Social Networks and Electoral Competition

Victoria Mooers

September 15, 2024

(Preliminary and Incomplete)

Click here for latest version.

Abstract

Does it matter if representative district borders reflect communities? In an effort to ensure equal representation, U.S. congressional districts are required to be roughly equal in population. However, congressional district boundaries need not reflect the geography of social networks. If voters share political information through their networks, a community that is more cohesive with its district may be at an informational advantage. This paper examines whether alignment between congressional districts and social networks ("congruence") impacts voters' knowledge of their representatives and their turnout in congressional elections. Using data from the Facebook Social Connectedness Index and changes in district boundaries due to redistricting, I find that increased congruence significantly increases voter knowledge and participation. For example, a 20pp increase in congruence raises the probability that a voter knows their representative's party by 6.3 percentage points, from a mean of 61.7%. Additionally, higher congruence reduces abstention in congressional elections. Using a model of information diffusion, I simulate the share of informed voters under counterfactual congressional district maps, developing a framework that can be used to evaluate proposed maps. These findings suggest that aligning political boundaries with community networks can improve democratic engagement.

1 Introduction

The principle of equal representation enshrined in the U.S. Constitution mandates that congressional districts within a state have roughly equal populations. However, these districts often do not align with the organic social networks of communities. This misalignment can have significant implications for democratic engagement. The existence of a "public sphere," where constituents can discuss shared electoral interests, has been postulated as necessary for a well-functioning democracy (Habermas et al. 1992). In cohesive districts, where social networks are well-aligned with political boundaries, voters may more easily learn from each other about their representatives. Conversely, people living in socially fragmented districts may be deprived of the opportunity to learn much from their fellow constituents, as most people in their networks are represented by someone else. In this sense, a community that is more cohesive with its district may be at an informational advantage.

Previous research has highlighted the importance of peer effects on political knowledge and participation, often focusing on individual interactions, capturing the direct spillovers from one individual to another (Fafchamps et al. 2019, Quintelier et al. 2012, Sinclair 2012), or taken the role of the overall environment as an object of study, with social networks forming one of several dimensions of the environment (Brown et al. 2023, Cantoni and Pons 2022). This paper seeks to capture the role of peer effects in the aggregate, separated from other contextual factors. In particular, I examine how the alignment between social networks and congressional districts impacts voter knowledge and electoral participation in the U.S.

Using the Social Connectedness Index (SCI) derived from Facebook data (Bailey, Cao, et al. 2018), I measure the extent to which counties' social networks align with congressional districts. Specifically, for each county, I measure the share of the county's friends that live in the same congressional district as the county. I refer to this measure as congruence (analogous to Snyder and Strömberg 2010). I find that congruence varies substantially across the country, with a mean of 41% but a range of 2% to 87%.

Using an event-study design, I leverage the natural experiment provided by redistricting to identify causal impacts while controlling for determinants of social networks and congressional district borders. I find that voters are indeed more informed when their social networks better align with their congressional districts. For example, I find that a 20pp increase in congruence increases the probability that a voter knows their representative's party by 6.3pp, from a mean of 61.7%. Additionally, I find that congruence generally does not impact placebo outcomes like familiarity with the governor or senators. I find that congruence is also associated with increased participation in House elections, suggesting that aligning political boundaries with community networks can foster more informed and engaged electorates.

Notably, an expansive literature has studied the impacts of social media and technology access on political knowledge and participation (Enikolopov, Makarin, et al. 2020, Manacorda and Tesei 2020, Di Tella et al. 2021, Guriev et al. 2021) and the responses of politicians (Bessone et al. 2022). In this paper, I am agnostic about whether information is spread online or offline, and instead focus on the role of the geography of connections. Indeed, the SCI is a useful measure of social networks because it closely reflects offline networks (Bailey, Cao, et al. 2018, Kuchler et al. 2022), in part because Facebook usage is relatively even across demographic groups (Auxier and Anderson 2021, Chetty et al. 2022) and also because Facebook friendships are persistent, accumulating throughout a lifetime. The SCI also captures friendship patterns that might be lost in other proxies for social networks, such as physical proximity.

Nonetheless, I test the robustness of my findings to alternative proxies for social networks by constructing an analogous measure of congruence based on commuting flows. Using commuting flows, I find similar but smaller impacts on voter knowledge. Commuting flows tend to represent much denser networks (the mean commuting congruence is 86%) and as such may not capture the role of more geographically distant ties.

This paper makes three primary contributions:

First, I document how well social networks are reflected by existing congressional district boundaries. There is large variation across the country in how socially connected people are to their congressional districts. Many demographic factors also are highly correlated with congruence, indicating that the possible advantages of congruence are not distributed equally across groups. However, changes in congruence following redistricting are largely uncorrelated with observable demographics, suggesting no obvious biases over time.

Second, I establish a causal relationship between social network congruence and voter knowledge and behavior. I contribute to the literature on the effect of political context on voter behavior, such as Cantoni and Pons 2022 and Brown et al. 2023, who measure the aggregate effects of place on outcomes like party registration and turnout behavior. By directly focusing on the alignment between congressional districts and social networks, I am able to separate the impacts of peer effects from other place-specific elements of the environment, such as institutions.

Third, I highlight the policy implications of designing congressional districts that reflect social networks. The findings suggest that more congruent districts could lead to better-informed electorates and higher participation in House races, thereby enhancing the quality of democratic representation. Using a model of information diffusion, I simulate the share of informed voters under counterfactual district maps, demonstrating a framework that can be used for evaluating the informational consequences of proposed maps. This contribution is particularly relevant for policymakers aiming to improve democratic engagement through redistricting practices.

2 Model

The following model formalizes the connection between my primary measure, congruence (the share of a county's friends that are in the same district as the county) and the share of voters in a county that are informed about their representative.

Consider pieces of news about congressional representatives arising and spreading in a population. In a given area, what is the equilibrium share of people who have learned some relevant (i.e., sufficiently recent) news about their representative? I represent this process using a mean-field approximation, applying Jackson and López-Pintado (2013)'s model of diffusion with homophily and heterogenous types.

2.1 Informedness states

Any individual can be informed (state 1) or uninformed (state 0) about their own congressional representative at any given point in time. Informed individuals can become uninformed, each piece of news eventually becomes irrelevant.[1] Individuals only care about news about their own representative, and as such will only become informed if they receive a piece of news about their *own* representative.

2.2 Types of individuals

Let $\mathcal{D} = [1, ..., D]$ be the set of all congressional districts and $\mathcal{C} = [1, ..., C]$ be the set of counties.[2] Individuals are characterized by the congressional district $d \in \mathcal{D}$ and county $c \in \mathcal{C}$ in which they reside; accordingly, there are $C \times D$ possible types of agents. An agent is of type (c, d) if they live in the intersection of county c and district d.

There is a continuum of agents $N = [0, 1]$, and they are partitioned by type: $n(c, d) \in [0, 1]$ denotes the fraction of agents of type (c, d). I consider a representative individual from each county. Empirically, for many types, $n(c, d) = 0$, because each county only intersects with one or a few districts.[3] Call $D(c)$ the set of districts that county c intersects with. Let the share of a county's population in each district be $q_{(c,d)} = \frac{n(c,d)}{\sum_{d' \in D(c)} n(c,d')}$; Q represents the matrix of these shares.

Because I will not be using data at a lower level than that of $\mathcal{C}$, I will consider a representative agent from each county. This agent is a population-weighted average of the types of agents within the county (see below).

[1] Equivalently, agents can forget.

[2] I focus on counties in order to align with the data in my application. However, the model can be written analogously for any other geography that intersects districts.

[3] There are 3,143 counties and county equivalents in the US, and 435 congressional districts. **[[X%]]** of counties are fully contained within one congressional district. LA County intersects with 17 congressional districts, the most of any county.

2.3 The random news-learning process

A random set of agents are initially informed, or seeded with a piece of news about their own representative. Once informed, agents forget the news (or the news becomes irrelevant to them) at a rate $\delta > 0$.

Each period, every agent meets with one friend to share news. They share news about their own representative if they are informed, and they share noise otherwise. However, assume that there are some frictions to communicating information, such that in any given meeting they share with probability $\alpha \in (0, 1]$.[4]

While an agent may receive a piece of news from any friend, they are only interested in the news about their own representative, as this is the only information that matters for their decision at the ballot box. Accordingly, an individual only enters state 1 (and thereby is labeled as informed and becomes capable of passing on information) if they receive information about their own representative, either through being initially seeded or through receiving news from an informed friend from the same district.

2.4 Friendship shares

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

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

where $\pi_{(i,j)(k,l)} \geq 0$ is the share of friends of the representative agent from county i and district j that are from county k and district l.[5] Accordingly, $\sum_{k=1}^{C} \sum_{l=1}^{D} \pi_{(i,j)(k,l)} = 1$. While $\pi_{(i,j)(k,l)}$ need not equal $\pi_{(k,l)(i,j)}$, assume that if $\pi_{(i,j)(k,l)} = 0$ then $\pi_{(k,l)(i,j)} = 0$, since friendships are mutual.

As described above, what ultimately matters for whether an individual becomes informed is whether they observe news from friends in their own district. News from friends outside their district is simply noise.

There is no ranking between districts and counties: counties can be fully within districts, districts can be fully within counties, or neither. While some states have implemented rules that require districts to follow county lines, not all states have. Further, in dense urban counties like Manhattan, the constitutional requirement that each congressional district within a state must as nearly as possible represent an equal number of people demands that there must be multiple congressional districts within Manhattan.

[4] 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 conveyer of the news fails to pass it on. I assume away from $\alpha = 0$ because it gives the trivial equilibrium where everyone is uninformed.

[5] Equivalently, if we were to consider a random agent from the county, the probability that an agent from county i and district j is friends with an agent from county k and district l.

The county-level SCI data contains information on the strength of social connection between county c and county k (for every county pair c and k). Let $\pi_{c,k}$ represent the share of county c's friends that live in county k, regardless of district. The SCI data can then be used to construct the county-county friendship matrix Π':

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

The SCI data does not contain information about how county c's friends are distributed within k. In particular, within county k, I do not know how c's friends are distributed across the districts in $D(k)$. This means that I do not generally know $\pi_{(c,d)(k,l)}$: I only know it in the cases where both county c is fully contained within district d, and county k is fully contained within district l.

To address this, I take population-weighted averages. I rely on the assumption that friendships are uniformly distributed within a county. Under this assumption, $\pi_{(c,d)(k,l)} = \pi_{c,k} \times q_{(k,l)}$. That is, I multiply the share of county c's friends that live in county k by the share of county k that lives in district l.

I can then calculate π_c, the share of county c's friendships that are between people in the same district, where $\pi_c = \sum_{d \in D(c)} \sum_{k \in C} \left(\pi_{c,k} \times q_{(c,d)} \times q_{(k,d)}\right)$. π_c represents the population-weighted average congruence for county c. Appendix Section A.1 provides the full derivation.

2.5 Timing

To summarize, it is easy to conceptualize the process as occurring in discrete periods that each proceed as follows:

1. We begin each period with some agents in state 1 (informed) and some agents in state 0 (uninformed).

2. News is shared. Assume that everyone shares news with one friend, and they share news about their own representative if they are informed and noise otherwise.

3. Anyone who received a piece of news about their own representative becomes informed with probability α.

4. At the end of each period, some informed people forget and become uninformed at rate δ.

2.6 Individual transitions from state to state

Let $\rho_c(t)$ denote the share of people in a given county pair that are informed at time t; we want to know the equilibrium value of this. Let $\rho_{(c,d)}(t)$ denote the share of type (c,d) people that are informed at time t. We can then write $\rho_c(t) = \sum_{d \in D(c)} q_{(c,d)} \rho_{(c,d)}(t)$. Let $\boldsymbol{\rho}(t) = \{\rho_c(t)\}_{c \in \mathcal{C}}$ represent the sequence of these functions.[6]

What will be most relevant to a given agent is the share of their friends in the same district that are informed at time t; this is not necessarily equal to the overall share of people in their county or in their district that are informed at time t, because their distribution of friends across counties in the same district is not necessarily equal to the overall distribution of people across counties in their district (and vice versa). As such, let $\tilde{\rho}_c(t)$ represent the average share of a randomly chosen individual from c's friends that are informed at time t, conditional on being from the same district as the individual:

$$\tilde{\rho}_c(t) = \frac{1}{\pi_c} \sum_{k \in C} \sum_{d \in D(c)} \left(q_{(c,d)} \times q_{(k,d)} \times \pi_{c,k} \times \rho_k(t) \right) \tag{1}$$

Recall that everyone only shares news with one friend each period, and they randomly sample among their friends. For an individual from c, on average a share π_c of their friends are in the same district (and thus can potentially share relevant news that informs them). Of those, on average $\tilde{\rho}_c(t)$ are informed. Accordingly, for an uninformed person, the probability of becoming informed from a friend in a given period is $\alpha \pi_c \tilde{\rho}_c(t)$: an agent's probability of becoming informed depends on their congruence, the probability that their friends in their district are informed, and the frictions in sharing information.

A sequence of transition functions $\{f_c(\boldsymbol{\rho})\}_{c \in \mathcal{C}}$ describe the dynamics. At any time $t \geq 0$ the state of the system keeps track of the informed and uninformed agents. Representing time $t \in \mathbb{R}_+$ as continuous, $\frac{d\rho_c}{dt} = f_c(\boldsymbol{\rho}(t))$. In particular,

$$f_c(\boldsymbol{\rho}(t)) = \underbrace{(1 - \rho_c(t))}_{\text{In state 0}} \underbrace{(\alpha \pi_c \tilde{\rho}_c(t))}_{\text{Rate } 0 \to 1} - \underbrace{\delta \rho_c(t)}_{\text{Rate } 1 \to 0 \times \text{In state 1}} \tag{2}$$

[6] In the data, while I can construct $\pi_{c,d}(d)$ and $\rho_{c,d}$ within any given year (CES respondents are linked to both a county and a district), measuring changes around redistricting only makes sense at the county level: county boundaries remain unchanged, but district boundaries and numbering changes – it is unclear what the same (c,d) would be after redistricting. As such, I focus on π_c and ρ_c (county-level objects) in order to more appropriately match the object I can estimate.

2.7 Steady states and dynamics

In the steady state, $\frac{d\rho_c}{dt} = f_c(\boldsymbol{\rho}(t)) = 0$ for all c, and therefore the steady state level is independent of time. Consequently, at the steady state

$$\begin{aligned} \delta\rho_c &= (1-\rho_c)(\alpha\pi_c\tilde{\rho}_c) \\ \Rightarrow \rho_c &= \frac{\alpha\pi_c\tilde{\rho}_c}{\alpha\pi_c\tilde{\rho}_c+\delta} \end{aligned} \tag{3}$$

If α and δ are both positive, the first order effect of an increase in π_c is an increase in ρ_c (in general, ρ_c is increasing in π_c so long as the change in π_c does not cause too large of decrease in $\tilde{\rho}_c$, see Appendix Section A.2.1).

Replacing (3) in (1)

$$\tilde{\rho}_c = \frac{1}{\pi_c}\sum_{k\in C}\sum_{d\in D(c)}\left(q_{(c,d)} \times q_{(k,d)} \times \pi_{c,k} \times \frac{\alpha\pi_k\tilde{\rho}_k}{\alpha\pi_k\tilde{\rho}_k+\delta}\right) := \phi_c(\tilde{\rho}_1, ..., \tilde{\rho}_C)$$

Each steady state of the model will correspond to a fixed point of ϕ_c. The system of equations $\{\phi_c\}_{c\in\mathcal{C}}$ implicitly characterizes the steady states $\{\rho_c\}_{c\in\mathcal{C}}$.

2.8 Swing Voter's Curse

3 Methods

3.1 Congruence: A Measure of District and Network Alignment

I study the consequences of geographical mismatch of social networks and political boundaries. In particular, I focus on the alignment of county-level social networks and congressional districts. I capture this alignment by constructing "congruence": for a given county, the share of county residents' friends that live in the same congressional district.

I use data representing the Facebook friendship graph to construct congruence for counties in the 48 contiguous U.S. states, and I show that congruence varies substantially across the U.S. I demonstrate that congruence is (not surprisingly) correlated with many determinants of social networks and district borders, but *changes* in congruence due to redistricting largely are not. Consequently, causal identification of the impacts of congruence can leverage these changes over time, as I discuss further in Section 3.2.

3.1.1 Facebook Social Connectedness Index: A Proxy for Social Networks

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 (Bailey, Cao, et al. 2018). For each pair of locations, $SCI_{i,j}$ is constructed as the relative probability of a friendship link between users in two geographic units i and j:

$$\text{SCI}_{i,j} = \frac{\text{Friendship Links}_{i,j}}{\text{Facebook Users}_i \times \text{Facebook Users}_j}$$

That is, the SCI is the number of friendship links between the two locations, normalized by the total number of possible connections between them.[7] I use the SCI for U.S. county-county pairs from the October 2021 snapshot.[8]

The SCI is an effective proxy for real-world social networks because it captures social ties that might not be revealed in geographic-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.[9] Second, Facebook usage rates across counties are uncorrelated with demographics like income (Chetty et al. 2022).[10] Nonetheless, in Section 6.2, I demonstrate that results are qualitatively similar if I use commuting flows as an alternative proxy for social networks.

[7] This makes the SCI independent of differences in the number of users in each location. Double the SCI means that it is twice as likely for a user in i to be friends with a user in j. 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.

[8] 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 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.

[9] For example, Enke et al. 2023 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).

[10] Similarly, at the time of the snapshot, survey-reported Facebook usage rates were also relatively even across demographic groups nationwide (Auxier and Anderson 2021). 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). 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 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).

3.1.2 Construction of Congruence from SCI

Whereas the SCI gives the relative probability of a friendship between two counties, congruence is the share of a county's friends that live in the same congressional district as the people in the county. Accordingly, to construct congruence I need to appropriately aggregate the SCI.

Recall from the model (Section 2.4) that congruence is defined as $\pi_c = \sum_{d \in D(c)} \sum_{k \in C} \left(\pi_{c,k} \times q_{(c,d)} \times q_{(k,d)}\right)$, where C represents the set of all U.S. counties; $D(c)$ represents the set of all congressional districts c intersects; $q_{(c,d)}$ and $q_{(k,d)}$ represent the share of county c and k's population that lives in district d, respectively; and $\pi_{c,k}$ is the share of county c's friends that live in county k. I use the SCI to construct the $\pi_{c,k}$ shares.

The county-county friendship shares can be represented as

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

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

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

(from here on referring to "Friendship Links" as "Links" for brevity), and then we could calculate

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

However, the number of Facebook users in each county is not made available, so this is not possible to back out. 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

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

which is feasible to calculate. I get each county's population from the 2020 Decennial Census. I also use the 2020 Decennial Census to construct the Q matrix of county-district population shares.

Finally, the SCI is only available for one snapshot of the social network, in 2021. In order to derive congruence in each year over the period, I hold the social network fixed and I re-calculate congruence with each congressional border change – consequently, all changes in congruence are due solely to changes in the location of the district border.[11] As discussed above, social networks are slow-changing, so this assumption is not unreasonable. Further, holding the social network fixed at its 2021 structure and projecting it back in time will primarily introduce measurement error. I construct congruence for each year from 2002 to 2022, i.e. the 107th-117th Congresses.

3.1.3 Examples of Congruence

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

Figure 1 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 of 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

[11] 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).

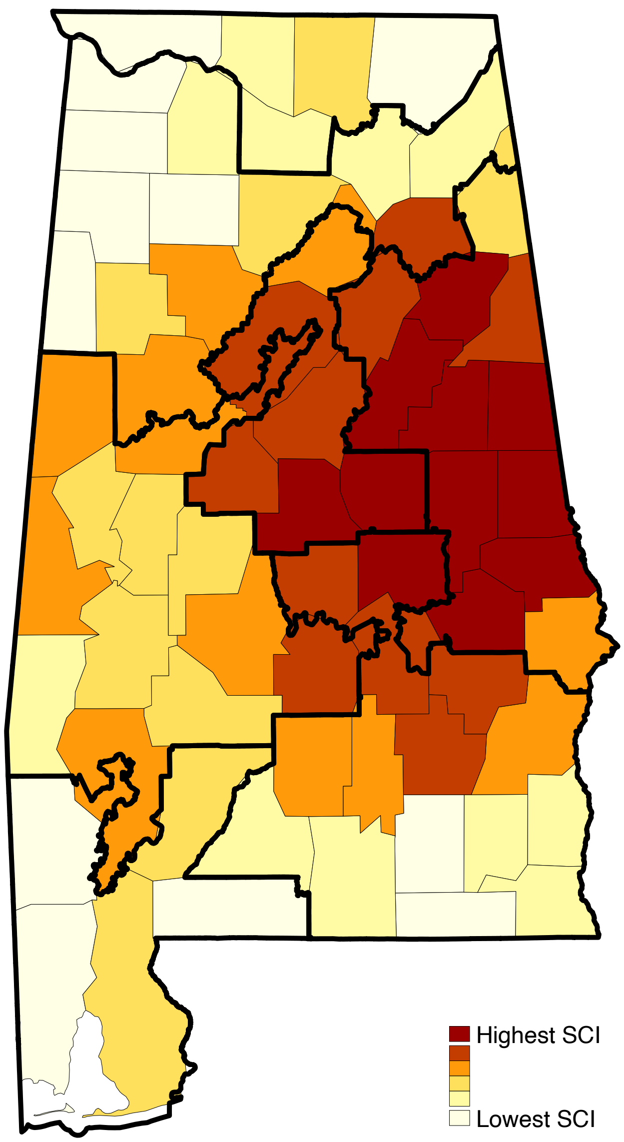

Figure 1: SCI of Coosa County, Alabama. Left: 117th Congress boundaries; Right: 118th Congress boundaries.

Coosa County is grouped in with much of its social network. This is reflected in Coosa County's congruence before and after redistricting. In Figure 2, the left map represents the congruence of each county of Alabama before the 2022 redistricting, while the right map represents congruence after. Again, there is an equal number of counties in each bin, so congruence levels should be interpreted as congruence relative to other counties in Alabama. As we might predict from the SCI maps, Coosa County is among the least congruent Alabama counties 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 becomes one of the most congruent counties in Alabama. In particular, Coosa County experiences a 39.3pp change in congruence, going from 16.6% under the old borders to 55.9% under the new borders.

In Appendix Section B.1.1, I show two more examples. Pondera County, Montana experienced the largest decrease in congruence in the 2022 redistricting, largely because Montana went from having one to two congressional districts thanks to recent population growth. Hocking County, Ohio also experienced a relative large decrease in congruence: it used to be in a congressional district that circumscribed the densest part of its social network, but it was moved to be on the edge of a district that splits its network in two.

3.1.4 Summary Statistics and Predictors of Congruence

Congruence varies substantially across the continental U.S. Congruence 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 congruence.

Among the continental 48 states over the full period, mean congruence is 41% with a standard deviation of 14pp; minimum congruence is 2% and maximum is 87%, while the 1st percentile is 8% and the 99th percentile is 67%. The middle 50% of counties have congruence between 32% and 51%, and the middle 80% of counties have congruence between 22% and 58%.[12]

Appendix Tables 6 and 7 summarize how various geographic and demographic features correlate with congruence, separately in 2010 and 2020. As might be expected given that social networks tend to follow state boundaries (Bailey, Cao, et al. 2018), counties in single district states[13] have higher congruence on average (53%). Additionally, 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; Whitaker

[12] 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.

[13] 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.

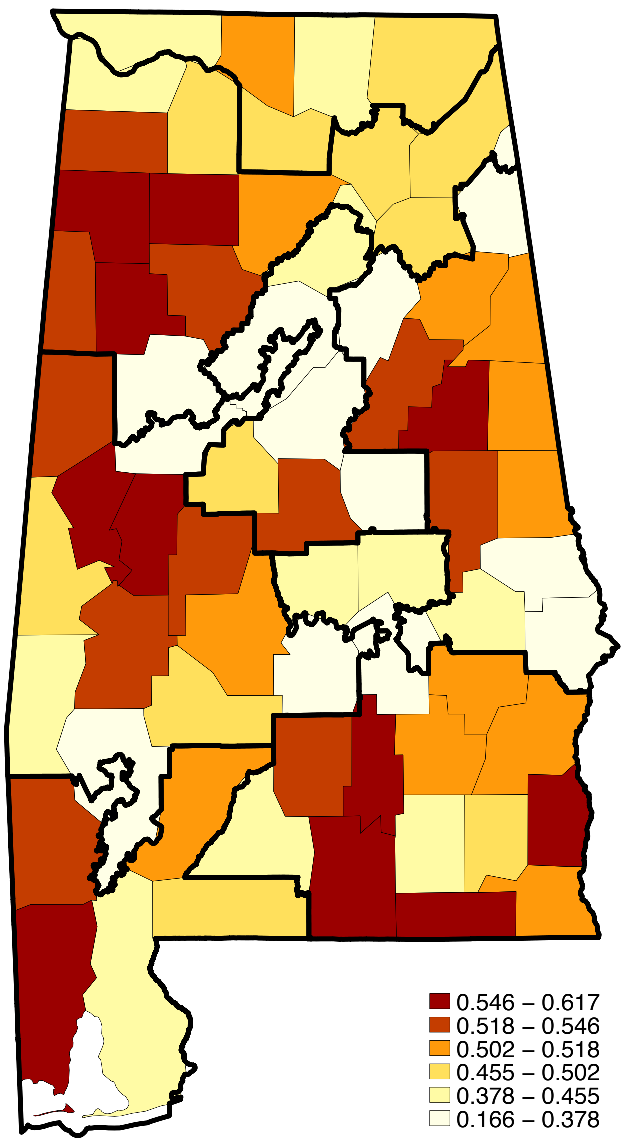

Figure 2: Congruence of Alabama Counties. Left: 117th Congress boundaries; Right: 118th Congress boundaries.

2017), 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 congruence.

The other determinant of congruence 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 congruence. 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 congruence 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 congruence in urban areas.

3.1.5 Changes in Congruence over Time

How is congruence changing over time? I examine changes in congruence due to redistricting following the 2010 and 2020 Decennial Censuses. Recall that I re-calculate congruence for each year by holding the social network fixed. I then calculate the change in congruence for a county following redistricting. In both years, the average change in congruence is nearly zero – 0.1pp in 2013, 0.3pp in 2023 – with a standard deviation of 6.5pp and 5.9pp respectively. In 2013, the biggest drop in congruence was by 36.6pp, while the biggest increase was by 32.7pp. In 2023, 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 congruence (specifically, an absolute change of less than 0.1pp), and following the 2020 Census, 467 counties (15%) experienced nearly no change in congruence. Thus, most counties experience some change in congruence, but very large changes are unusual.

In Appendix Tables 10 and 11, I show the predictors of changes in congruence. Most county characteristics are not correlated with changes in congruence, and most 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 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 congruence and the share with income below the poverty line is negatively correlated. Accordingly, I control for these characteristics in my regressions. However, I show below that in the sample I use for survey outcomes, there is no correlation between changes in congruence and respondents' demographics.

3.2 Redistricting

Congruence measures variation in the match between social networks and congressional district boundaries, but it is not itself exogenous. Congruence 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 congruence, 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 congruence, I measure the impact on outcomes of a change in congruence due to congressional redistricting. I use an event study design, focusing on the redistricting that followed the 2010 Census. 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 congruence. 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 congruence) 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:

$$y_{ict} = \lambda_t + \sum_{\tau=2006}^{\tau=2010} \beta_\tau \Delta\text{Congruence}_c \mathbb{I}(\tau = t) + \sum_{\tau=2014}^{\tau=2022} \beta_\tau \Delta\text{Congruence}_c \mathbb{I}(\tau = t) + X_{ct}\delta + Z_{ict}\gamma + \varepsilon_{ict} \quad (4)$$

where where y_{ict} is the outcome for a given individual i in county c in year t, $\Delta\text{Congruence}_c$ is the change in congruence 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.

4 Outcomes Data and Descriptive Statistics

I study the impact of congruence 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.

4.1 Voters' Information

I test whether voters in more congruent counties 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 congruence 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. Brief descriptions of these variables are in in Table 1. 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.[14] 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. Table 2 shows that, as expected, fewer people select their representative's party than claim to have heard of them, and fewer still select the correct party (though, among those who select a party, the overwhelming majority select the correct party).

As when constructing the congruence measure, I only include respondents in the 48 contiguous states.[15] 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 45% for the average county).[16]

4.2 Voter Turnout and Vote Shares

I test impacts of congruence 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.

[14] 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.

[15] 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.

[16] 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%.

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.
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 1: Descriptions for CES Outcome Variables

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 2: CES Data: Summary Statistics

I begin with CES survey responses in order to study 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 congruence 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.

In order 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 construct the same voting outcomes at the county-by-CD-level, 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), and combine this 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). In my main results I focus on the data from Dave Leip's Election Atlas, but I include analogous results from the other data sources in the Appendix.

Vote counts at the county-by-CD level for all offices is useful to have for constructing one of my primary voting outcomes, "roll-off" (Miller 2022; Snyder and Strömberg 2010). Roll-off is 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

Variable	Description
Roll-Off	$\frac{\text{\# Votes in Top-of-Ticket Race} - \text{\# Votes in House Race}}{\text{\# Votes in Top-of-Ticket Race}}$ From Dave Leip's Election Atlas (county-level); 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 over age 18. 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 3: 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 4: Voting Outcomes: Summary Statistics

share of the top-of-ticket votes cast. 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).[17] The question answered by roll-off is: conditional on turning out to vote for an up-ballot race, how likely is a voter to bother to vote in the House election? This controls for the costs of turning out to vote (it has already been paid by showing up for the top-of-ticket race) and also controls for changes in turnout driven by the top-of-ticket race.

4.3 Campaign Contributions

To be added.

[17] 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.

5 Results

Section 5 presents my main results that congruence increases voters' knowledge about their representatives, and accordingly decreases roll-off in House elections.

5.1 Voters

5.1.1 Voters' Knowledge about Representatives

I estimate positive and significant coefficients for the impact of congruence on voters' knowledge about their representatives. Figure 3 presents the event studies for outcomes "Heard of Incumbent," "Selected Party," and "Selected Correct Party" constructed from the CES data. I find that a change in congruence has an immediate and persistent impact on voters' knowledge. I focus on even years of the CES survey until 2022 (the last year before the next national redistricting event).[18] The β and dashed line on the figures indicate the estimated aggregate effect, from the specification of the form:

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

In the event studies shown, I include district-by-year fixed effects, DMA-by-year fixed effects, and individual demographic controls from the CES; 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. Congruence 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 congruence.

The event studies show that the change in voter knowledge due to changes in congruence in redistricting most strongly takes effect in the first survey after redistricting (2014). Impacts are relatively stable over time. The otherwise 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 congruence by 20pp

[18] 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.

would increase the probability that a respondent in that county has heard of their representative by 1.4pp (recall from Table 2 that the mean is 93.2%). The same change in a county's congruence would increase the probability a respondent in that county selects a party by 5.8pp (from mean 68.6%) and selects the correct party by 6.3pp (from mean 61.7%).[19]

As discussed in more detail in Section 6, I find similar results when using commuting flows as an alternative proxy of social networks. Additionally, I do not find evidence that congruence increases voters' knowledge on placebo outcomes (i.e., the same three outcomes but for the respondent's governor and senators).

5.1.2 Voters' Choices

How does information translate into vote choices? I examine the impact of congruence 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).[20] I run event studies analogous to equation 4 to examine the impact of congruence on voting-related outcomes; however, here I treat 2010 as the base year, as 2012 elections occur under the new district boundaries, and accordingly congruence 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 4, I show event studies for these three

[19] Recall from Section 3.1.5 the a one standard deviation change in congruence following redistricting is roughly 5pp. An increase in congruence by 5pp increases the probability of having heard of the incumbent by 0.5pp, of selecting a party by 1.2pp, and of selecting the correct party by 0.9pp. The largest changes in congruence following redistricting are normally by around 30pp. Such an increase in congruence increases the probability of having heard of the incumbent by 2.9pp, of selecting a party by 7.3pp, and of selecting the correct party by 5.4pp.

[20] 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.

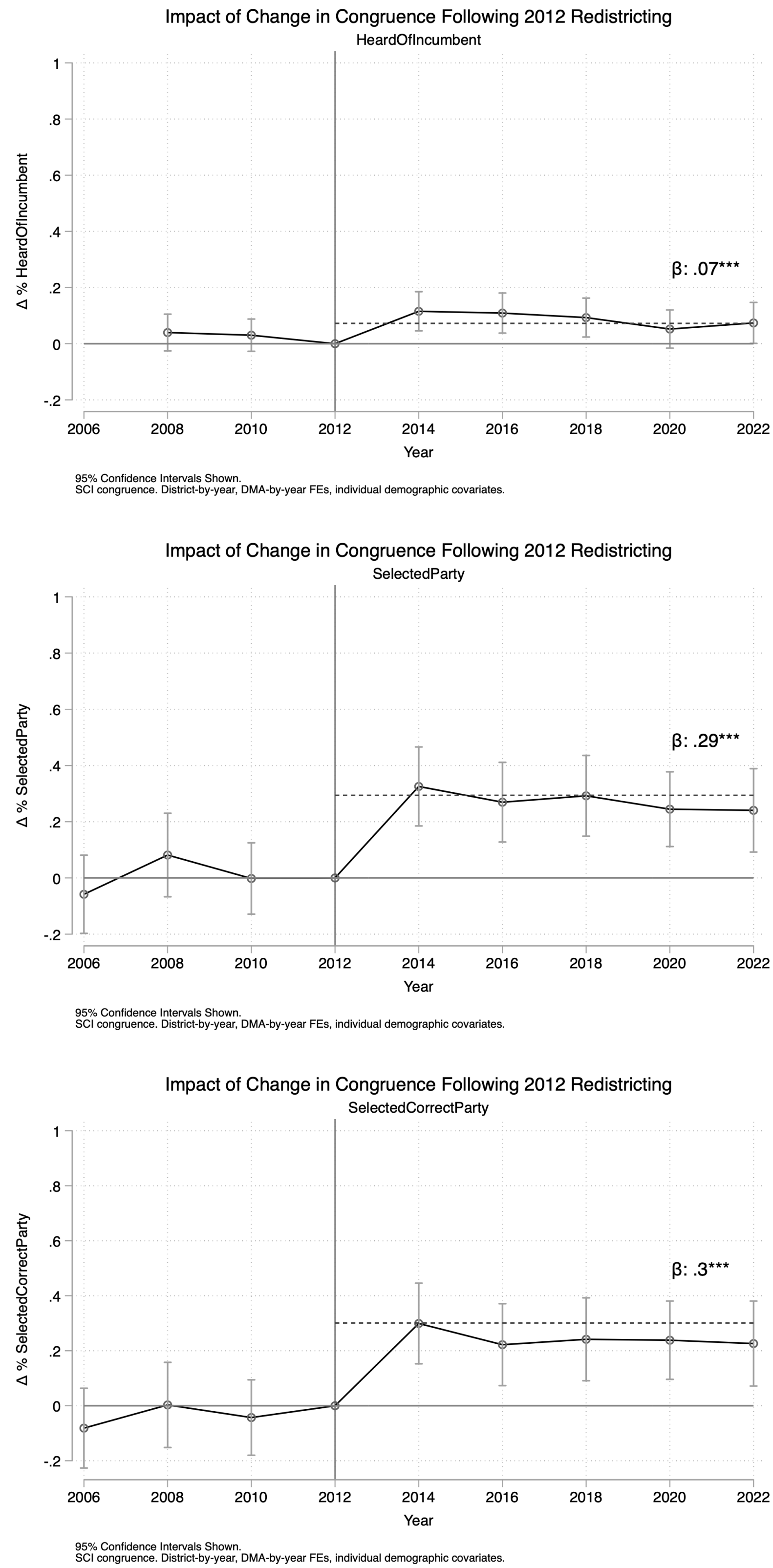

Figure 3: Dynamic Effects of Change in Congruence on Voter Familiarity with Representative

outcomes; I always restrict to cases in which an incumbent exists.[21]

Results are noisy, but an increase in congruence is marginally associated with an increase in preference for the incumbent. However, this increased preference for the incumbent does not come at the cost of preference 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.[22] Here again results are noisy, and I find no impacts on votes for the incumbent or votes for opponents (see Appendix Figure 15). Figure 5, however, suggests that there is a decrease in subjects who report *not* voting in the House election. 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.

Third, I examine impacts on turnout generally (not just for the House election). I look at whether voters are validated as having turned out in the general election, whether voters self-report having turned out in the general election, and whether voters are validated as having turned out in the primary election. On all three, I find no evidence of impacts of congruence (see Appendix Figure 14).

Together, these results suggest that while congruence may not have substantial impacts on self-reported votes for the incumbent or overall turnout behavior, conditional on already turning out to vote, congruence may increase extensive margin participation in House elections. To test this, I next turn to actual vote count data.

Vote Count Data Figure 6 shows the impact of the change in congruence following redistricting on roll-off, as measured using county-level vote counts from Dave Leip's Election Atlas. 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.

[21] 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.

[22] 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.

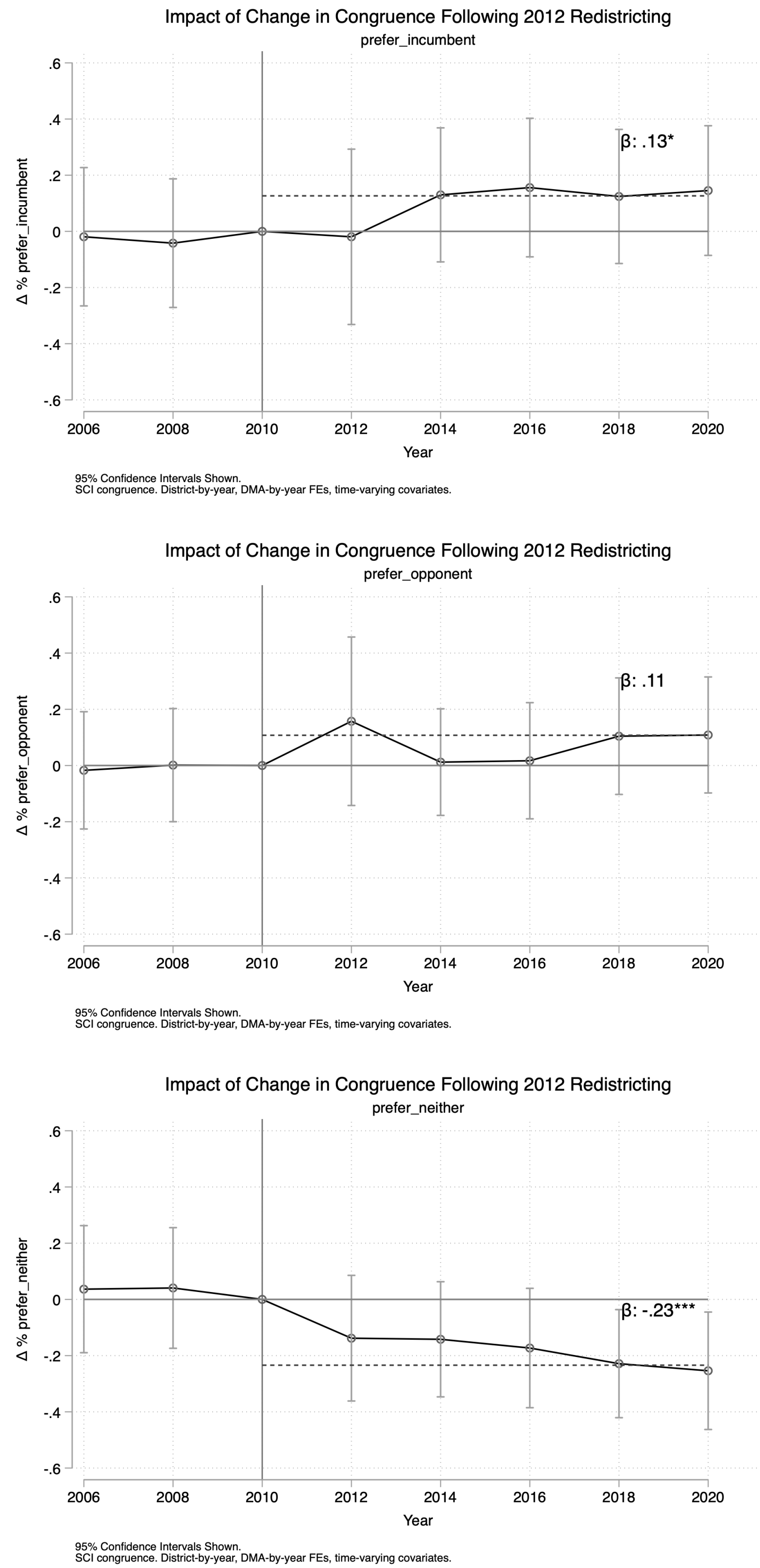

Figure 4: Dynamic Effects of Change in Congruence on Voter Preferences in CES

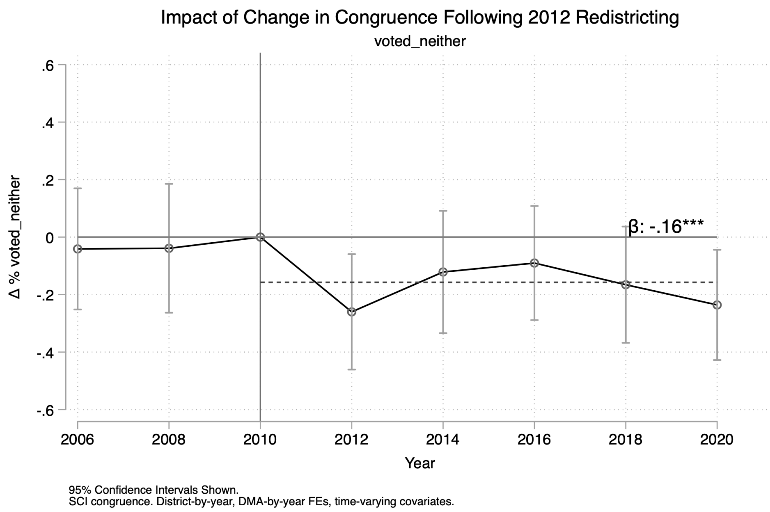

Figure 5: Dynamic Effects of Change in Congruence on Voter Choices Reported in CES - Roll-Off

The negative impact indicates that congruence reduces roll-off: if a county becomes more congruent, 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 20pp increase in congruence reduces rolloff by 0.8pp, or about one-fifth.

5.2 Campaign Contributions

To be added.

6 Robustness

I find in my main specification that congruence has a positive effect on voters' familiarity with their representatives. I explore the robustness of this finding by testing whether congruence impacts placebo outcome and by constructing an alternative measure of congruence using commuting flows.

6.1 Placebo Outcomes

I test whether congruence impacts voters' familiarity with their governor and senators: because these offices are elected through statewide elections, and consequently congressional district borders are not relevant for

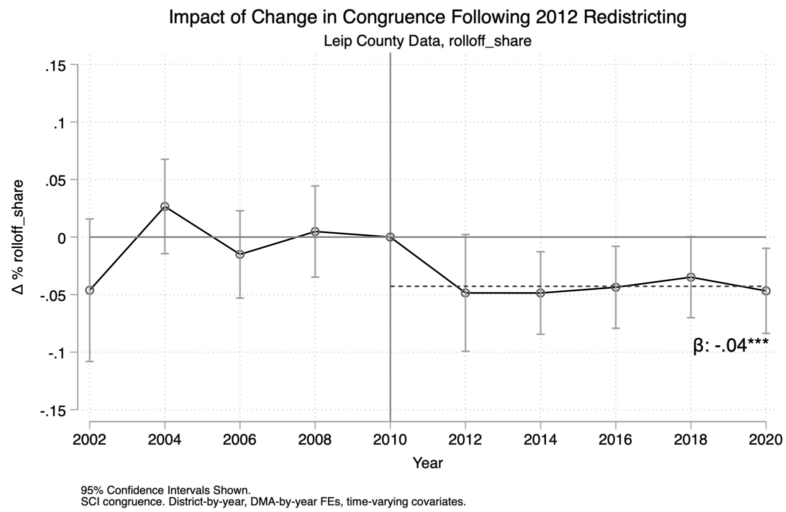

Figure 6: Dynamic Effects of Change in Congruence on Roll-Off

them, congruence should not impact them.[23] 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 15.

I find scant evidence for an impact of congruence on these nine outcomes. Results are reported in Appendix Tables 16-18.[24]

6.2 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

[23]These are not perfect placebos: a county that is more congruent with its congressional district may also be more likely to be more congruent with its state (consider a couple rural districts in a rural state). However, this would push me to find significant estimates of the impact of congruence on the placebo outcomes, even if congruence is appropriately capturing the alignment of voters' social networks with their congressional districts.

[24]While point estimates are significant at the 5% level in 2 cases and at the 10% level in 2 cases, over the 36 specifications ran this is not far beyond (though still more than) what is expected from random chance. One of these instances is when considering whether respondents select the correct party for senator 1, and for this estimates become much smaller and insignificant as soon as controls are added. The other three instances occur when considering whether respondents select a party for the governor. This is somewhat more concerning, but it should be noted that (1) estimates again become slightly smaller and less significant as controls are added, and (2) point estimates are negative, which is the opposite direction of the effects found when considering representatives.

average number of people that commute between two counties, and I construct commuting congruence 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),

$$\text{Commuting Congruence}_i = \frac{\sum_{j \in J} \text{Commuters}_{i,j}}{\sum_{k \in K} \text{Commuters}_{i,k}}$$

In Appendix Table 20, I report results for the effect of commuting congruence on voters' familiarity with their representatives. Estimates are of smaller magnitudes but otherwise are similar: across specifications, commuting congruence has a positive effect on measures of voters' familiarity with their representatives, with significance at at least the 5% level for all but one estimate. 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 congruence likely also reflect use of Facebook to share news about representatives.

6.3 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 congruence levels. Accordingly, we can identify the impact of congruence 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

$$y_{ct} = \alpha_c + \mu_{pt} + \beta \text{Congruence}_{c,t} + X'_{ct}\delta + \varepsilon_{ct}$$

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 other 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).

With the border pairs design, I find very similar results as in the redistricting design, except estimates on "Selected Party" become insignificant after adding DMA-by-year fixed effects.

7 Counterfactuals

Can congressional district maps be drawn to increase the share of informed voters? What congressional map characteristics are associated with higher congruence?

To answer these questions, I estimate parameters governing the model of information diffusion, and I then use the model to simulate the share of informed voters and changes in roll-off under different congressional maps.

7.1 Estimation

Recall that for a given county c, the steady-state share of informed voters is given by

$$\rho_c = \frac{\alpha\pi_c\tilde{\rho}_c}{\alpha\pi_c\tilde{\rho}_c + \delta}$$

Further, if we let $\lambda = \delta/\alpha$, we can write

$$\rho_c = \frac{\pi_c\tilde{\rho}_c}{\pi_c\tilde{\rho}_c + \lambda}$$

$\pi_c$can be calculated from the data, so the only unknown parameter determining the steady state share informed is λ. Consequently, I focus on estimating λ.

From the reduced form results, I have estimated the impact of a change in congruence on the change in share informed (i.e., $\hat{\beta} = \frac{\text{cov}(\Delta\pi,\Delta\rho)}{\text{var}(\Delta\pi)}$). Consequently, I can estimate λ by using indirect inference **[[add citation]]** to find the λ^* that gives a simulated $\tilde{\beta}$ that most closely matches $\hat{\beta}$. I do this as follows:

1. Draw a value of λ (say, from a discretized grid). Vectors π^{before} and π^{after} are observed. Draw a random initial vector of $\rho^{\text{before}}(0)$.

2. Given λ, π^{before}, and $\rho^{\text{before}}(0)$, simulate $\rho^{\text{before}}(1), ..., \rho^{\text{before}}(t)$, until convergence to the steady state is reached (recalling that the model has a unique stable steady state). This gives the simulated steady-

	(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					
Congruence	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					
Congruence	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					
Congruence	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, District x Year	County, Pair x Year, District x Year, DMA x Year	County, Pair x Year, District x Year, DMA x Year	County, Pair x Year, District x Year, DMA x Year	County, Pair x Year, District 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 5: Effect of Congruence on Voter Familiarity with Representative, within Border Pairs

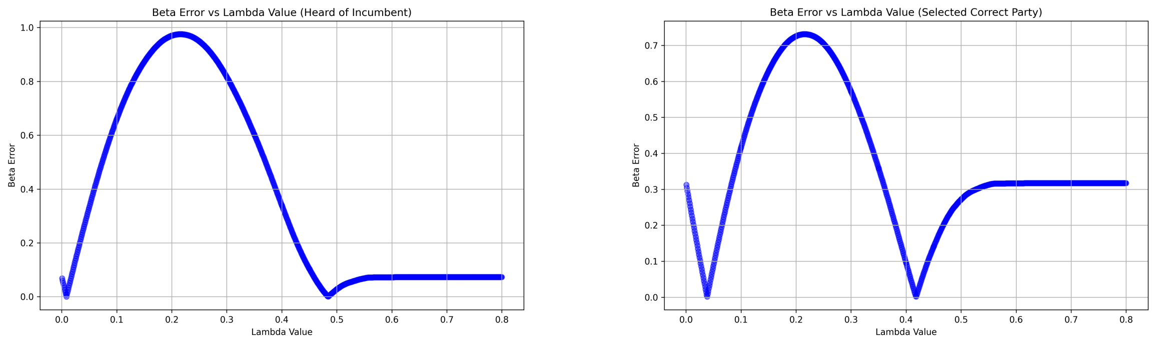

Figure 7: Difference Between $\hat{\beta}$ and $\tilde{\beta}$ for Given λ

state value of the share informed before the change in congruence.

3. Repeat the process for π^{after} and $\rho^{\text{after}}(0)$.

4. Calculate $\tilde{\beta}(\lambda) = \frac{\text{cov}(\Delta\pi,\Delta\rho)}{\text{var}(\Delta\pi)}$, where $\Delta\pi = \pi^{\text{after}} - \pi^{\text{before}}$ and $\Delta\rho = \rho^{\text{after}} - \rho^{\text{before}}$.

5. Repeat for many values of λ and choose the λ^* that minimizes the difference between $\tilde{\beta}(\lambda^*)$ and the $\hat{\beta}$ estimated in my regressions on the CES data.

I estimate the λ that corresponds to a model where the "share informed" corresponds to the share who have heard of their incumbent (call it λ^{Heard}), and I also estimate the λ when the "share informed" is defined as the share who correctly select the incumbent's party (call it λ^{Correct}).

Using a grid search, I find that $\lambda^{\text{Heard}} = 0.008$ and $\lambda^{\text{Correct}} = 0.039$. The below figures show the differences between the simulated and empirical β over a range of λ values, over 800 values of λ. Observe that in each case, there are two possible values of λ that minimize the error. This is because the steady-state can be written as quadratic in the vector of ρvalues.. However, the larger value of λ can be ruled out in each case, as it corresponds to a steady state where the share informed is very near zero. Indeed, the difference between $\hat{\beta}$ and $\tilde{\beta}$flattens above this because for large enough λ, all λs larger than it lead to convergence to the steady state where $\rho = 0$, and as such no change in congruence can lead to any change in ρ (because ρ is always 0).

7.2 Counterfactuals

With these estimates of λ in hand, I can then calculate the share of informed voters under counterfactual maps.

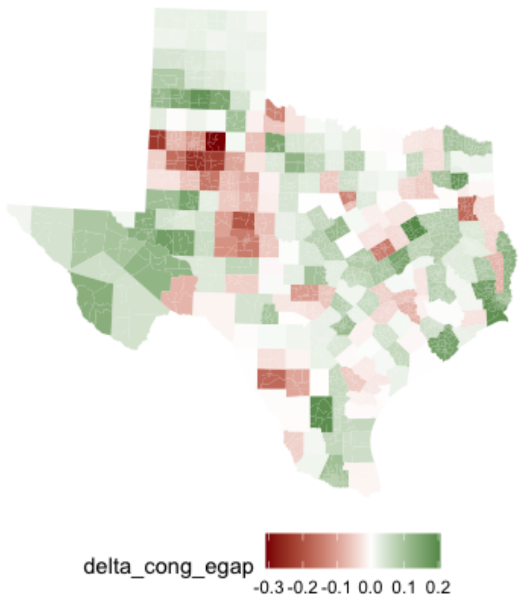

Figure 8: Change in (left) congruence, (right) share informed, from current map to efficiency gap minimizing map

McCartan et al (2022) simulate 5,000 congressional district maps for each of the 50 states. They use Monte Carlo simulation, and the maps are constrained to follow the given state's redistricting laws. They calculate commonly used measures of gerrymandering for each map. I use maps from their database to calculate congruence, shares informed, and abstention shares under counterfactual maps. For a given map, I calculate congruence. I then know the vector π for the map; λ has been estimated; and so I can simulate ρ under the map.

As an example, consider two congressional district maps for Texas: the current district boundaries, and the district boundaries among the 5,000 Texas maps in the McCartan et al (2022) database that minimize the efficiency gap **[[add efficiency gap cite]].**

The below figure shows the differences between the current map and the efficiency gap minimizing map, with the differences in congruence on the left and the differences in the share informed (in particular, the share who can correctly select their representative's party) on the right. Most notably, under the efficiency gap minimizing map, counties in urban areas such as Austin, Dallas, and Houston would become more congruent. In turn, they would also experience increases in the share of informed voters.

Under the current map, the simulated Texas-wide mean share informed is 79.98%; this increases the 81.03% under the efficiency gap minimizing map.

7.2.1 Abstention

The Swing Voter's Curse (Feddersen and Pesendorfer 1996) models the role of information in the decision to abstain. A goal of the Swing Voter's Curse is to explain roll-off -- among those who have already paid the cost of voting and turned up at the voting booth, why do some still skip down-ballot elections? Feddersen and Pesendorfer 1996argue that it can be rational for a voter to abstain from an election if they are uninformed. This is because, in a common interest setting, the uninformed voter is only pivotal when they are voting against the informed voters, i.e. when they vote for the "wrong" state of the world. The set-up in Feddersen and Pesendorfer 1996can be directly applied in this setting to simulate abstention rates.In large populations, Feddersen and Pesendorfer 1996give a simple approximation for the overall fraction who abstain, where p_i is the share of independents, p_0 and p_1 are the shares of 0 and 1 partisans respectively, and q is the share informed:

$$p_i(1-q) - |p_0 - p_1|$$

$p_i(1-q)$ is the share of uninformed independents, the only people who abstain in this model (everyone else votes either for their preference if they are partisan, or the correct state if they are an informed independent).However, this approximation only applies when $p_i(1-q) \geq |p_0 - p_1|$, i.e. when the partisan advantage of one party is not too large. Otherwise, all uninformed independents vote. The primary predictions are that abstention is decreasing as the share informed increases; abstention is increasing in the share of independents, and abstention is decreasing as the partisan advantage of one side increases. Similarly, the margin of victory can be approximated as the following (when he share of uninformed independents is greater than the partisan gap – otherwise, the margin of victory depends on the state of the world, which we do not observe):

$$MV = \frac{p_i q}{1 - p_i(1-q) + |p_0 - p_1|}$$

The primary predictions about the margin of victory are that it increases as the share of informed voters increases; and it increases as the share of independents increases.In particular, this means that an increase in the share informed results in **both** an increase in the margin of victory **and** a decrease in the abstention rateThis is because the uninformed independents employ a mixed strategy (everyone else has a pure strategy) in which they vote at the optimal rate to offset the partisan advantage.

To apply this model to my setting, I use CES self-reports of "strong partisan" vs. "independent" to approximate partisan shares at the county level. Accordingly, for a given county, I know every parameter required for the approximation, and can thus apply the approximation to simulate the abstention rate.

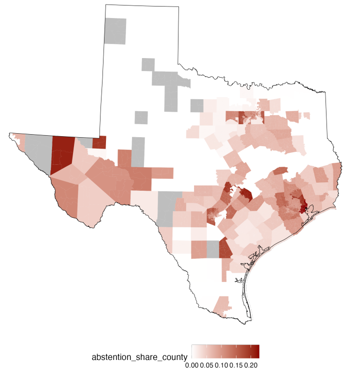

Figure 9: Simulated County Abstention Rates in Texas Under Current Map

I calculate the district-wide shares of partisans as well as the district-wide share informed, because these are the parameters that would matter for a given uninformed independent's voting decision (as opposed to the county-wide shares). But, the county-level share informed determines the abstention rate that would be observed in a given county.

Not all counties have a respondent in the CES in all years, especially for low-population counties. I calculate average partisan shares over 2012-2020 (the districting period) to adjust for this, but some very low population counties still aren't represented in any survey over this period. It should also be noted that the CES is not representative at the county level, only at the state level.

Using this method, I simulate abstention rates under the current district map in Texas, in the figure below. Darker red indicates higher abstention rates. Grey counties do not have any respondents in the CES over the period. White counties are generally in districts where the partisan gap is large enough that abstention is predicted to be zero.

8 Conclusion

Counties across the U.S. vary substantially in their social cohesiveness with their congressional district – their congruence. While people living in the average county share a representative with about half of the county's friends, this varies from 2% to 87%. I show that congruence increases voters' familiarity with their representative: when a county becomes more congruent due to redistricting, voters are more likely to recognize the name and know the party of their representative. I find similar results 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 congruence decrease rates of abstaining in the House election.

The impacts of congruence on electoral outcomes will inform policy by providing new dimensions on which to assess proposed district boundaries. This evidence is especially important as detailed social network data, like the SCI, has become publicly available for the first time in recent years, and consequently usable for gerrymandering.

References

Angelucci, C., Cagé, J., & Sinkinson, M. (2020). *Media Competition and News Diets* (Working Paper No. 26782). National Bureau of Economic Research. https://doi.org/10.3386/w26782

Auxier, B., & Anderson, M. (2021). 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

Bessone, P., Campante, F., Ferraz, C., & Souza, P. (2022). *Social Media and the Behavior of Politicians: Evidence from Facebook in Brazil* (tech. rep. w30306). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w30306

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

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

Di Tella, R., Gálvez, R. H., & Schargrodsky, E. (2021). *Does Social Media cause Polarization? Evidence from access to Twitter Echo Chambers during the 2019 Argentine Presidential Debate* (Working Paper No. 29458). National Bureau of Economic Research. https://doi.org/10.3386/w29458

Eckman, S. J. (2021). *Apportionment and Redistricting Process for the U.S. House of Representatives* (tech. rep. R45951). Congressional Research Service.

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

Enikolopov, R., Makarin, A., & Petrova, M. (2020). Social Media and Protest Participation: Evidence From Russia. *Econometrica*, *88*(4), 1479–1514. https://doi.org/10.3982/ECTA14281

Enikolopov, R., Petrova, M., & Zhuravskaya, E. (2011). Media and Political Persuasion: Evidence from Russia. *American Economic Review*, *101*(7), 3253–3285. https://doi.org/10.1257/aer.101.7.3253

Enke, B., Fisman, R., Freitas, L. M., & Sun, S. (2023). *Universalism and Political Representation: Evidence from the Field* (tech. rep. 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*.

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

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

Habermas, J., Burger, T., & Lawrence, F. (1992). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. The MIT press.

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

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

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

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

MIT Election Data and Science Lab. (2021). *County presidential election returns 2000-2020*. https://doi.org/10.7910/DVN/VOQCHQ

Prat, A., & Strömberg, D. (2005). *Commercial Television and Voter Information* (CEPR Discussion Paper No. 4989).

Quintelier, E., Stolle, D., & Harell, A. (2012). Politics in Peer Groups: Exploring the Causal Relationship between Network Diversity and Political Participation. *Political Research Quarterly*, *65*(4), 868–881. https://doi.org/10.1177/1065912911411099

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.

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

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

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). Partisan differences in social media use show up for some platforms, but not Facebook.

Whitaker, L. P. (2017). *Congressional Redistricting Law: Background and Recent Court Rulings* (tech. rep. R44798). Congressional Research Service.