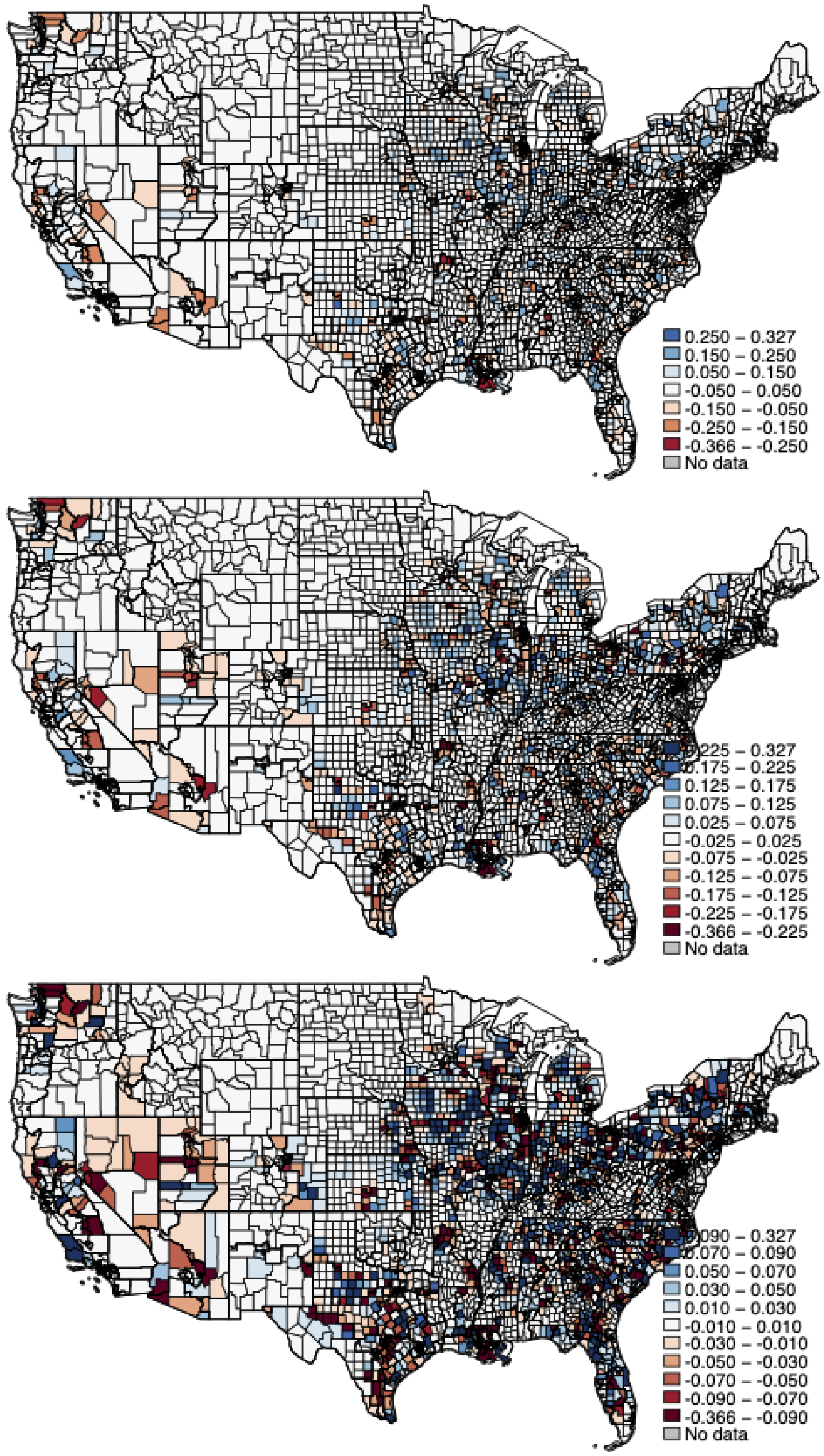


Figure C5: Changes in District Homophily: Equally Spaced Bins (Bins get progressively finer, zooming in on smaller variations.)

C.2 Placebo Outcomes for Voter Information

Table C1 provides summary statistics for the nine outcome variables used for placebo tests. These distributions are generally similar to those for House representatives (Table B3), though respondents are generally more likely to recognize and select the correct party for their Senators and Governor. The nine figures that follow show the results of the placebo tests. In general, district homophily does not significantly predict the placebo outcomes.

Variable	**Observations**	**Mean (%)**	**SD (pp)**
Heard of Governor	549,740	96.3	18.8
Selected Governor Party	608,985	81.3	39.0
Selected Correct Gov. Party	608,985	74.7	43.5
Heard of Senator 1	549,244	94.9	22.0
Selected Senator 1 Party	608,414	75.0	43.3
Selected Correct Sen. 1 Party	608,414	67.7	46.7
Heard of Senator 2	549,246	94.9	22.0
Selected Senator 2 Party	608,402	74.3	43.7
Selected Correct Sen. 2 Party	608,402	66.9	47.0

Table C1: CES Data: Summary Statistics for Placebo Outcomes

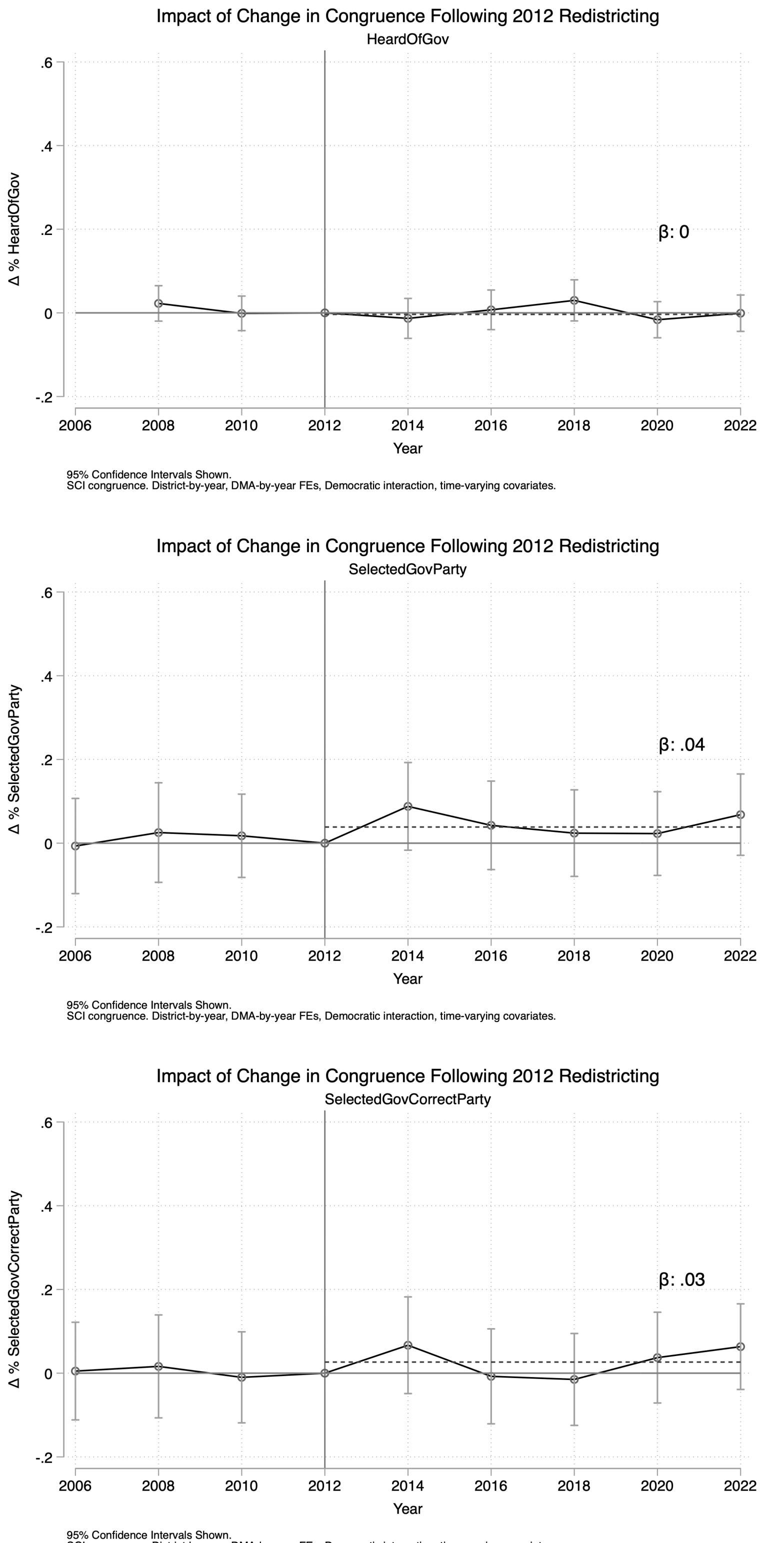


Figure C6: Effect of District Homophily on Knowledge of Governor

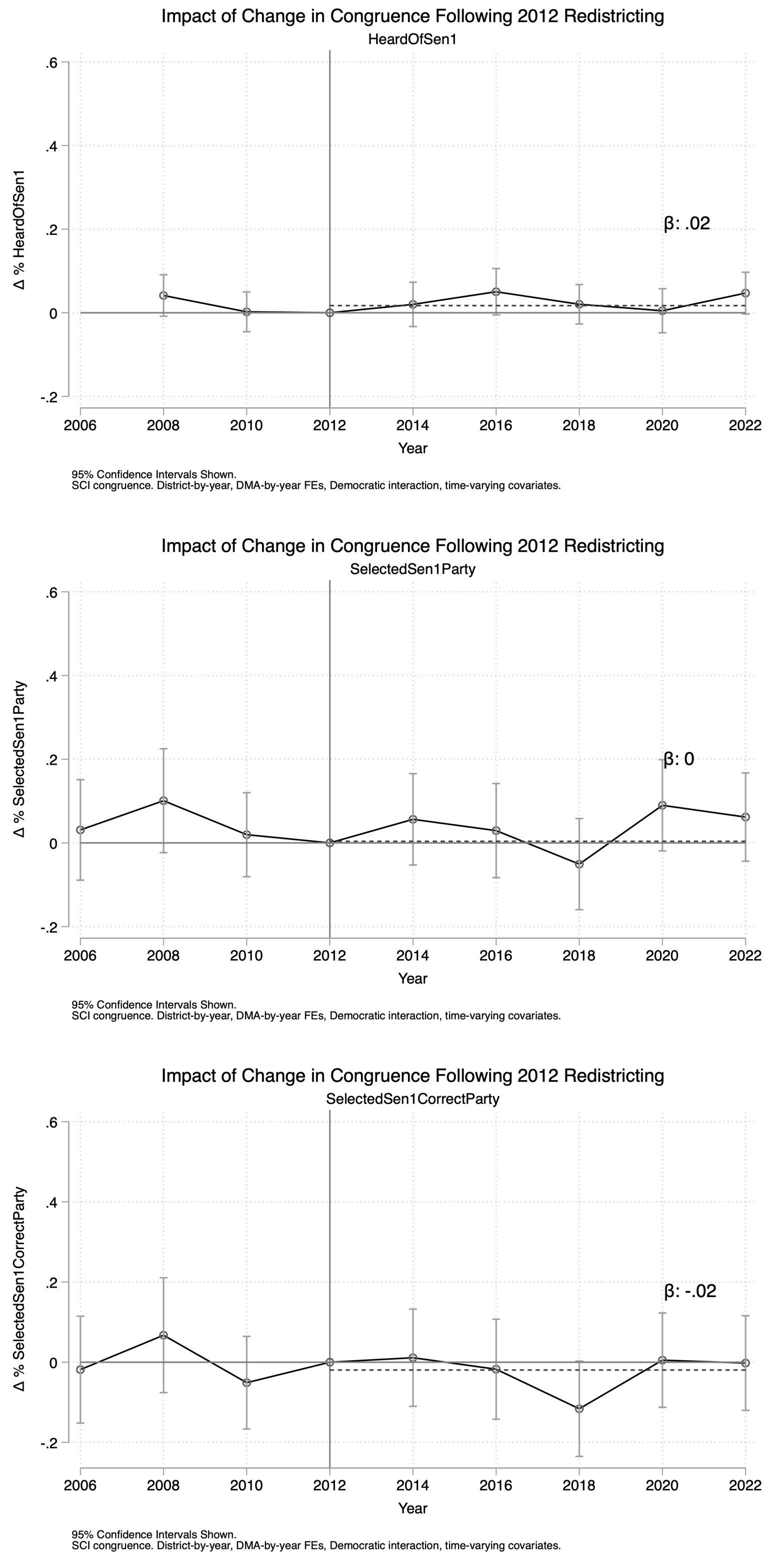


Figure C7: Effect of District Homophily on Knowledge of Senator 1

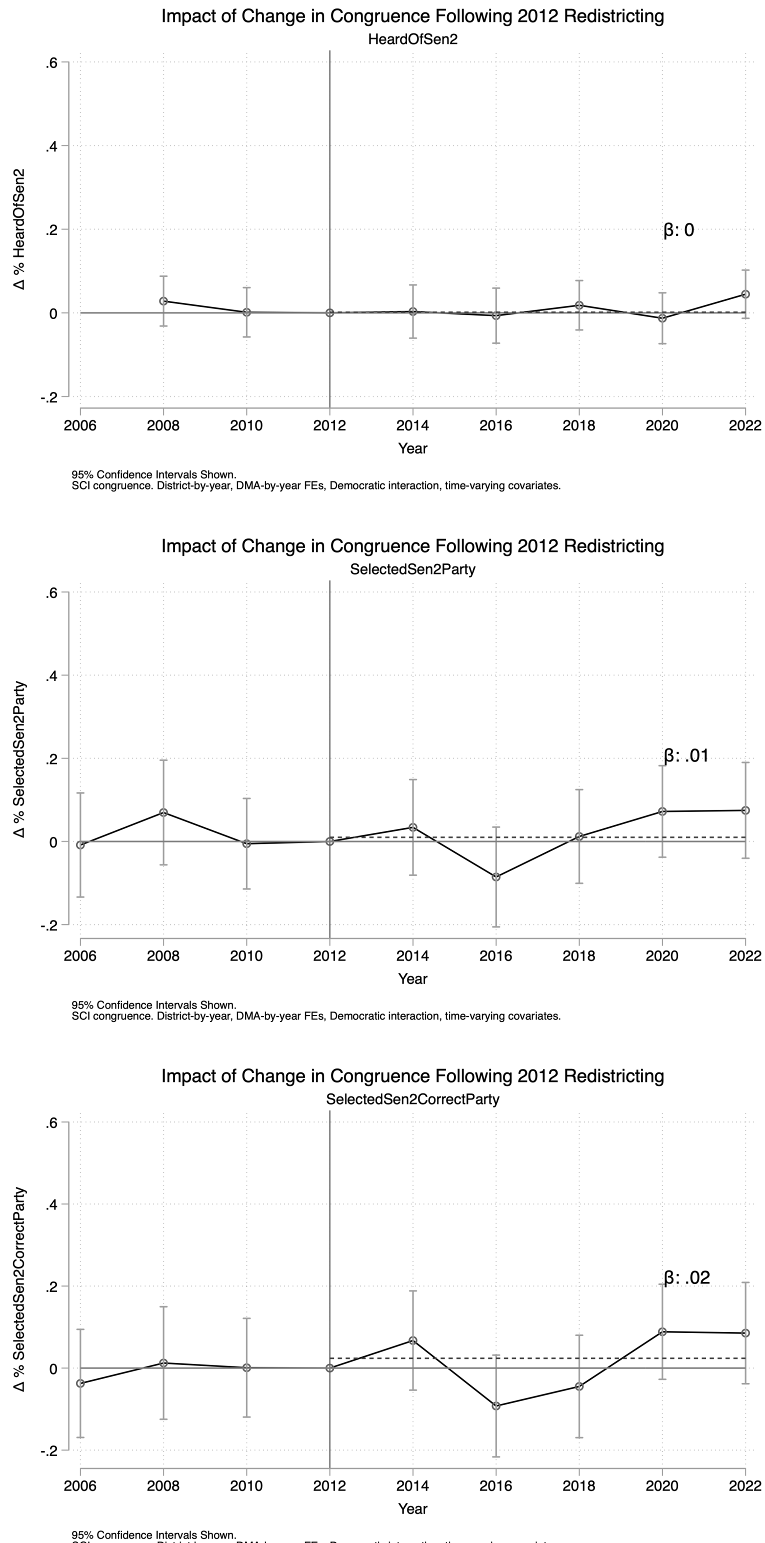


Figure C8: Effect of District Homophily on Knowledge of Senator 2

C.3 Commuting Flows as a Proxy of Social Networks

The following figures present the effect of district homophily on the main outcomes, when district homophily is constructed using commuting flows, rather than the SCI.

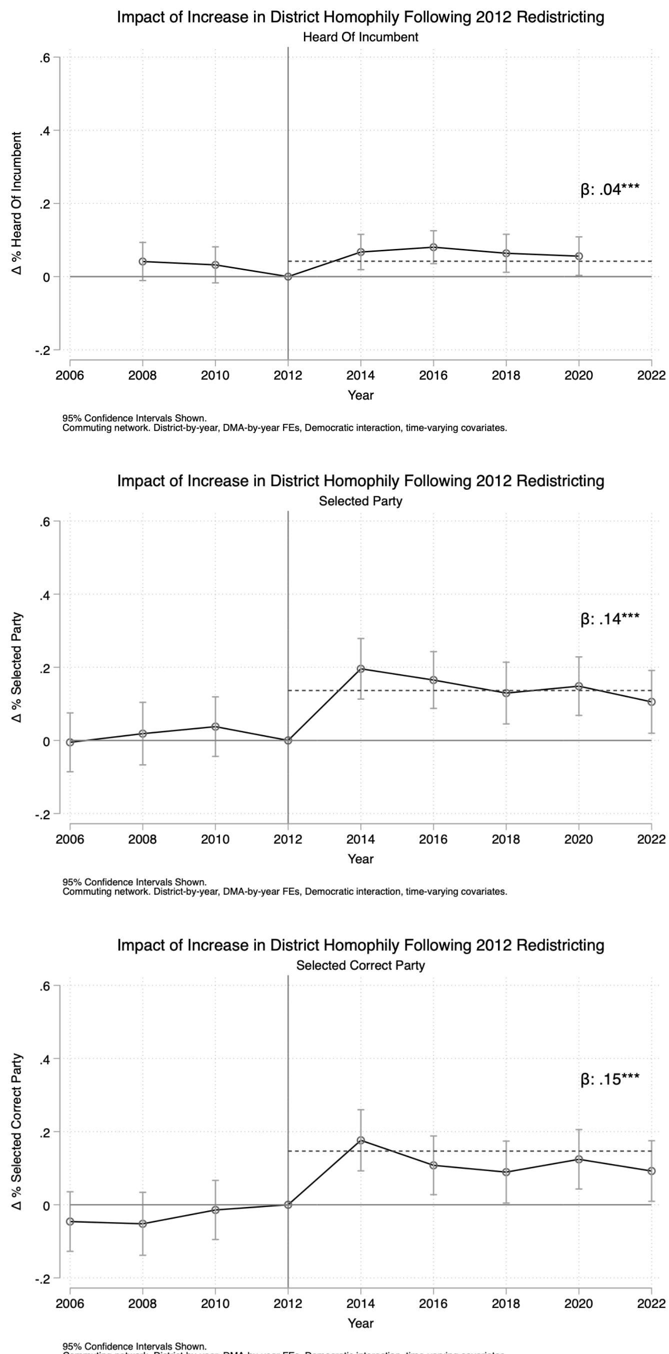


Figure C9: Effect of Commuting District Homophily on Voters' Familiarity with Representatives

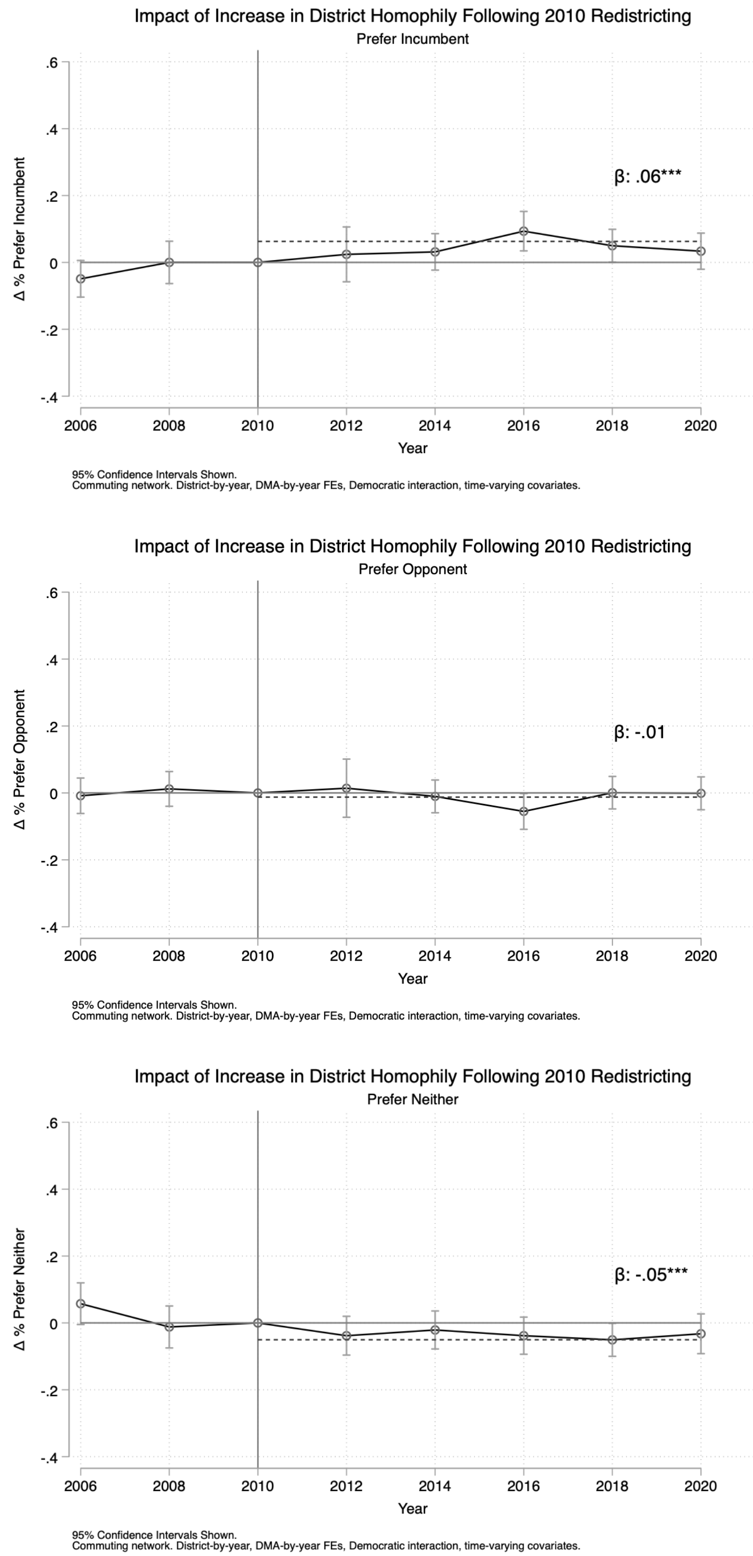


Figure C10: Effect of Commuting District Homophily on Voters' Preferences Among House Candidates

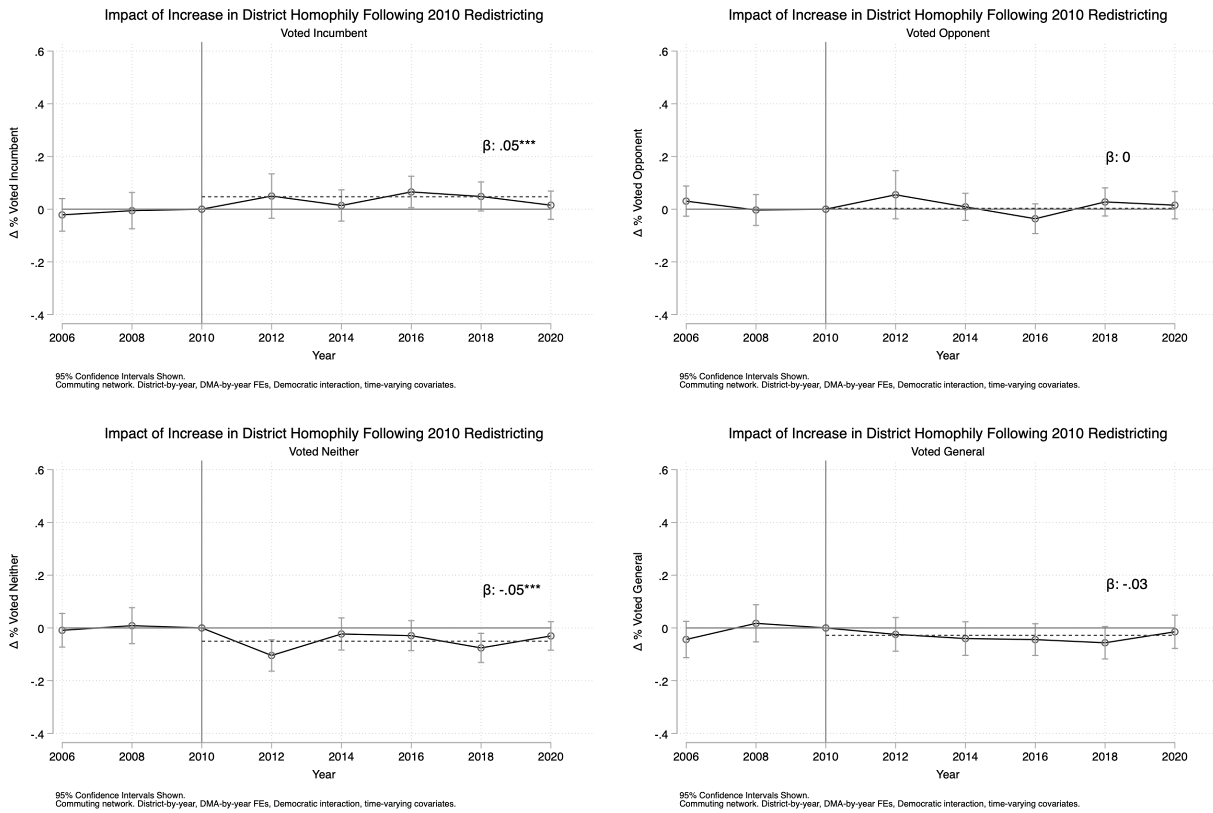


Figure C11: Effect of Commuting District Homophily on Voters' Self-Reported Voting Outcomes (Post-Election Survey)

C.4 Zip Code District Homophily

The following figures present the effect of district homophily on the main outcomes, when district homophily is constructed using zip-code-level social networks.

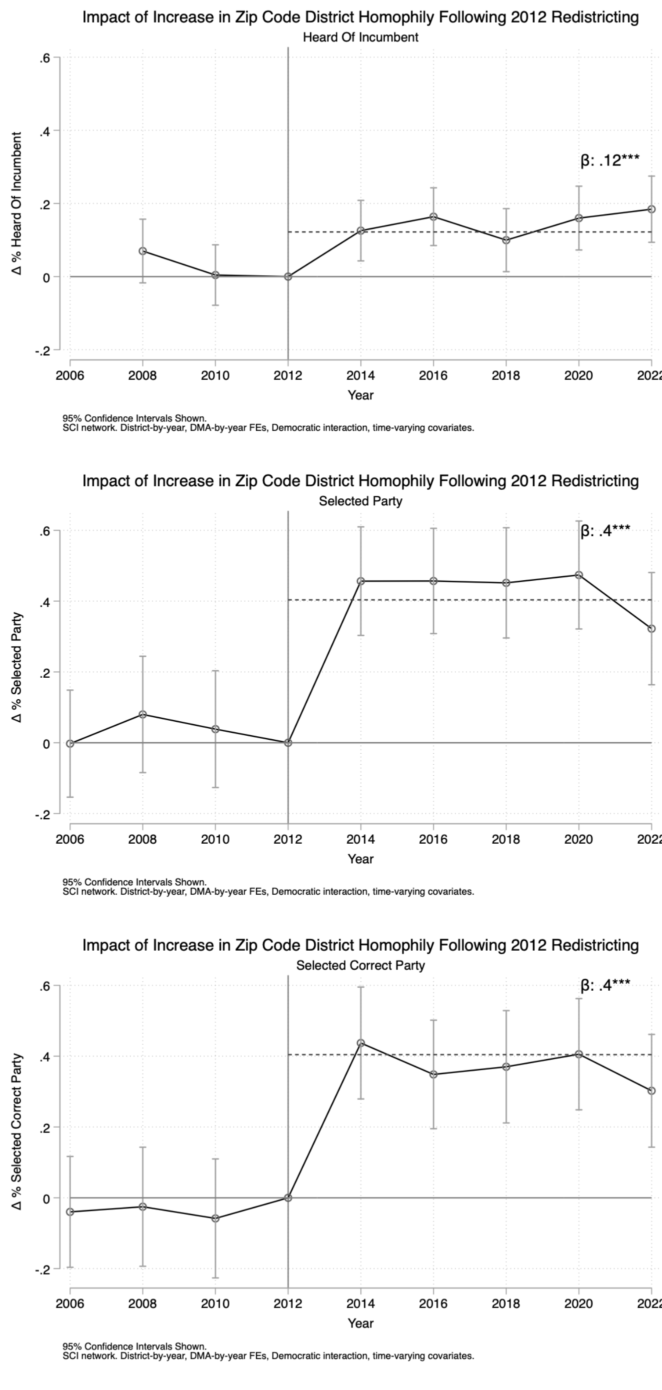


Figure C12: Effect of Zip Code District Homophily on Voters' Familiarity with Representatives

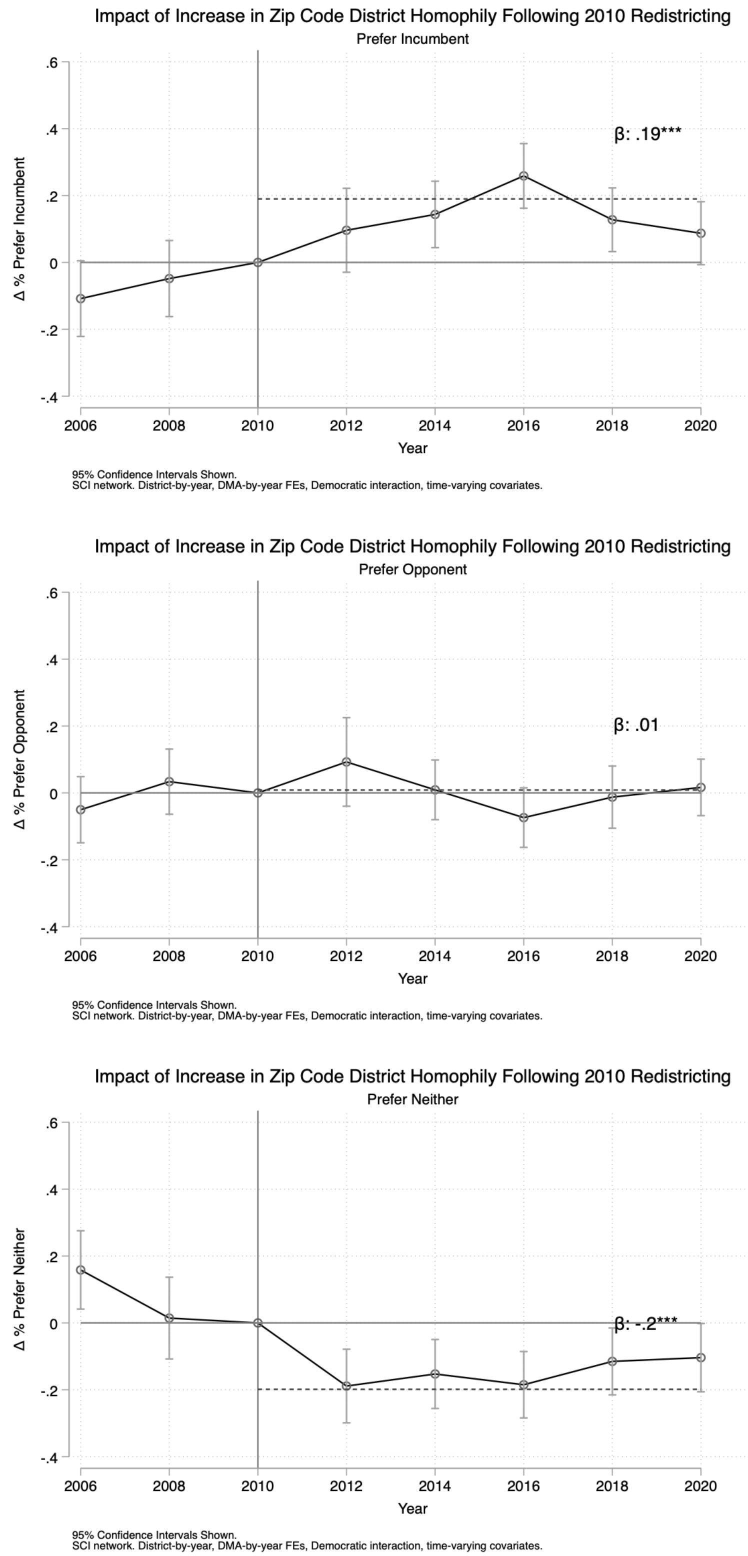


Figure C13: Effect of Zip Code District Homophily on Voters' Preferences Among House Candidates

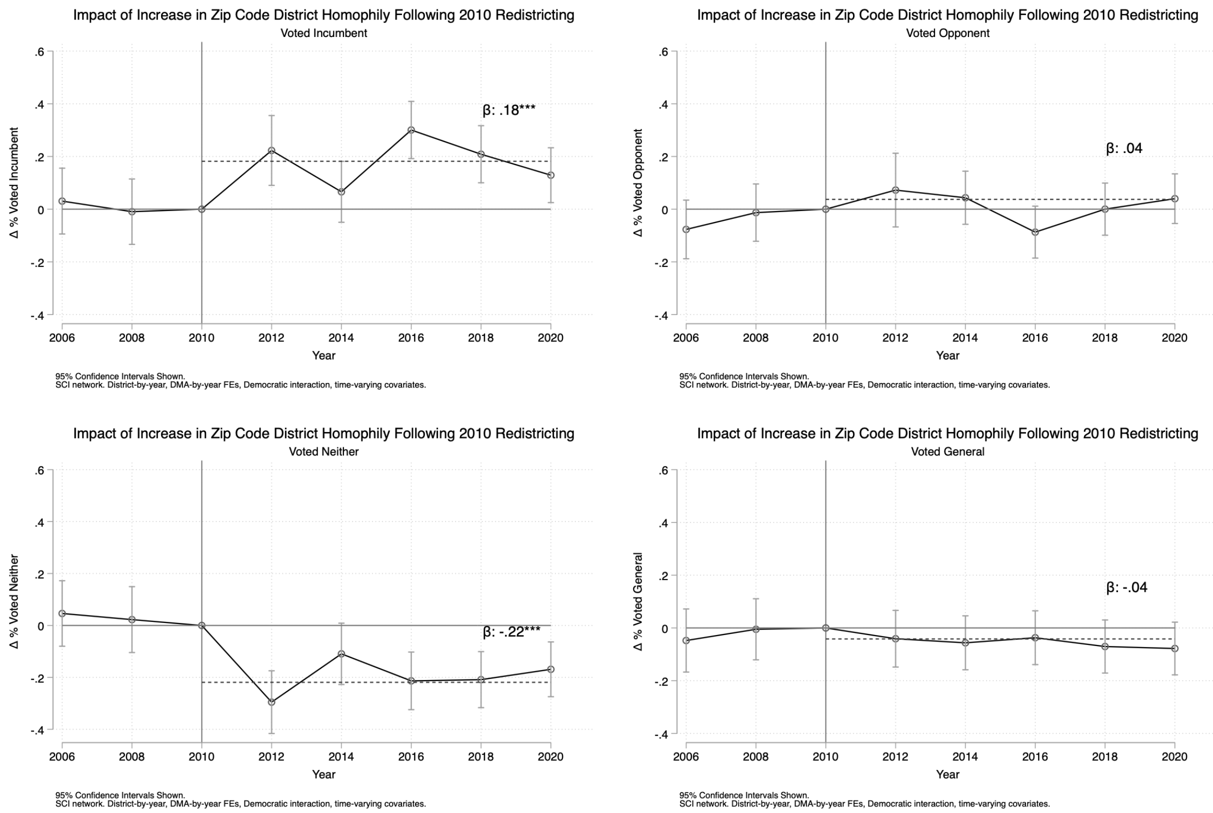


Figure C14: Effect of Zip Code District Homophily on Voters' Self-Reported Voting Outcomes (Post-Election Survey)

C.5 Border Pairs Design

Table C2 presents the effects of district homophily when variation comes from looking within pairs of counties that lie on opposite sides of a district border.

	(1) County & Pair x Year FEs Only	(2) Add State x Year FEs	(3) Add DMA x Year FEs	(4) Add Dem. Exposure	(5) Add Individual Demographic Controls	(6) Add County-Year Controls
			Heard of Incumbent			
District Homophily	0.254***	0.254***	0.296***	0.294***	0.253***	0.256***
	(0.080)	(0.080)	(0.091)	(0.091)	(0.093)	(0.091)
	[0.002]	[0.002]	[0.001]	[0.001]	[0.007]	[0.005]
Obs	22,094	22,094	21,508	21,508	21,508	21,508
R^2	0.606	0.606	0.699	0.699	0.718	0.720
			Selected Party			
District Homophily	0.368***	0.368***	0.214	0.217	0.162	0.136
	(0.141)	(0.141)	(0.155)	(0.156)	(0.151)	(0.151)
	[0.009]	[0.009]	[0.168]	[0.165]	[0.284]	[0.367]
Obs	25,798	25,798	25,126	25,126	25,126	25,126
R^2	0.620	0.620	0.709	0.709	0.740	0.742
			Selected Correct Party			
District Homophily	0.608***	0.608***	0.399**	0.406**	0.346**	0.343**
	(0.161)	(0.161)	(0.176)	(0.176)	(0.165)	(0.160)
	[0.000]	[0.000]	[0.023]	[0.022]	[0.036]	[0.032]
Obs	25,798	25,798	25,126	25,126	25,126	25,126
R^2	0.631	0.631	0.716	0.717	0.749	0.752
Dem. Exposure				X	X	X
Ind. Controls					X	X
County x Year Controls						X
FEs	County, Pair x Year	County, Pair x Year, State x Year	County, Pair x Year, State x Year, DMA x Year	County, Pair x Year, State x Year, DMA x Year	County, Pair x Year, State x Year, DMA x Year	County, Pair x Year, State x Year, DMA x Year

*** p<0.01, **p<0.05, * p<0.1
Standard errors clustered at the county level in parentheses. P-values in square brackets.
"Heard of Incumbent" not available in 2006, 2007, or 2009. Individual controls include gender, race, education, age categories, and whether the respondent is affiliated with the same party as their representative. County-by-year controls include population and shares by race, age categories, gender, and county urban population share.

Table C2: Effect of District Homophily on Voter Familiarity with Representative, within Border Pairs

C.6 CES Survey Responses

In the post-survey, respondents are asked who they voted for. Based on these self-reports, I find that district homophily increases the share of general election voters reporting voting for the incumbent, without affecting the share who vote for the challenger. Instead, as described in the main text, the share of general election voters who report skipping the House election decreases.

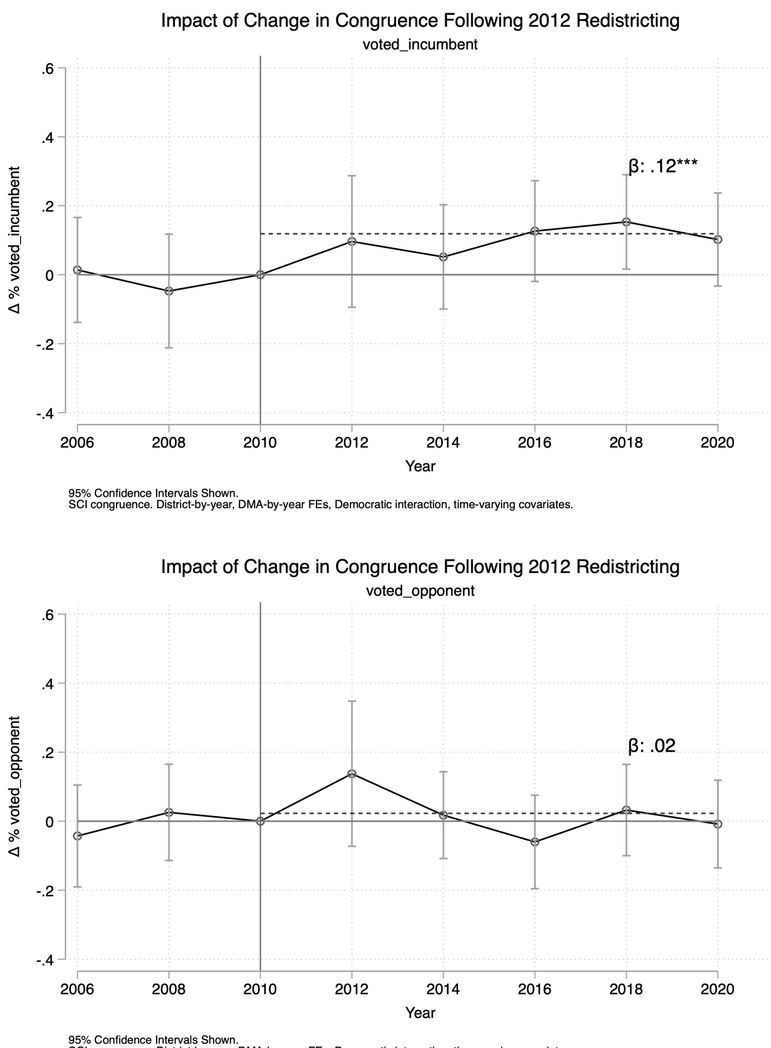


Figure C15: Effect of District Homophily on Voter Choices: Incumbent vs Challenger

C.7 Vote Count Data

Using county-level vote count data, I look at the impact of a change in district homophily on actual turnout in the top-of-ticket (i.e., President, Senate, or Governor) election and turnout in the House election. I measure this as total votes in the respective election divided by the share of the county's population that is age 18 or older (i.e., the voting age population). While the voting age population does not account for those who are ineligible to vote (e.g., non-citizens or people with a felony record in some states), I do control for the share of the population that are non-citizens in the set of county-by-year covariates. The sample includes only counties that have at least 50% of population in one district, and district-by-year fixed effects reflect that district. (Results are qualitatively similar when restricting to counties 100% in a single district.)

Here we find no impacts of district homophily on turnout in the top-of-ticket race and in the House race. However, this does not necessarily contradict the finding of district homophily decreasing roll-off: When I construct roll-off as $\frac{\text{Votes in Top-of-Ticket Election} - \text{Votes in House Election}}{\text{Voting Age Population}}$, I find a significant 2pp decrease in roll-off. This is because the estimates of House turnout are too imprecise to detect an impact of this scale, and also have a slight negative pre-trend; in the roll-off estimates, controlling for the top-of-ticket election both reduces the noise in the estimates and eliminates much of the pre-trend.

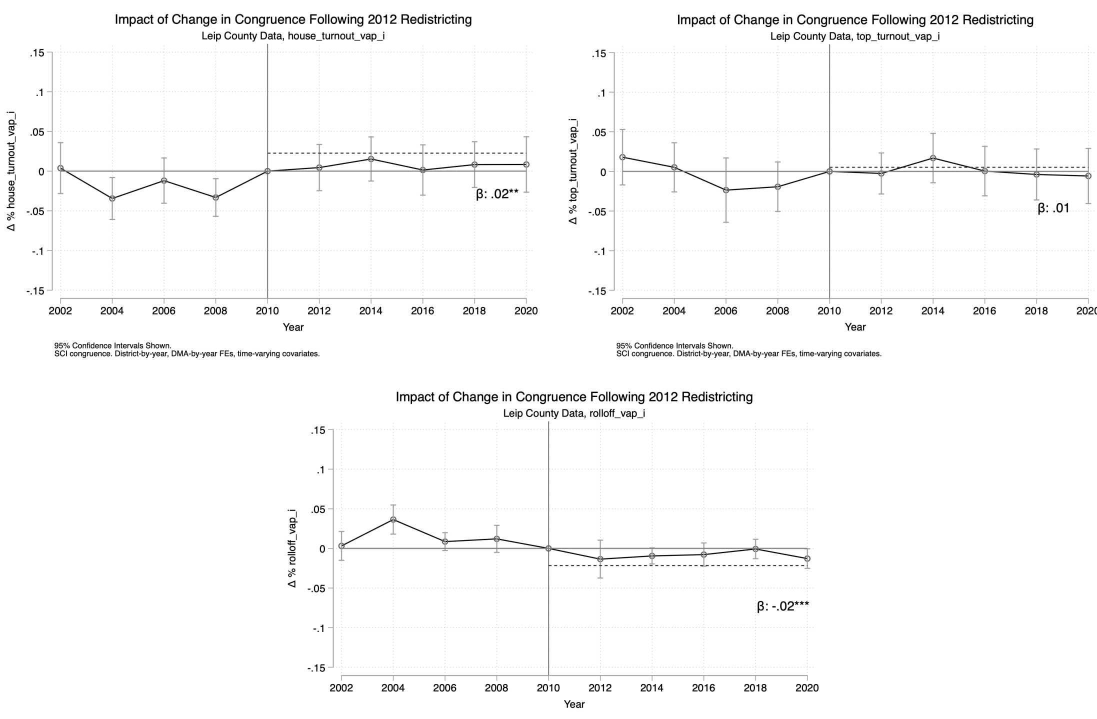


Figure C16: Effect of District Homophily on Turnout

C.8 Campaign Contributions

District Homophily has no impact on total contributions to House candidates; instead, it increases contributions to in-district candidates at the cost of contributions to out-of-district candidates. I also find similar results when restricting to only primary elections; only general elections; or excluding large donations (i.e., excluding Census tracts where the average donation per contributor exceeds $1,000).

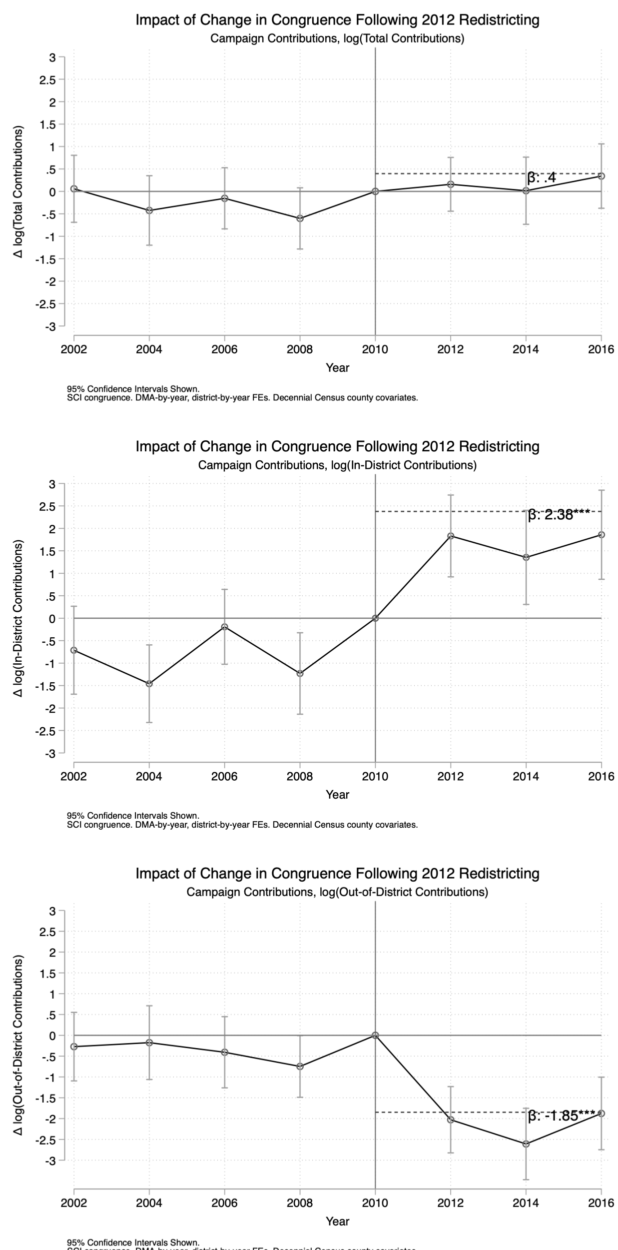
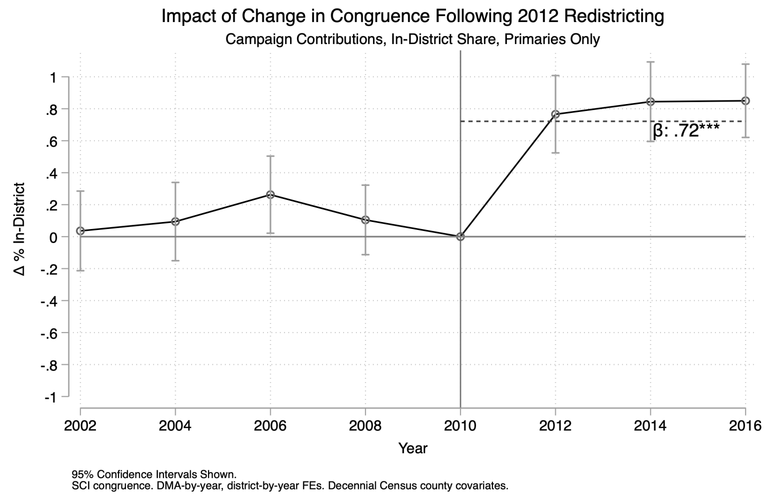
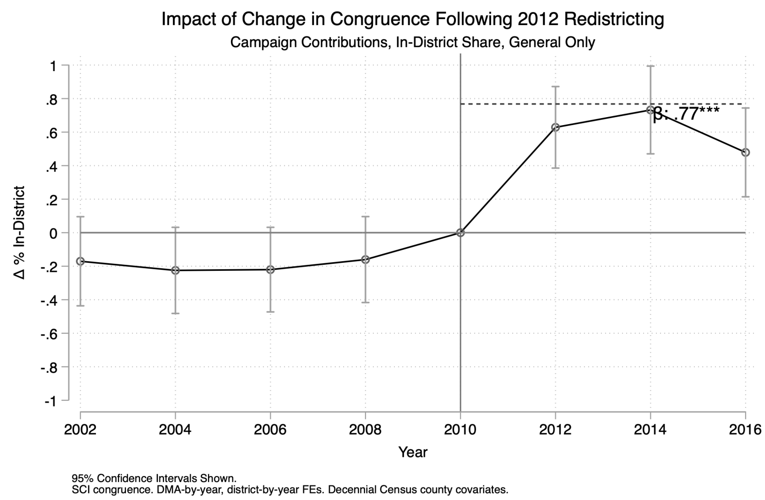


Figure C17: Effect of District Homophily on Log of Campaign Contributions





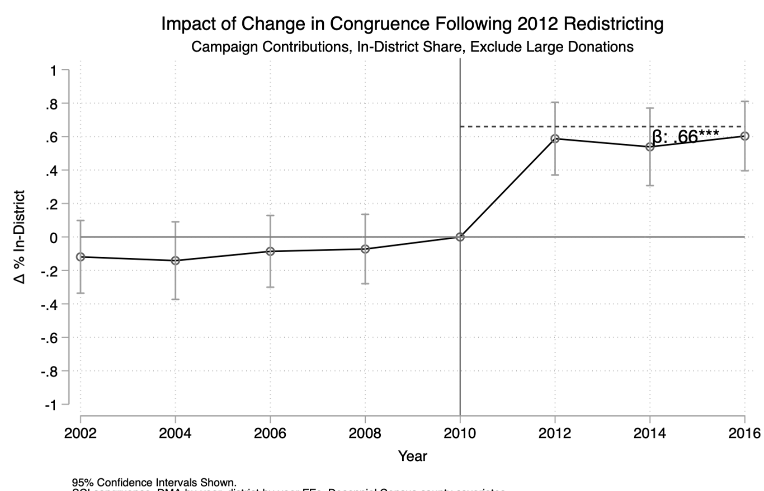


Figure C18: Effect of District Homophily on Share of Campaign Contributions to In-District Candidates

C.9 Distribution of District Homophily Across Simulated Maps

This section explores the features of the distribution of district homophily across the simulated maps.

C.9.1 Range of Average State-wide District Homophily

My event studies demonstrate that district homophily positively impacts voters' familiarity with their representatives and decreases the probability voters abstain in House elections. As such, if increased voter knowledge and participation are policy goals, when considering implications for policy, an immediate question is whether—and where—it is feasible to draw maps that increase district homophily state-wide, or if district homophily is largely zero sum such that any increase for one part of the state leads to a decrease in another part of the state.

To study this, for each state, I measure the range of average state-wide district homophily across the simulated maps. Specifically, under each of a state's 5,000 simulated maps, I calculate the population-weighted average district homophily of the state (using district homophily constructed at the county level). I then take the difference between the maximum district homophily map and the minimum district homophily map. Figure C19 plots this range for each of the contiguous 48 states. (The five single-district states among the 48 have no simulated maps, as there are no alternative ways to draw the district, and are shown in gray.)

Again, Figure C19 does not plot the range of district homophily across *all* possible maps, but rather across the distribution of simulated maps. Thus, bigger differences are certainly possible, but the figure conveys which states have a political geography and social network structure that leave the map drawer with more leverage to affect overall district homophily.

Figure C19 reveals that there is variation across states in terms of this range. The median state is Indiana, with a range of 4pp (the mean is 4.01pp). Oregon has a range just below Indiana's, and Michigan's is just above. Louisiana has the maximum range of 7.09pp, while Nevada has the minimum range of 0.48pp. The remaining of the top five states, with ranges above 6pp, are Maine, South Carolina, Illinois, Utah, and West Virginia. Meanwhile, the remaining states at the bottom of the distribution, with ranges of about 2pp or less, are California, Texas, Florida, Kansas, and Arizona.

The map reveals that there is no clear correlation between the range of district homophily and factors like partisan-leaning. Each state's idiosyncratic features—its share of out of state friends, the number of districts the state must be split into (which scales with population), the state's redistricting laws—interact to shape the range of statewide-average district homophily.

For the median state, applying the event study results directly, going from the minimal to maximal

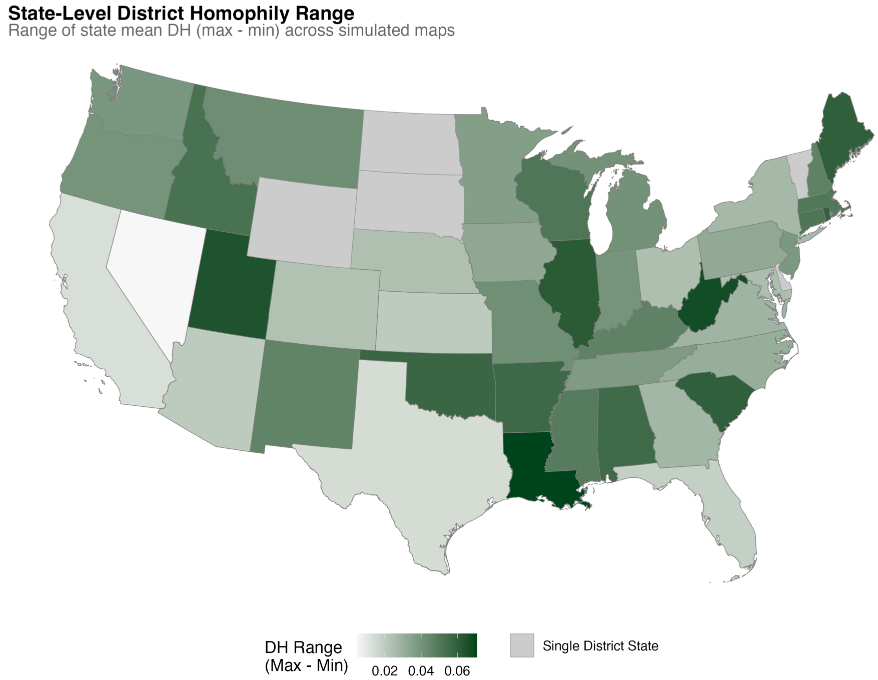


Figure C19: Range of State-Level District Homophily across Simulated Maps

simulated map translates into a 1.3pp increase in the overall share of informed voters (defining "informed" as selecting the correct party). For Louisiana, the possible increase is 2.3pp. On the one hand, the figure reveals that district homophily is not zero sum: there are maps with higher averages and maps with lower averages. On the other hand, the scope for possible outcomes is clearly relatively constrained: for example, a policy that attempts to raise the national average share of informed voters via drawing maps would only be able to increase the average by a few percentage points, at least while obeying other constraints on redistricting, such as existing state redistricting laws.

C.10 Correlates of District Homophily

As the preceding figures reveal, the distribution of district homophily within the simulated maps depends on a variety of factors unique to each state's social networks and political geography. Consequently, within the simulated maps, regressions of gerrymandering measures (partisan bias, efficiency gap, and the dissimilarity index between Democrats and Republicans) on district homophily (including state fixed effects) reveal no correlation. Nonetheless, within a given state district homophily can correlate with different measures of gerrymandering and bias, in ways that reflect the state's population distribution.

As an example, I simulate district homophily across all 5,000 Texas maps in the McCartan, Kenny,

Outcome	$\bar{\pi}_{\mathbf{c}}$ **(Unweighted)**	$\bar{\pi}_{\mathbf{c}}$ **(Weighted)**
Compactness (Edge)	0.093***	0.132***
Compactness (Polsby-Popper)	0.497***	-0.013
County Splits	-38.841***	-82.224***
Municipal Splits	-62.116***	-395.18***
Efficiency Gap	0.497***	0.290*
Partisan Bias	0.389***	1.237***
E(Dem. Seats)	-27.313***	-3.457
Dissimilarity Index (Democrat vs Republican)	-0.075***	0.665***
Dissimilarity Index (Black vs Other)	-0.036	0.047
Dissimilarity Index (Hispanic vs Other)	-0.049	1.076***

Table C3: District Homophily vs. Measures of Gerrymandering in Texas

Simko, Kuriwaki, et al. 2021 dataset. Table C3 reports the correlation between district homophily and several common measures of gerrymandering across these maps. I provide the correlation with the average district homophily under the map, both unweighted (treating every county equally) and weighted by the population of the county. Some of these correlations are clearly unique to the geography and culture of Texas, such as higher district homophily being associated with a stronger Republican bias (as indicated by the positive coefficients for partisan bias and the efficiency gap, and negative coefficients for the expected number of Democratic seats). Meanwhile, other features may be more common across states, such as district homophily being positively correlated with increased compactness and fewer county or municipal splits.

While I discuss the demographic correlates of district homophily under the enacted maps in Section 2.1.5, by using “real world” maps these correlations necessarily include the consequences of any biases in the map drawing process. As such, I can look at the correlation between district homophily and demographics within the simulated maps, in order to understand the structural patterns that persist even under neutrally drawn maps. In Tables C4 and C5, I regress the mean district homophily of the simulated maps at the state level and the county level, respectively, on several demographic shares. At both the state and county level, the mean district homophily of the simulated maps is significantly negatively correlated with the share who are not White Non-Hispanic, the share Democrat, and the share urban; at the state level, it is also significantly negatively correlated with the share Hispanic, and at the county level, it is marginally significantly negatively correlated with the share Black. These patterns are similar to those reported in Section 2.1.5, and consistent with the observation that urban areas will have lower district homophily because of their diffuse social networks, with many connections to other urban areas.

Dependent Variable:	State Mean District Homophily				
	(1)	(2)	(3)	(4)	(5)
Constant	0.2935***	0.3771***	0.4305***	0.6787***	0.6213***
	(0.0258)	(0.0286)	(0.0342)	(0.0556)	(0.0589)
% Black	0.1023				
	(0.1937)				
% Hispanic		-0.6313**			
		(0.2665)			
1 - % White			-0.4036***		
			(0.1141)		
% Dem.				-0.7791***	
				(0.1080)	
% Urban					-0.0651***
					(0.0106)
Observations	41	41	41	41	41
R^2	0.00844	0.33838	0.26440	0.50033	0.40515
Adjusted R^2	-0.01699	0.32141	0.24554	0.48752	0.38989

Clustered (state) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table C4: Mean State-wide Average District Homophily of Simulated Maps vs. State Demographic Shares

Dependent Variable:	County Mean District Homophily				
	(1)	(2)	(3)	(4)	(5)
% Black	-0.0953*				
	(0.0549)				
% Hispanic		-0.1787			
		(0.1536)			
1 - % White			-0.1864***		
			(0.0655)		
% Dem.				-0.3265***	
				(0.0514)	
% Urban					-0.0458***
					(0.0040)
State FEs	Yes	Yes	Yes	Yes	Yes
Observations	2,562	2,562	2,562	2,562	2,562
R^2	0.28488	0.29805	0.32318	0.39234	0.49760
Within R^2	0.00729	0.02558	0.06046	0.15647	0.30258

Clustered (state) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table C5: Mean County District Homophily of Simulated Maps vs. County Demographic Shares

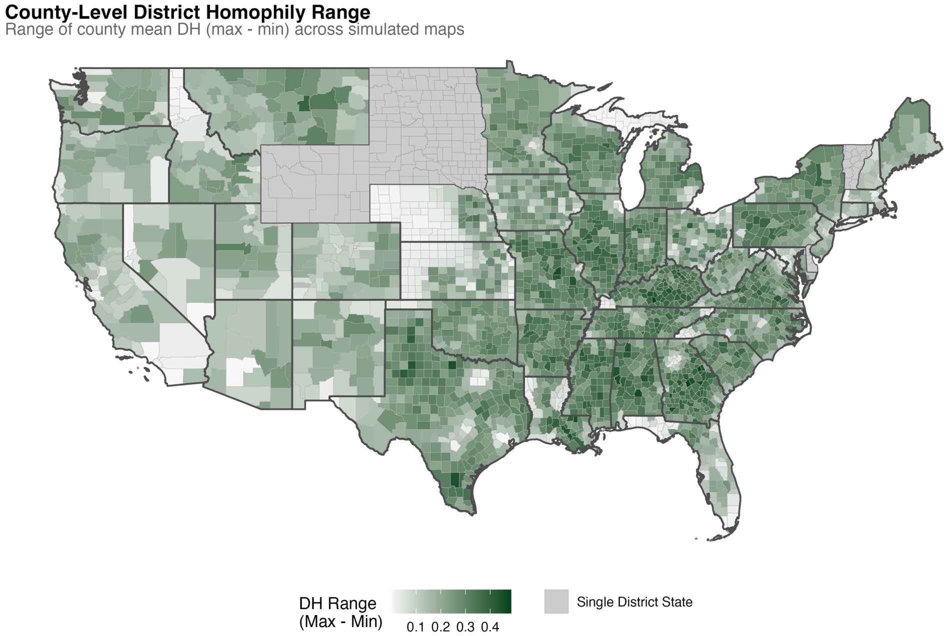


Figure C20: Range of County-Level District Homophily across Simulated Maps

C.10.1 Range of County-Level District Homophily

While the range of district homophily at the state level can be relatively narrow, often when studying district maps we are most interested in inequality within states. Figure C20 plots the range of district homophily across simulated maps for each county.

The figure reveals that there is much more leverage available at the county level. The scale of the figure is five times as wide as in Figure C19. While some counties have small ranges, most counties have a range of at least 10pp, and some counties have ranges as high as 50pp. Again, directly applying the event study estimates for a back-of-the-envelope calculation, this implies that in most counties, going from the minimal to maximal map for that county in terms of district homophily can increase the share of informed voters by at least 3.3pp; in some counties, by over 16pp.

Furthermore, within a state, there can be substantial variation in the ranges of counties. This suggests that, in some states, there is also more scope to draw more or less equal maps, in terms of county district homophily. Figure C21 plots the range of Gini coefficients for each state, measured in terms of inequality in county district homophily in each state under each map. In some states, the level of inequality across counties is relatively fixed, while in other states there is a large range in Gini coefficients across the distribution of

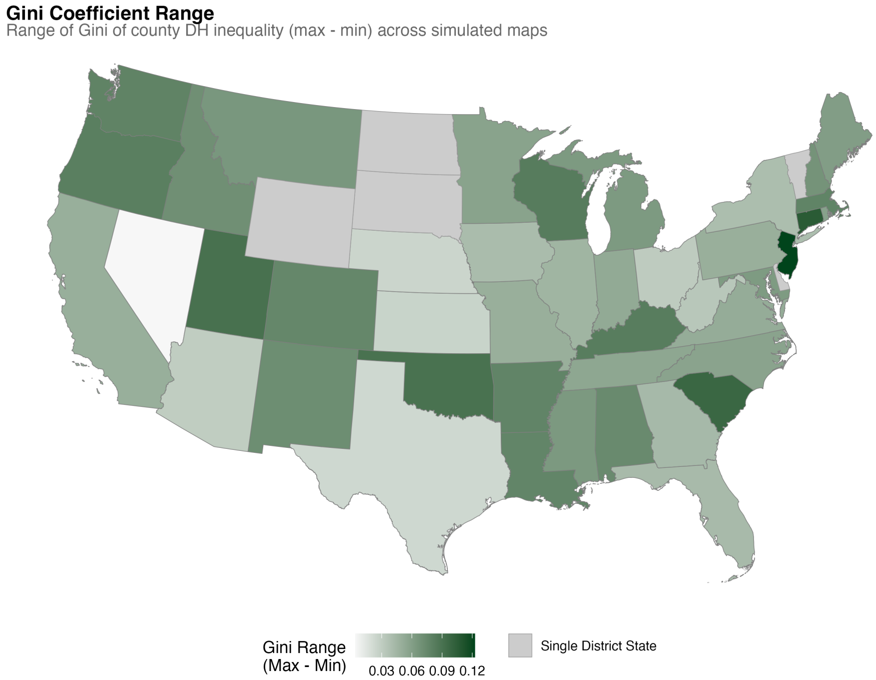


Figure C21: Range of Gini Coefficient of County District Homophily

maps—suggesting that a map drawer would have a large amount of discretionary power over which counties have higher or lower district homophily.

C.11 Comparisons of Enacted Maps to Simulated Maps

The following figures plot, for each continental state that has more than one congressional district, the distribution of the average statewide district homophily across the 5,000 simulated maps. The red dashed line reflects the district homophily under the map that was enacted following the 2020 Decennial Census. Note that the scales of the axes are different on each figure, as possible state district homophily ranges vary substantially.

Figure C22: Enacted Map vs. Distribution of District Homophily Across Simulated Maps (I)

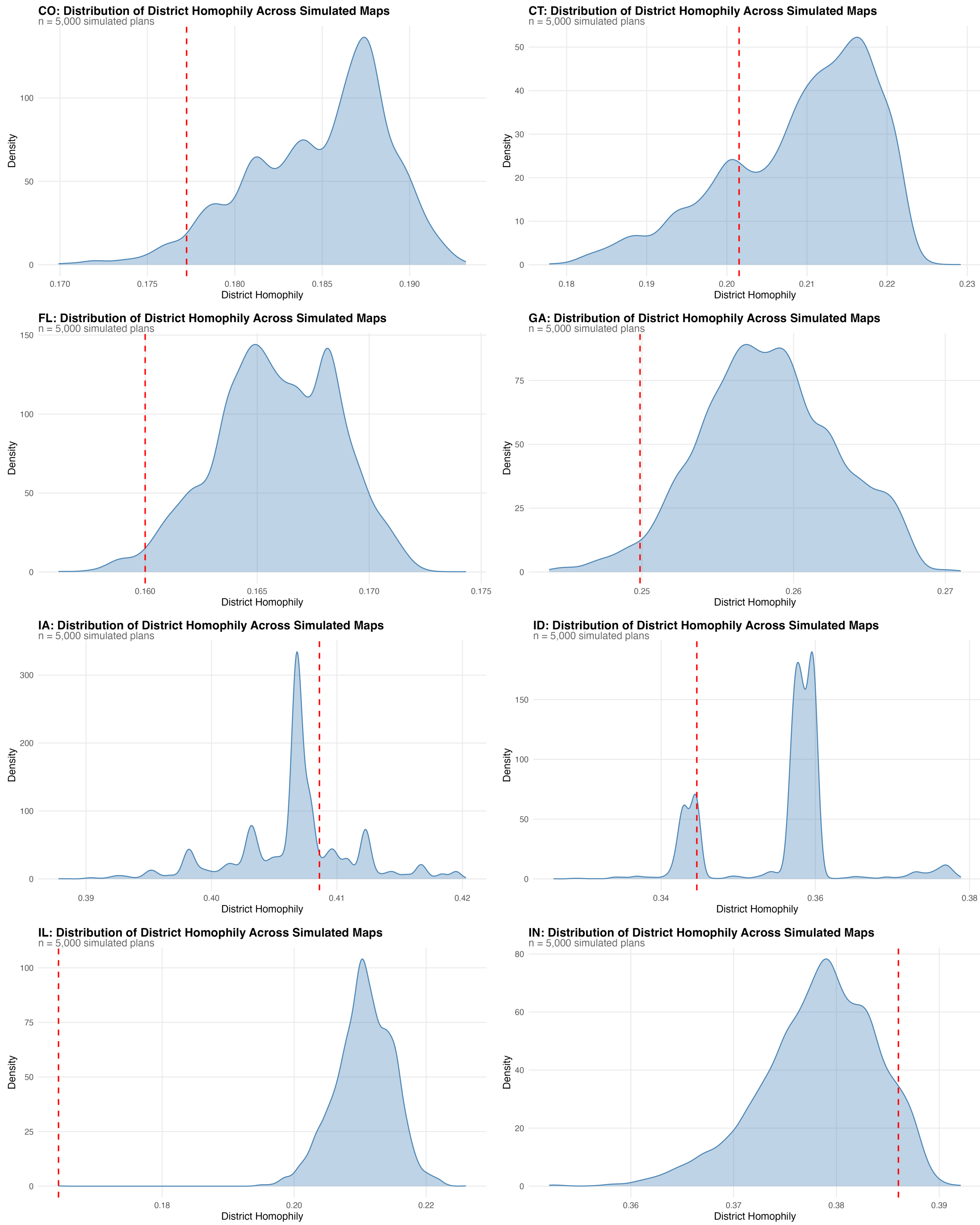


Figure C23: Enacted Map vs. Distribution of District Homophily Across Simulated Maps (II)

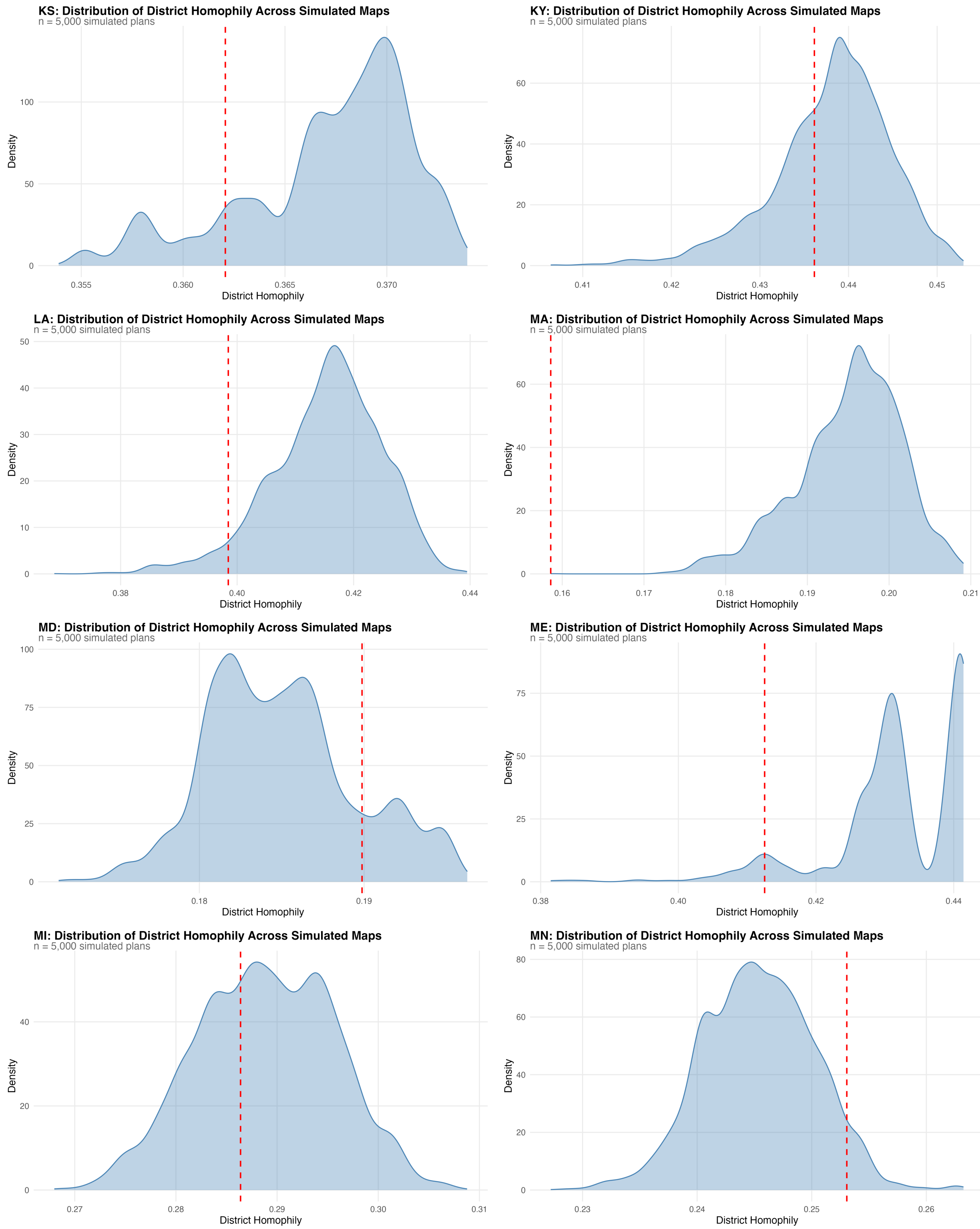


Figure C24: Enacted Map vs. Distribution of District Homophily Across Simulated Maps (III)

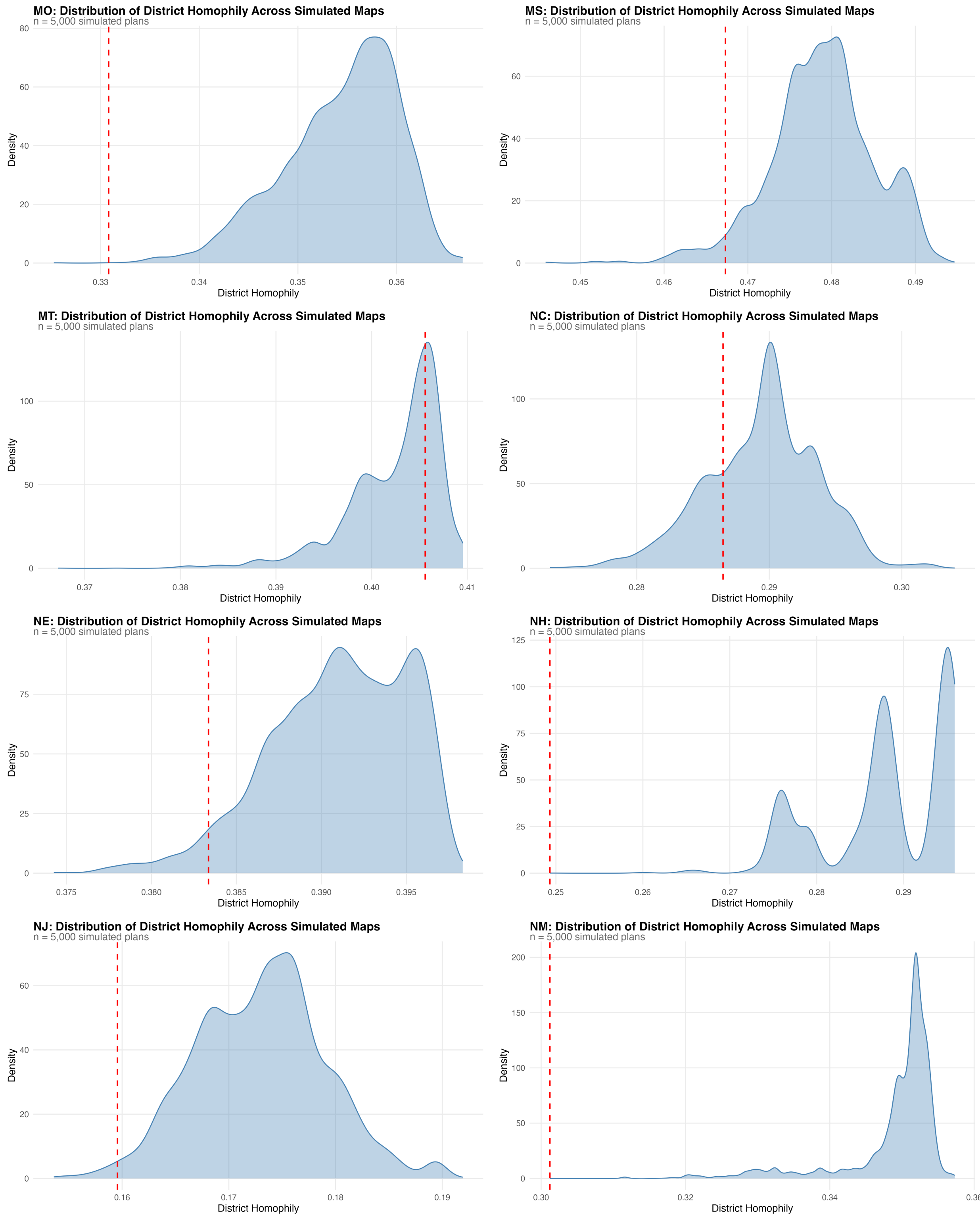


Figure C25: Enacted Map vs. Distribution of District Homophily Across Simulated Maps (IV)

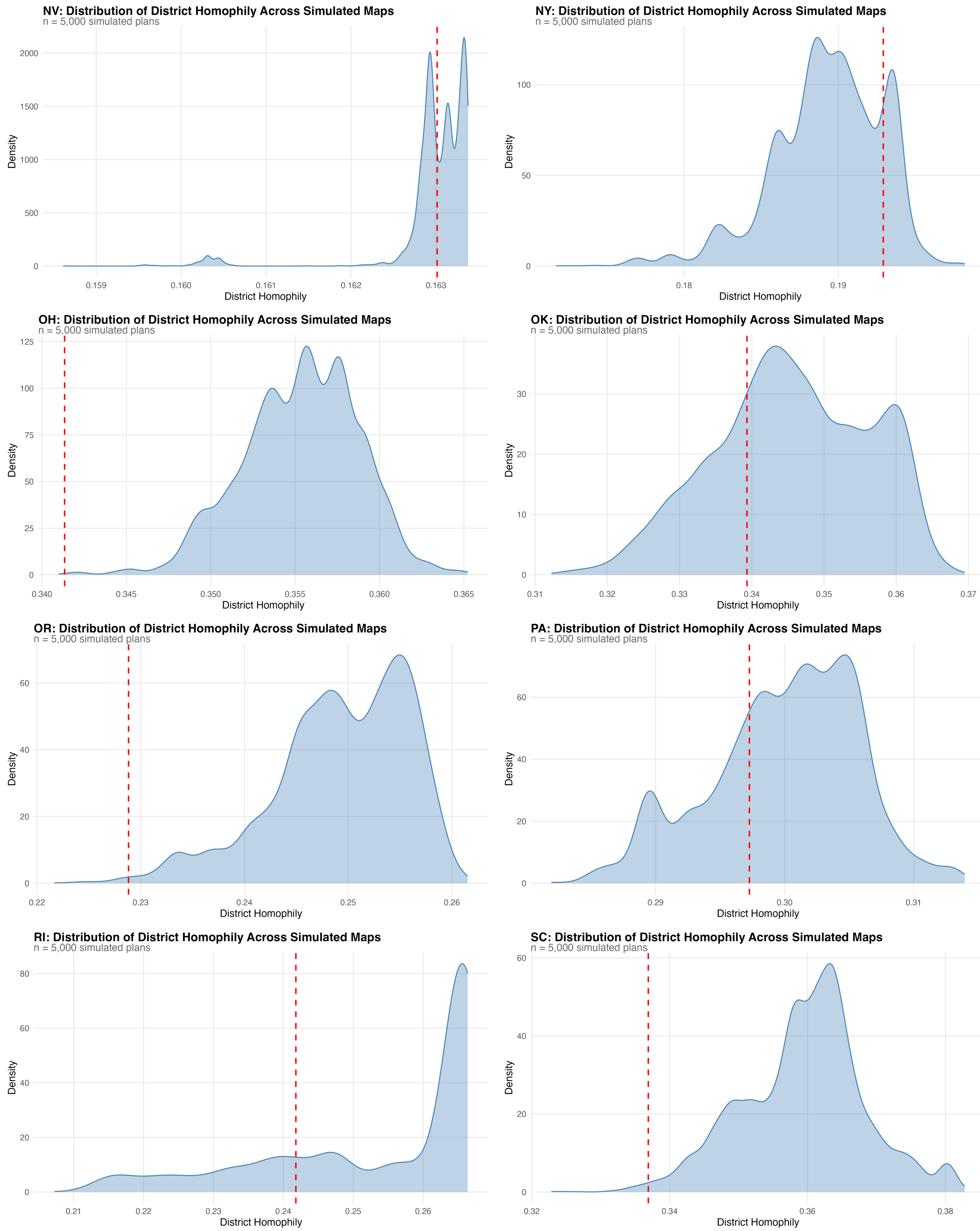


Figure C26: Enacted Map vs. Distribution of District Homophily Across Simulated Maps (V)

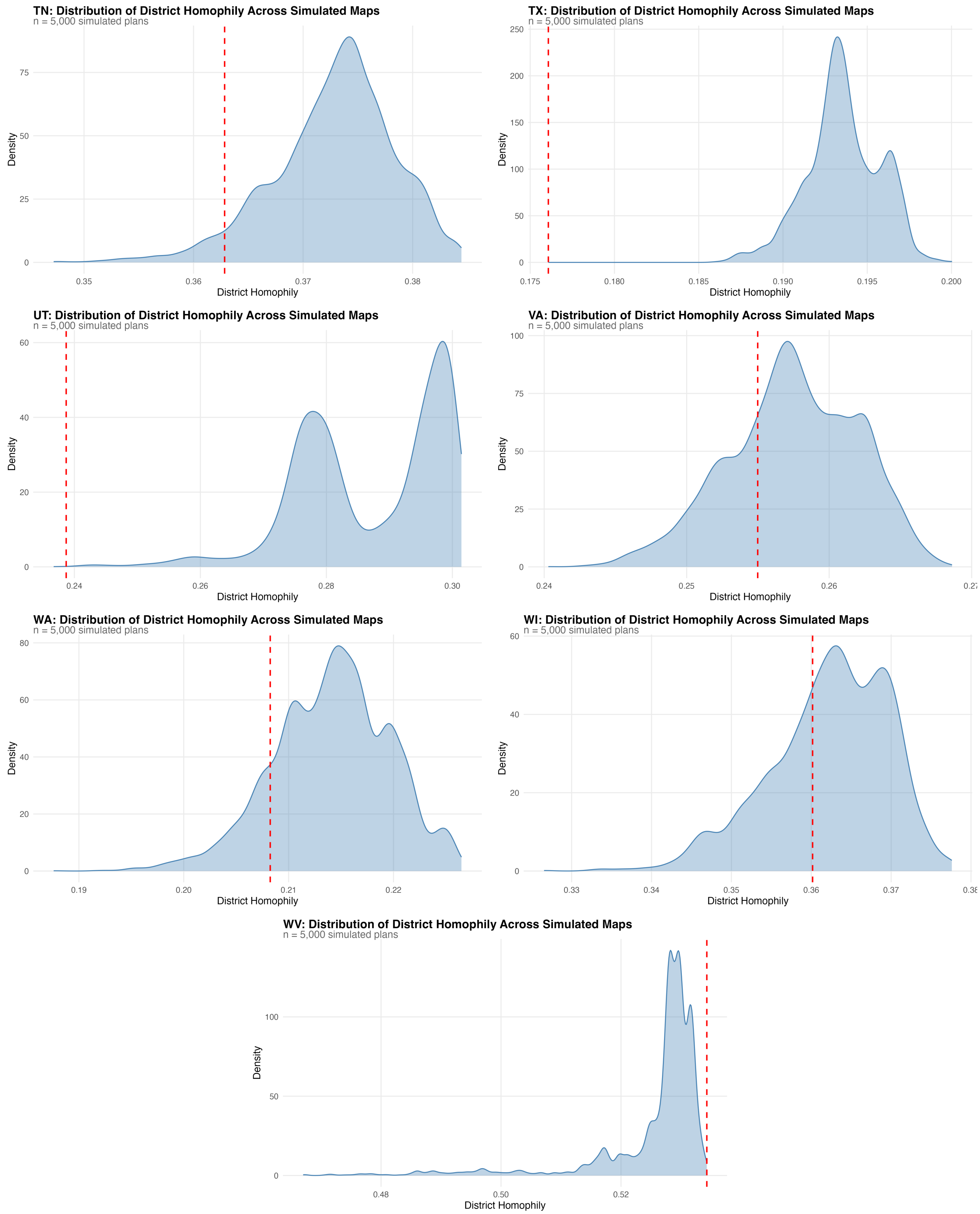


Figure C27: Enacted Map vs. Distribution of District Homophily Across Simulated Maps (VI)

The following figures plot the Gini coefficient of county district homophily under a given map on the x-axis and the statewide average district homophily on the y-axis. Each blue point represents a simulated map, and the red point represents the map that was enacted following the 2020 Decennial Census. There is one figure for each continental state that has more than one congressional district.

The figures demonstrate the distance of the enacted map from the average district homophily-equality in district homophily frontier. The "ideal" map has a higher y-value (higher average) and a lower x-value (more equal across counties), such that it is (under such norms) preferable to be in the upper left corner.

Figure C28: Inequality in District Homophily vs. Average District Homophily (I)

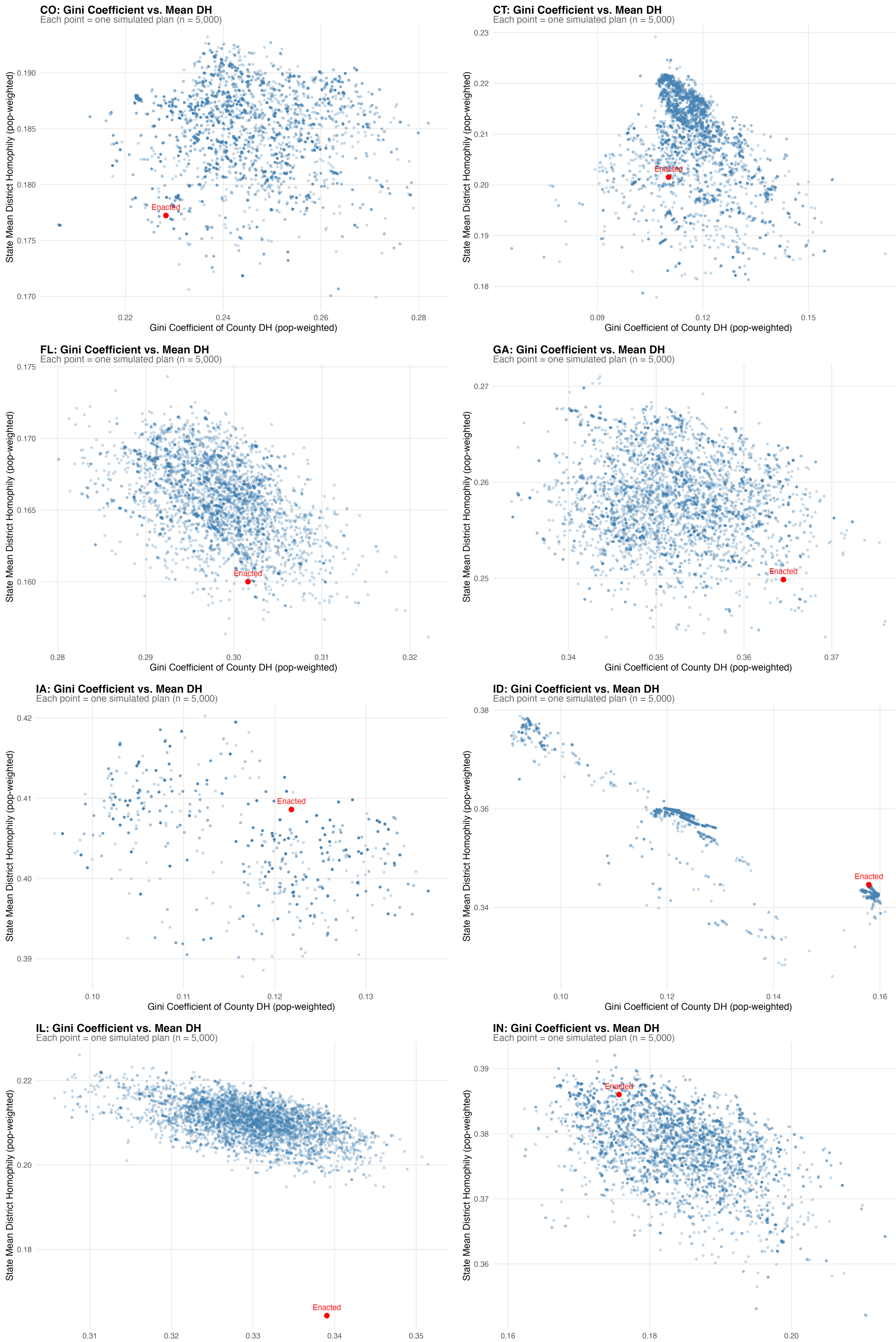


Figure C29: Inequality in District Homophily vs. Average District Homophily (II)

Figure C30: Inequality in District Homophily vs. Average District Homophily (III)

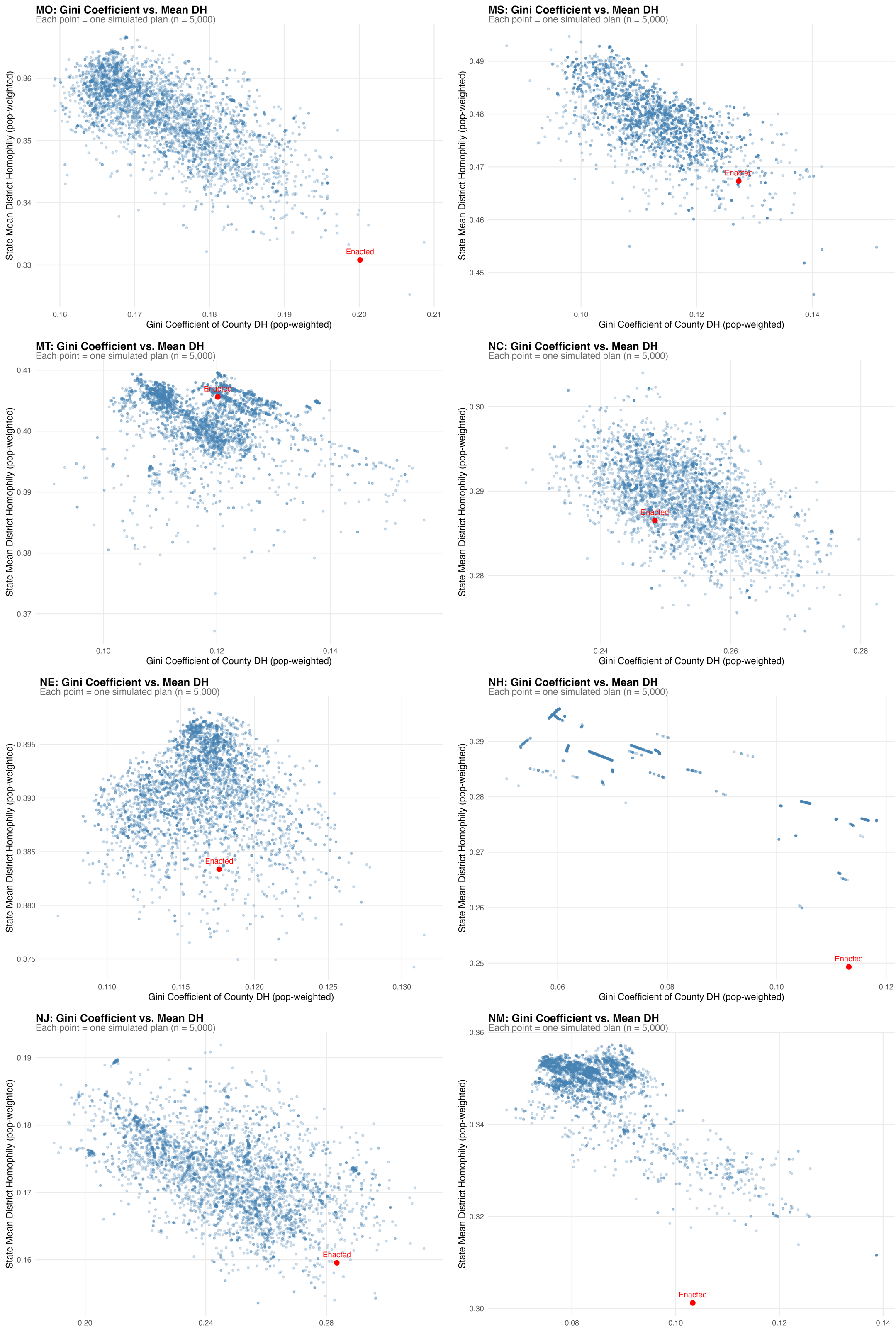


Figure C31: Inequality in District Homophily vs. Average District Homophily (IV)

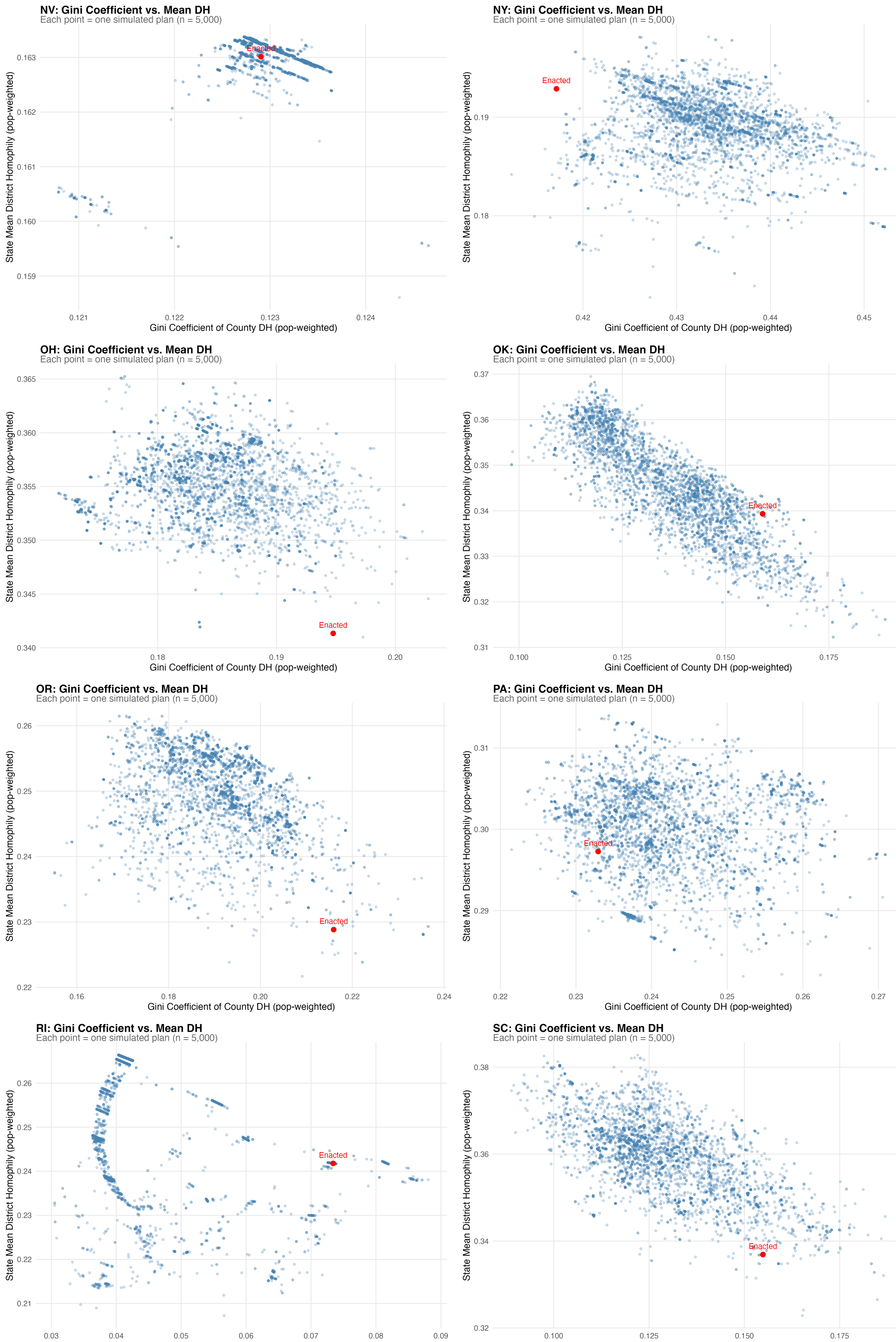


Figure C32: Inequality in District Homophily vs. Average District Homophily (V)

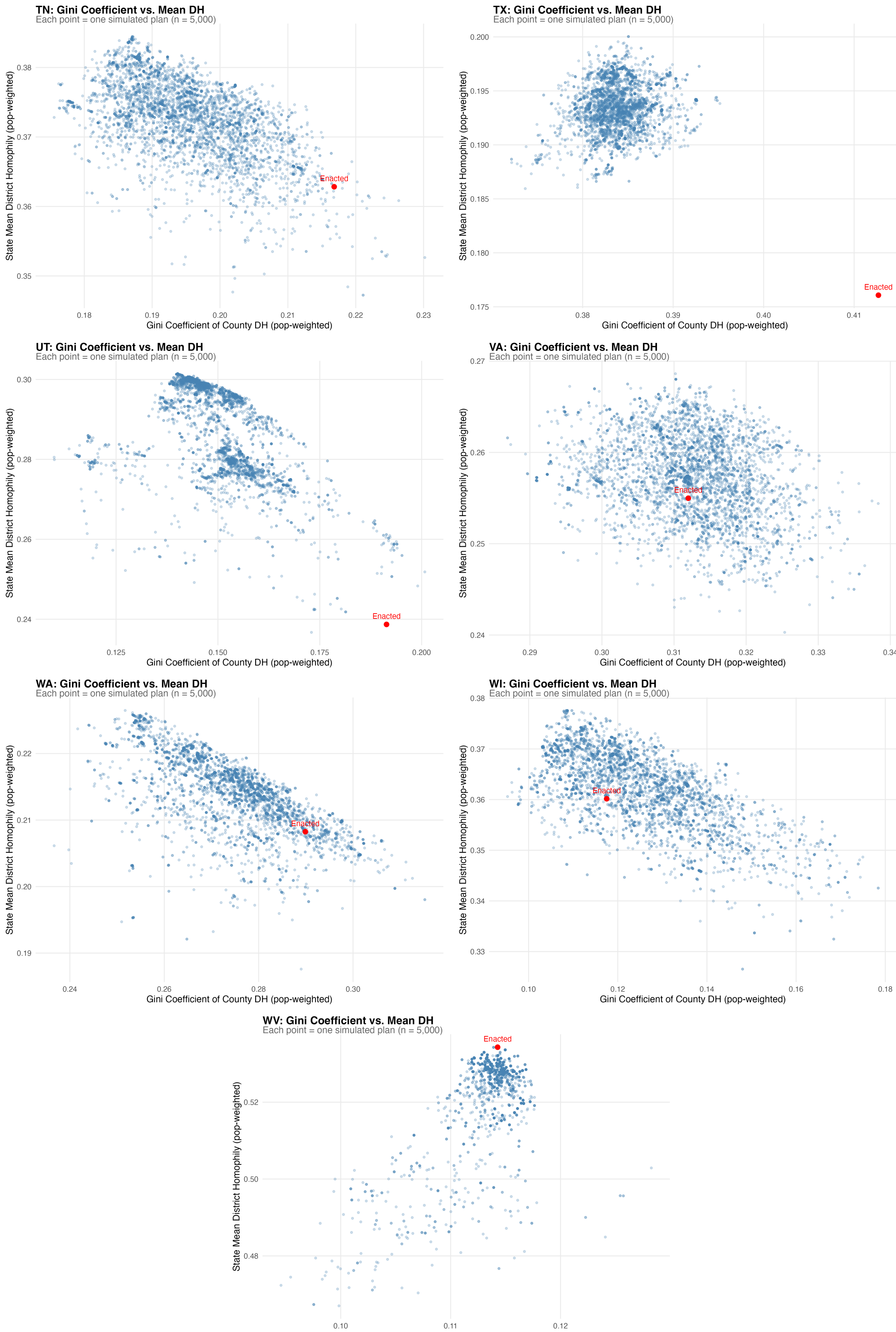


Figure C33: Inequality in District Homophily vs. Average District Homophily (VI)