

Audit Report: Workflow c52a4b6842704aac8c4258c71def4e29

Overview

- Source JSON: reports\workflow-report-c52a4b6842704aac8c4258c71def4e29.json
- Run ID: c52a4b6842704aac8c4258c71def4e29
- Papers input: papers\NetworksElectoralCompetition.pdf
- Started at: 2026-02-15T21:37:57.144102+00:00
- Finished at: 2026-02-15T21:43:52.685368+00:00
- Duration: 0:05:55.541266

Effective Configuration

- Chat model: gpt-5-nano
- Embedding model: text-embedding-3-large
- Top K: 10
- Chunk words / overlap: 350 / 75
- Batch size: 64
- Database URL configured: True

Step Outcomes

- prep: completed
- ingest: num_pdfs=1, num_papers=1
- enrich: openalex=0, citec=0
- econ_data: fetched
- agentic: completed
- index: indexed
- report_store: pending

Agentic Summary

- Status: completed
- Main question: What is the key contribution?
- Report question set: both
- Structured questions generated: 84
- Confidence mean/median: 0.29217296599402814 / 0.3038366165053296
- Confidence labels: low=77, medium=6, high=0

Final Answer

- Core contribution: A novel roll-off metric that quantifies abstention by top-of-ticket voters in House races (defined as the difference between top-ticket votes and House votes, as a share of top-ticket votes), linking to Swing Voter's Curse literature. This provides a direct measure of the information diffusion/participation gap within districts. (Source: page 18)

- Empirical strategy to identify diffusion within districts: A border-pairs design that compares county-pair deviations from their pair means, restricting to counties fully within a single district and collapsing to the county level to isolate within-district border variation. The key specification is $y_{ct} = \alpha_c + \alpha_{pt} + \beta_{c,t} + X_{ct} + \gamma_{ct}$, with $\beta_{c,t}$ the coefficient of interest; estimated with state-by-year fixed effects and no district-by-year fixed effects due to data limits. (Source: page 28)
- Data and identification leveraging donors: Campaign contributions to House candidates (Kuziemko et al. 2023; built from FEC via Bonica 2014) are geocoded to flag whether a contributor lives in the same congressional district as the candidate, enabling analysis of geographic proximity as a channel for information diffusion. (Source: page 18)
- Conceptual mechanism: Information diffusion within districts is shaped by district homophily, which summarizes network structure and governs the diffusion process and the equilibrium share of informed voters. (Source: page 39)
- Data construction and variables (appendix grounding): Appendix B documents vote-count construction (county-by-CD from multiple data sources) and variable descriptions for CES and voting outcomes (e.g., Heard of Representative, Selected Party, Voted in General Election); Appendix C highlights border-pairs results and district-homophily effects on self-reported voting outcomes (C2; C5, C11, C14). (Source: Appendix B, B.1–B.5; Appendix C, page references 54, 61, 64)

Sub-Answers

Sub-answer 1

- Question: ResponseTextConfig(format=ResponseFormatText(type='text'), verbosity='medium')
- Question tokens estimate: 19
- Answer:
- Roll-off measure definition and purpose
 - Definition: difference between top-of-ticket votes and House votes, as a share of top-of-ticket votes; captures share of voters who paid the voting cost for top-ticket but abstain from House (roll-off). Also linked to Swing Voter’s Curse literature. (Source: page 18, words 275–496)
- Campaign contributions data and identification strategy
 - Data: campaign contributions to House candidates from Kuziemko et al. 2023 (Oct), built from FEC data via Bonica 2014.

- Identification: geocode to flag whether contributor lives in the same congressional district as the candidate. (Source: page 18)
- Border-pairs empirical design and estimation
 - Design: compare county-pair deviations from their pair mean; restrict to counties fully within one district; collapse to county level to keep within-district border variation.
 - Key specification: $y_{ct} = \alpha_c + \alpha_{pt} + \gamma_{c,t} + X_{ct} + \epsilon_{ct}$ (11); $\gamma_{c,t}$ is the coefficient of interest; include state-by-year fixed effects; no district-by-year fixed effects due to data limits. (Source: page 28, 0–349)
- Conceptual framework: information diffusion within districts
 - Main idea: district homophily summarizes network structure, shaping the diffusion of information and the equilibrium share of informed voters. (Source: page 39, 0–327)
- Appendix B: data construction and variable descriptions
 - B.1: construction of vote-count measures (county-by-CD; sources include Harvard MIT data for various years, Dave Leip’s Atlas, Daily Kos)
 - B.2: variable descriptions for CES and voting outcomes (e.g., Heard of Representative, Selected Party, Voted in General Election, etc.)
 - B.3–B.5: descriptive statistics for CES outcomes and voting measures (Source: page 47–49)
- Appendix C: data and figures (highlights)
 - C2: results for border-pairs design; C5, C11, C14 figures illustrating district homophily effects on self-reported voting outcomes (Source: page 64; figure references on 54, 61, 64)

If you’d like, I can extract specific equations or variable definitions into a compact reference list. (Cited sources: page 18; page 28; page 39; Appendix B pages 47–49; Appendix C pages 54, 61, 64)

Structured Q&A Appendix

This section mirrors `agentic.report_questions` for audit traceability.

A01: What is the main research question of the paper?

- Category: A) Research question / contribution
- Confidence: low (0.3130893618247843)
- Retrieval method: local
- Evidence type: text
- Data source: Facebook friendship graph; CES survey data (Cooperative Election Study)

- Answer:

The main research question is how the alignment between social networks and political boundaries (district homophily) affects voters' political knowledge and voting behavior.

- Quote snippet:

How does the alignment between social networks and political boundaries impact voters' political knowledge and behavior?

- Citation anchors:
 - page=2, words=0-30, section=Introduction, note=States the main research question about the alignment of social networks and political boundaries and its impact on knowledge and behavior.

A02: What is the paper's primary contribution relative to the existing literature?

- Category: A) Research question / contribution
- Confidence: low (0.32002973932191364)
- Retrieval method: local
- Evidence type: textual_evidence_from_introduction
- Data source: Facebook friendship graph (county-level); alternative measure using commuting flows
- Answer:

The paper's primary contribution relative to the existing literature is twofold: (1) it provides causal estimates of how social network structure within districts influences voter knowledge, turnout, and campaign contributions at the scale of nearly the entire U.S.; and (2) it bridges the social-learning literature with political-geography models by showing that district borders—and the resulting alignment of social networks with districts—shape turnout through information diffusion.

- Quote snippet:

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.

- Citation anchors:
 - page=3, words=275-478, section=Introduction, note=Statement of two primary contributions, including large-scale causal estimates of social networks on voter knowledge.

A03: What is the central hypothesis being tested?

- Category: A) Research question / contribution

- Confidence: low (0.28644707638844313)
- Retrieval method: local
- Evidence type: empirical test of information diffusion via district homophily on voter knowledge and behavior
- Data source: Cooperative Election Study (CES)
- Assumption flag: True
- Assumption notes: Assumes district homophily drives information diffusion within districts, influencing knowledge and behavior; uses redistricting and district-border designs to identify effects.
- Answer:

District homophily—the alignment between social networks and congressional districts—is the central hypothesis being tested; the study investigates whether higher district homophily enhances diffusion of political information, increasing voters’ knowledge about their representatives and affecting voting-related behavior (turnout, candidate preferences).

- Quote snippet:
I study the impact of district homophily on voters’ knowledge and political behavior.
- Citation anchors:
 - page=6, words=0-349, section=Empirical Strategy and Networks Data, note=Defines district homophily and argues about how district boundaries influence information flow.
 - page=20, words=0-349, section=3.1 Voters’ Information, note=Explicitly states the study of district homophily’s impact on voters’ knowledge and behavior.
 - page=28, words=0-349, section=Conceptual Framework: Information Diffusion within Districts, note=Presents a diffusion model explaining how district homophily affects information diffusion.

A04: What are the main outcomes of interest (dependent variables)?

- Category: A) Research question / contribution
- Confidence: low (0.3439635433139609)
- Retrieval method: local
- Evidence type: textual_summary
- Data source: Cooperative Election Study (CES)
- Answer:

The main outcomes are voters’ knowledge about their representative (three binary knowledge measures), turnout (e.g., roll-off and turnout in top-of-ticket vs. House), and voting choices (preferences for incumbent, opponent, or neither, plus actual votes for incumbent/opponent/neither).

- Quote snippet:

I construct three binary variables to assess how familiar respondents are with their current representative.

- Citation anchors:
 - page=16, words=0-349, section=3 Outcomes Data / 3.1 Voters' Information, note=Defines the knowledge outcomes (three binary variables) used to assess familiarity with the current representative.
 - page=22, words=0-349, section=3.2 Voter Turnout and Voting Choices, note=Describes voting-related outcomes including preferences and actual votes; connects knowledge to turnout and vote choices.
 - page=47, words=0-207, section=Appendix B, note=B.1–B.3 describe variable construction for knowledge and voting outcomes.
 - page=49, words=0-276, section=Appendix B, note=Table B3 and related outcome variables for CES voting data (Voted for Incumbent/Opponent/Neither; Voted in General Election).

A05: What are the key treatment/exposure variables (independent variables)?

- Category: A) Research question / contribution
- Confidence: low (0.2478938536040581)
- Retrieval method: local
- Evidence type: appendix/methods description of exposure measures and table
- Data source: Cooperative Election Study (CES) data
- Table/Figure: Table C2
- Answer:

The key treatment/exposure variables are: (1) District Homophily — the change in district homophily experienced by a county between 2012 and 2013 (\bar{c}), used as the main exposure; and (2) Democratic Exposure — a county-level measure approximating the share of a county's social connections that voted Democratic, constructed by weighting each county's friends by the Democratic vote share and aggregating across connections.

- Quote snippet:

\bar{c} is the change in district homophily experienced by county c between 2012 and 2013.
- Citation anchors:
 - page=page 15, words= \bar{c} -between 2012 and 2013, section=main text, note=Definition of district homophily change as exposure.
 - page=page 15, words=I multiply the share of the county's friends-connected to, section=main text, note=Construction of Democratic Exposure measure.

- page=page 65, words=District Homophily-Dem. Exposure, section=appendix, note=Table showing District Homophily and Democratic Exposure as predictors.

A06: What setting/context does the paper study (country, market, period)?

- Category: A) Research question / contribution
- Confidence: low (0.3136748328106495)
- Retrieval method: local
- Evidence type: Setting description about the country and period
- Data source: U.S. county-year outcomes data across the continental United States (district homophily, information diffusion)
- Answer:

United States; the paper studies the redistricting that followed the 2010 Census and the 2012 U.S. congressional elections, using county-year data across the continental United States.

- Quote snippet:
the redistricting that followed the 2010 Census. Appendix Figure C4 provides a map of these changes.
- Citation anchors:
– page=15, words=0-349, section=Redistricting context, note=Cites the 2010 Census redistricting timeline and the 2012 elections.

A07: What is the main mechanism proposed by the authors?

- Category: A) Research question / contribution
- Confidence: low (0.2681809436308259)
- Retrieval method: local
- Evidence type: theoretical framework / diffusion model
- Data source: Facebook friendship graph (county-level) with an alternative commuting-flow measure
- Assumption flag: True
- Assumption notes: Assumes information diffusion is driven by district homophily; diffusion dynamics governed by (transmission) and network structure.
- Answer:

The main mechanism is diffusion of political information through social networks that is constrained by district homophily: information spreads mainly among people who live in the same district, because social ties cross district boundaries unevenly, shaping how voters become informed about their elected officials.

- Quote snippet:

Voters learn political information through their social networks (Lazarsfeld et al. 1944, Druckman et al. 2018).

- Citation anchors:
 - page=6, words=0-349, section=6 Conceptual Framework: Information Diffusion within Districts, note=Introduces diffusion mechanism, social networks, district homophily, and the ρ parameter for information transmission.
 - page=31, words=0-349, section=6.4 Timing, note=Describes the step-by-step diffusion process and how the probability ρ governs becoming informed when meeting an informed friend.
 - page=34, words=0-349, section=6 Conceptual Framework: Information Diffusion within Districts, note=Defines district homophily ρ and shows how it mechanically enters the diffusion process.

A08: What alternative mechanisms are discussed?

- Category: A) Research question / contribution
- Confidence: low (0.24243899016897702)
- Retrieval method: local
- Evidence type: textual_mention_of_alternative_mechanisms
- Answer:

Social pressure and recruitment.

- Quote snippet:

which also explores other mechanisms such as social pressure (Gerber et al. 2008, Sinclair et al. 2012) and recruitment (Klofstad 2007).
- Citation anchors:
 - page=4, words=0-349, section=Main text, note=Mentions alternative mechanisms in peer effects: social pressure and recruitment.

A09: What are the main policy implications claimed by the paper?

- Category: A) Research question / contribution
- Confidence: low (0.34575101077752535)
- Retrieval method: local
- Evidence type: Policy implication
- Data source: Facebook friendship graph (SCI) data used to construct district homophily.
- Answer:

The main policy implications are that social-network-informed districting (district homophily) can be used by policymakers to draw fairer congressional districts and improve information diffusion and turnout, but it also creates opportunities for partisan gerrymandering to exploit the data.

- Quote snippet:

enabling its use by policymakers to draw fairer districts, but also by partisan gerrymanderers who may seek to exploit it.

- Citation anchors:
 - page=6, words=0-349, section=Policy implications, note=Direct statement about policymakers using district homophily data to draw fairer districts and the risk of exploitation by gerrymanderers.

A10: What is the welfare interpretation (if any) of the results?

- Category: A) Research question / contribution
- Confidence: low (0.31141557800835046)
- Retrieval method: local
- Evidence type: empirical results (CES data; event studies)
- Data source: CES data
- Table/Figure: Figure 4 (event-study results on knowledge outcomes) and related figures
- Assumption flag: True
- Assumption notes: Interprets increased knowledge and lower abstention as welfare gains from better information and participation; no direct welfare metric is provided.
- Answer:

The welfare interpretation is that district homophily yields positive welfare effects for voters within a district by increasing their knowledge about their representatives and reducing abstention in House elections; it also shifts political donations toward in-district candidates, reflecting greater emphasis on district-level representation. However, the interpretation hinges on assuming that more information and higher turnout constitute welfare gains and that the donation shift does not offset welfare in other districts.

- Quote snippet:

district homophily increases voters' knowledge about their representatives, and accordingly decreases abstention in House elections.
- Citation anchors:
 - page=19, words=0-349, section=Section 4: Results, note=Summary of main results: district homophily increases voters' knowledge and decreases abstention; shifts donations to same-district candidates.
 - page=19, words=275-406, section=Section 4: Results, note=Reinforces that district homophily affects knowledge and donations.

A11: What are the main limitations acknowledged by the authors?

- Category: A) Research question / contribution
- Confidence: low (0.26469760725311464)
- Retrieval method: local
- Evidence type: limitations and modeling/identification assumptions

- Assumption flag: **True**
- Assumption notes: Assumptions 1–4 governing the identification and stability of the mapping Q s: (1) Limited fragmentation (dominant district exists), (2) Simple boundaries (counties intersect at most two districts), (3) Non-trivial out-of-state connections ($(\Pi_s) < 1$), (4) Network non-degeneracy (rows of Π s are linearly independent). These restrict generalizability and identifiability to near-observed maps and within-state configurations.
- Answer:

The authors acknowledge several key limitations: (i) diffusion parameters (e.g., μ , ν , ρ) are not yet estimated and are left for future work; (ii) Identification relies on strong, but restrictive, assumptions about the map topology (limited fragmentation, dominant districts, and simple boundaries) and on non-degenerate, partially open-state networks; (iii) the analysis is constrained to within-state counties and to local perturbations near the observed map; (iv) despite these limitations, some conditions (e.g., limited fragmentation) hold for the vast majority of counties, but the framework may not generalize beyond them.

- Quote snippet:
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.
- Citation anchors:
 - page=34, words=0-349, section=main text, note=Future work to estimate diffusion process parameters; parameter estimation is an acknowledged limitation.
 - page=44, words=275-491, section=appendix, note=Assumptions 1-4: Limited fragmentation, Simple boundaries, Non-trivial out-of-state connections, Network non-degeneracy; these are key limitations for identifiability.
 - page=45, words=0-349, section=appendix, note=Discussion of within-state focus and open-system leakage; emphasizes constraints used in modeling.
 - page=46, words=0-349, section=appendix, note=Technical conditions (I– Π s invertibility, local injectivity) that underpin identifiability; these reflect methodological limitations.

A12: What does the paper claim is novel about its data or identification?

- Category: A) Research question / contribution
- Confidence: low (0.31288778489299973)
- Retrieval method: local
- Evidence type: claims of novelty in data scope and identification strategy

- Data source: Facebook friendship graph (county-level) with an alternative commuting-flow measure
- Answer:

The paper claims novelty in using national-scale data on social networks (Facebook friendship ties) to measure district homophily and in employing a plausibly causal identification that leverages changes in district homophily over time due to redistricting, enabling causal estimates of how district alignment affects voter knowledge and behavior.

- Quote snippet:

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.

- Citation anchors:
 - page=3, words=0-349, section=Introduction, note=States the paper's contribution and the scale of data (nearly the entire U.S.)
 - page=3, words=275-478, section=Introduction, note=Explicitly describes two primary contributions, including causal estimates at national scale
 - page=4, words=275-492, section=Data/Methods, note=Describes use of national data on social networks and alternative commuting-flow measure
 - page=6, words=0-349, section=Identification/Strategy, note=Notes plausibly causal identification of impacts leveraging changes in district homophily over time due to redistricting

B01: What is the identification strategy (in one sentence)?

- Category: B) Identification strategy / causal design
- Confidence: low (0.34214973075803773)
- Retrieval method: local
- Evidence type: Appendix derivations and conceptual framework supporting local identifiability via an invertible mapping
- Assumption flag: True
- Assumption notes: Assumptions include out-of-state leakage ($\Pi_s < 1$), simple boundaries, limited fragmentation, and network non-degeneracy, which ensure local injectivity of the district homophily mapping.
- Answer:

Identification is achieved through a local, invertible mapping from the county-district assignment to district homophily (i.e., an invertible Jacobian), so redistricting-driven variation uniquely identifies the diffusion of information within districts.

- Quote snippet:

Thus, $Av = 0 \Rightarrow v = 0$. Finally, since $\det(DF(Q))$ is a continuous function of Q , there exists an open neighborhood around the clean map where the Jacobian remains invertible.

- Citation anchors:
 - page=43, words=0-349, section=A.2 Local Uniqueness of District Homophily, note=Local mapping / Jacobian invertibility supports identification.
 - page=45, words=0-349, section=Appendix sections 3-4: Non-trivial out-of-state connections; Network non-degeneracy, note=Assumptions yield $\ker(A)=0$, enabling identification.
 - page=46, words=0-349, section=Appendix, note=Conclusion: open neighborhood where Jacobian remains invertible; local identification.

B02: Is the design experimental, quasi-experimental, or observational?

- Category: B) Identification strategy / causal design
- Confidence: low (0.29602763018678624)
- Retrieval method: local
- Evidence type: Quasi-experimental natural experiment with event-study design around redistricting; uses a border-pairs design as robustness check.
- Data source: Facebook friendship graph; Cooperative Election Study (CES); Dave Leip's Election Atlas
- Assumption flag: True
- Assumption notes: Assumes redistricting exogenously changes district homophily and is not driven by pre-treatment trends; event-study design tests pre-trends and supports causal interpretation.
- Answer:

Quasi-experimental (natural experiment) using a redistricting event and an event-study design to identify causal effects.

- Quote snippet:

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.
- Citation anchors:
 - page=15, words=0-349, section=Methods: Event Studies / Redistricting, note=Describes redistricting as the treatment and uses an event-study design for causal identification.
 - page=28, words=0-349, section=Identification: Border-pairs design, note=Details the border-pairs design as an alternative quasi-experimental identification strategy.
 - page=20, words=0-349, section=Results: Voter knowledge and

behavior (event studies), note=Presents event-study evidence that changes in district homophily due to redistricting affect outcomes.

B03: What is the source of exogenous variation used for identification?

- Category: B) Identification strategy / causal design
- Confidence: low (0.3451500942468334)
- Retrieval method: local
- Evidence type: Identification strategy description (use of redistricting-induced variation in district homophily for causal identification).
- Data source: Appendix Figure C4; redistricting data after the 2010 Census.
- Assumption flag: True
- Assumption notes: Assumes the redistricting after the 2010 Census is exogenous to county-level outcomes and provides the variation in district homophily used for identification.
- Answer:

The source of exogenous variation is the redistricting that followed the 2010 Census, which changed district homophily across counties and is exploited via an event-study design.

- Quote snippet:
on the redistricting that followed the 2010 Census.
- Citation anchors:
 - page=15, words=0-349, section=Identification strategy, note=Describes focusing on a single redistricting event following the 2010 Census to identify effects via changes in district homophily.

B04: What is the treatment definition and timing?

- Category: B) Identification strategy / causal design
- Confidence: low (0.11053292476972407)
- Retrieval method: local
- Evidence type: textual evidence from provided context
- Assumption flag: True
- Assumption notes: Assumes 2012 as the last year before treatment for outcomes related to the current representative; uses 2012 as the last pre-treatment year in the event-study specification.
- Answer:

The treatment is the change in district homophily experienced by a county due to redistricting after the 2010 Census, defined as the change in district homophily between 2012 and 2013. The timing is tied to the 2012 redistricting for the November 2012 elections (with new representatives taking office in

January 2013); the analysis treats 2012 as the last year before treatment for outcomes related to the current representative, while outcomes tied to the next representative may treat 2010/2011 as the pre-treatment period.

- Quote snippet:

The Census was conducted in April 2010, and states needed to draw new congressional district borders in time for the November 2012 elections.

- Citation anchors:
 - page=15, words=0-40, section=main, note=Definition of treatment as the change in district homophily due to redistricting; timing anchored to 2012 election.
 - page=15, words=41-110, section=main, note=States needed to draw new borders for the November 2012 elections; representatives first elected under new borders assumed office in January 2013; discussion of last year before treatment.

B05: What is the control/comparison group definition?

- Category: B) Identification strategy / causal design
- Confidence: low (0.2139641910281757)
- Retrieval method: local
- Evidence type: methodology description
- Answer:

The control/comparison group is counties that do not experience a change in district homophily ($\bar{c} = 0$) during the redistricting window; treated units are counties with a change in district homophily due to redistricting, with comparison made via an event-study around the 2010 Census redistricting.

- Quote snippet:

\bar{c} is the change in district homophily experienced by county c between 2012 and 2013

- Citation anchors:
 - page=14, words= \bar{c} is the change in district homophily experienced by county c between 2012 and 2013- \bar{c} is the change in district homophily experienced by county c between 2012 and 2013, section=2.2 Redistricting, note=Defines the treatment variable that drives the event-study design.
 - page=14, words=I use an event study design-I use an event study design, section=2.2 Redistricting, note=States the identification strategy as an event-study around redistricting.
 - page=15, words=last year before the treatment-last year before the treatment, section=2.2 Redistricting, note=Describes the pre-treatment period relative to the redistricting event.

B06: What is the estimating equation / baseline regression specification?

- Category: B) Identification strategy / causal design
- Confidence: low (0.3290107117845091)
- Retrieval method: local
- Evidence type: Equation (11) in the border pairs design section
- Assumption flag: True
- Assumption notes: Counties are restricted to those fully within one district; pair-by-year fixed effects are used as part of the border-pairs specification.
- Answer:

Baseline regression (border pairs design): $y_{ct} = \alpha_c + \mu_{pt} + \gamma_{c,t} + X_{ct} + \epsilon_{ct}$

- Quote snippet:

$$y_{ct} = \alpha_c + \mu_{pt} + \gamma_{c,t} + X_{ct} + \epsilon_{ct}$$

- Citation anchors:
 - page=28, words=0-349, section=Equation (11) specification, note=Baseline regression for border pairs: $y_{ct} = \alpha_c + \mu_{pt} + \gamma_{c,t} + X_{ct} + \epsilon_{ct}$

B07: What fixed effects are included (unit, time, two-way, higher dimensional)?

- Category: B) Identification strategy / causal design
- Confidence: low (0.34033278422637275)
- Retrieval method: local
- Evidence type: text
- Answer:

Fixed effects include year (time) fixed effects and several high-dimensional unit-by-time effects: county-by-year fixed effects, pair-by-year fixed effects; plus higher-dimensional effects by state-by-year and DMA-by-year. District-by-year fixed effects are not used due to data limitations.

- Quote snippet:

To address this concern, I use the boundaries of the Nielsen Designated Market Areas and include DMA-by-year fixed effects.

- Citation anchors:
 - page=28, words=0-349, section=Methods: Fixed Effects, note=States-by-year fixed effects are included; district-by-year fixed effects are not due to data limits.
 - page=15, words=0-349, section=Methods: Fixed Effects, note=DMA-by-year fixed effects are used to address media-market confounds;

- year fixed effects τ are part of the model.
- page=65, words=0-288, section=Appendix Table C2, note=Table includes County-Year fixed effects and Pair-by-Year fixed effects; also indicates DMA-by-Year and State-by-Year fixed effects in other columns.

B08: What standard errors are used (robust, clustered; at what level)?

- Category: B) Identification strategy / causal design
- Confidence: low (0.32153160926730573)
- Retrieval method: local
- Evidence type: text and table notes indicating county-level clustering
- Answer:

Clustered standard errors at the county level.

- Quote snippet:
Errors ict are clustered at the county level.
- Citation anchors:
 - page=15, words=0-349, section=Main text (Eq. 8), note=Errors ict are clustered at the county level.
 - page=65, words=0-288, section=section appendix, note=Standard errors clustered at the county level in parentheses.

B09: What is the key identifying assumption (parallel trends, exclusion restriction, ignorability)?

- Category: B) Identification strategy / causal design
- Confidence: low (0.3383880309349816)
- Retrieval method: local
- Evidence type: Identifying assumption for causal interpretation in the event-study design; relies on parallel trends with a pre-trends check.
- Assumption flag: True
- Assumption notes: The causal claim relies on a parallel-trends assumption, validated conceptually by a pre-trends visual check in the event-study framework.
- Answer:

Parallel trends (the pre-treatment trends are assumed to be the same across units in the absence of the redistricting-induced change in district homophily).

- Quote snippet:
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.

- Citation anchors:
 - page=15, words=0-90, note=Mentions event-study design and a visual test of pre-trends in changes in district homophily.

B10: What evidence is provided to support the identifying assumption?

- Category: B) Identification strategy / causal design
- Confidence: low (0.26433262000983937)
- Retrieval method: local
- Evidence type: Theoretical proposition with Jacobian analysis
- Assumption flag: True
- Assumption notes: The result relies on 1) Limited fragmentation, 2) Simple boundaries, 3) Non-trivial out-of-state connections, and 4) Network non-degeneracy. Under these, there exists $\bar{s} > 0$ so that $s < \bar{s}$ implies local injectivity of $Q_s \rightarrow \bar{s}$, i.e., distinct nearby maps yield distinct district homophily.
- Answer:

The identifying assumption is supported by a formal theoretical result in the Appendix: Proposition 1 (State-conditional local injectivity of district homophily). It shows that, under four assumptions (Limited fragmentation, Simple boundaries, Non-trivial out-of-state connections, and Network non-degeneracy), there exists an $\bar{s} > 0$ such that if $s < \bar{s}$, the mapping from county-district assignments Q_s to district homophily \bar{s} is locally injective at Q_0 s. In words, small changes in the assignment lead to distinct changes in district homophily, so district homophily identifies the underlying assignment locally.

- Quote snippet:

Proposition 1 (State-conditional local injectivity of district homophily). Fix a state s .
- Citation anchors:
 - page=44, words=0-349, section=section appendix, note=Proposition 1: state-conditional local injectivity of district homophily; statement and assumptions.
 - page=45, words=0-349, section=section appendix, note=Outline of assumptions and proof structure supporting injectivity.
 - page=46, words=0-349, section=section appendix, note=Detail of the Jacobian-based argument showing local injectivity under the stated assumptions.

B11: Are there event-study or pre-trend tests? What do they show?

- Category: B) Identification strategy / causal design
- Confidence: low (0.327939027982952)

- Retrieval method: `local`
- Evidence type: `event-study and placebo/pre-trend tests`
- Data source: Cooperative Election Study (CES) data
- Answer:

Yes. The paper conducts event-study analyses around redistricting and pre-trend/placebo checks. The event-study results show that changes in district homophily most strongly affect voter knowledge in the first post-redistricting survey (2014) and remain relatively stable over time; placebo tests show no significant predictions for placebo outcomes, supporting the validity of pre-trend assumptions.

- Quote snippet:

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.

- Citation anchors:
 - page=page 20 words 0-349, words=0-349, section=Event-study results, note=The event studies show strongest effect in 2014 and stability over time.
 - page=page 55 words 0-152 section appendix, words=0-152, section=Appendix C.2 Placebo Outcomes, note=Placebo tests show district homophily does not predict placebo outcomes.
 - page=page 15 words 0-349, words=0-349, section=Pre-trends mention, note=Mentions pre-trends in changes in district homophily as part of the analysis.

B12: What instruments are used (if IV)? Define instrument and first stage.

- Category: B) Identification strategy / causal design
- Confidence: low (0.1320711091565386)
- Retrieval method: `local`
- Evidence type: No instrumental variables described; model uses exogenous information rate (c,d) and transmission
- Data source: Not applicable (no IVs used; model uses exogenous information rate (c,d) and transmission)
- Answer:

Not applicable: the context does not use instrumental variables. It models information diffusion with exogenous news arrival at rate (c,d) and a transmission probability ; no IV or first-stage is defined.

- Quote snippet:

Agents can learn news exogenously (such as from the media) at a rate (c,d) , which is allowed to vary by type.

- Citation anchors:
 - page=30, words=0-349, section=6.3 Information Sharing Process, note=Defines exogenous information arrival rate (c,d) and transmission dynamics
 - page=31, words=0-349, section=6.5 Individual Transition Probabilities, note=Details transition probabilities and information transmission; no instrument mention
 - page=32, words=0-349, section=6.5, note=Equations related to information transmission; no instrument description

B13: What is the first-stage strength (F-stat, partial R^2 , relevance evidence)?

- Category: B) Identification strategy / causal design
- Confidence: low (0.30078694597347216)
- Retrieval method: local
- Evidence type: First-stage regression result showing the relevance of district homophily to voter familiarity (Heard of Incumbent) with significance and model fit
- Data source: CES Data
- Table/Figure: Table C2
- Answer:

The first-stage is strong: district homophily significantly predicts the endogenous exposure (Heard of Incumbent). The coefficient is 0.254 with a standard error of 0.080 ($p = 0.002$). Across specifications, R-squared ranges from about 0.606 to 0.720, and the implied F-stat (approx. t^2) is around 10, indicating a strong relation.

- Quote snippet:

Heard of Incumbent District Homophily 0.254*** 0.254*** (0.080)
[0.002]
- Citation anchors:
 - page=65, words=0-288, section=appendix, note=Table C2: Effect of District Homophily on Voter Familiarity with Representative, within Border Pairs; District Homophily coefficient 0.254*** (SE 0.080); p-value 0.002; R2 values 0.606–0.720 across specifications.

B14: If RDD: what is the running variable and cutoff? bandwidth choice?

- Category: B) Identification strategy / causal design
- Confidence: low (0.29829687505789665)
- Retrieval method: local

- Evidence type: `information not found in context`
- Answer:

Not specified in the provided excerpts; the text discusses redistricting designs and border-pairs comparisons, but it does not define an RDD running variable, cutoff, or bandwidth.

- Quote snippet:
The specification for this design is $y_{ct} = \alpha + \mu_{pt} + \gamma_{c,t} + X_{ct}$
 $+ \epsilon_{ct}$ (11)
- Citation anchors:
 - page=28, words=0-349, section=border pairs design (methods), note=Describes border-pairs specification and does not mention an RDD running variable/cutoff or bandwidth.

B15: If DiD: what is the timing variation (staggered adoption)? estimator used?

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.3247377337456784)`
- Retrieval method: `local`
- Evidence type: `text`
- Answer:

An event-study style differences-in-differences estimator with relative-time indicators around the treatment. The specification uses year fixed effects and a set of relative-time dummies $I(\tau = t)$ multiplied by the change in district homophily (γ), capturing staggered timing of adoption.

- Quote snippet:
The event studies thus take the following form: $\tau = 2010 \text{ X } \tau = 2022$
 $X y_{ict} = \alpha + \gamma I(\tau = t) + \gamma I(\tau = t) + X_{ict} + Z_{ict} + \epsilon_{ict}$
- Citation anchors:
 - page=15, words=0-60, section=Methods / Event-study specification, note=Presents the event-study form: $y_{ict} = \alpha + \gamma I(\tau = t) + \dots$; discusses fixed effects options.
 - page=19, words=275-406, section=Methods / Event studies, note=Notes that event studies reflect timing of knowledge and pre-/post-election context; supports event-study approach.
 - page=28, words=0-349, section=Design / Border-pairs and fixed effects, note=Describes a DiD-like design and fixed effects, illustrating variations on timing controls (district-by-year FE mentioned as an option).

C01: What dataset(s) are used? (name sources explicitly)

- Category: C) Data, sample, and measurement
- Confidence: low (0.3257372819857322)
- Retrieval method: local
- Evidence type: dataset_sources
- Data source: CES (Cooperative Election Study; formerly CCES); Dave Leip's Election Atlas; Harvard Election Data Archive; MIT Election Data and Science Lab; Daily Kos
- Answer:

CES (Cooperative Election Study; formerly CCES), Dave Leip's Election Atlas, Harvard Election Data Archive, MIT Election Data and Science Lab, and Daily Kos.

- Quote snippet:

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.

- Citation anchors:
 - page=16, words=0-349, section=3.1 Voters' Information, note=Mentions the Cooperative Election Study (CES) as the data source for voter information.
 - page=17, words=275-542, section=3.2 Voter Turnout, note=References data from Dave Leip's Election Atlas for turnout analyses.
 - page=47, words=0-207, section=Appendix B Data Descriptions, note=Lists precinct-level data sources (Harvard Election Data Archive; MIT Election Data and Science Lab) plus Dave Leip's Election Atlas and Daily Kos.

C02: What is the unit of observation (individual, household, firm, county, transaction, product)?

- Category: C) Data, sample, and measurement
- Confidence: low (0.33683153978314195)
- Retrieval method: local
- Evidence type: text
- Data source: CES survey data
- Answer:

Individual (respondents in the CES survey).

- Quote snippet:

In these 48 states and across all 17 years of the CES, there are 612,085 respondents (552,307 excluding 2006, 2007, and 2009).

- Citation anchors:
 - page=17, words=275-542, section=CES data description, note=States the CES-based sample includes 612,085 respondents.

C03: What is the sample period and geographic coverage?

- Category: C) Data, sample, and measurement
- Confidence: low (0.24987723552890045)
- Retrieval method: local
- Evidence type: Descriptive data coverage from CES dataset
- Data source: Congressional Election Study (CES)
- Answer:

Sample period: across 17 years of the Congressional Election Study (CES); geographic coverage: the 48 contiguous states (excluding Alaska, Hawaii, D.C., and territories).

- Quote snippet:

In these 48 states and across all 17 years of the CES
- Citation anchors:
 - page=page 13, words=0-349, section=2.1.5 Summary Statistics and Predictors of District Homophily, note=Describes geographic coverage to continental 48 states and full period
 - page=page 15, words=0-349, section=3.2 Voter Turnout, note=Notes sample across all 17 years of the CES

C04: What are the sample restrictions / inclusion criteria?

- Category: C) Data, sample, and measurement
- Confidence: low (0.25810184105962247)
- Retrieval method: local
- Evidence type: context
- Answer:

Sample restrictions / inclusion criteria include: (1) Respondents are drawn from the 48 contiguous U.S. states. (2) Exclude respondents with missing responses to the candidate party recognition question (<2% per year). (3) Exclude respondents who did not answer the relevant demographic questions used as controls. (4) Exclude in 2006–2007 respondents assigned to counties not in their state of residence. (5) Exclude 2020 North Carolina respondents who were incorrectly assigned to districts.

- Quote snippet:

I only include respondents in the 48 contiguous states. I exclude missing responses to the candidate party recognition question (<2%). I also exclude respondents in 2006 and 2007 assigned to counties that are not in their state of residence.

- Citation anchors:
 - page=17, words=0-349, section=main text, note=I only include respondents in the 48 contiguous states; excludes 2006/2007 county-state mismatches.
 - page=17, words=0-349, section=main text, note=Exclude missing responses to the candidate party recognition question (<2%).
 - page=17, words=275-542, section=main text, note=Exclude respondents who did not answer the relevant demographic questions.
 - page=19, words=275-406, section=main text, note=In the 2020 survey, 925 NC respondents were assigned to incorrect districts; excluded.

C05: What is the sample size (N) in the main analysis?

- Category: C) Data, sample, and measurement
- Confidence: low (0.2534666829948621)
- Retrieval method: local
- Evidence type: statistic
- Data source: CES data (Congressional Election Study)
- Answer:

612,085 respondents.

- Quote snippet:

612,085 respondents (552,307 excluding 2006, 2007, and 2009).
- Citation anchors:
 - page=page 17, words=275-542, section=CES data (main analysis), note=Mentions 612,085 respondents across 17 years of the CES.

C06: How is the key outcome measured? Any transformations (logs, z-scores, indices)?

- Category: C) Data, sample, and measurement
- Confidence: low (0.3102709813360862)
- Retrieval method: local
- Evidence type: measurement_description
- Data source: CES survey data (Congressional Election Survey) and county-level voting data (Dave Leip's Election Atlas)
- Answer:

The key outcomes are binary indicators (0/1). No log or z-score transformations are described; effects are interpreted as changes in probability from a 0-to-1 change in district homophily.

- Quote snippet:

District homophily is measured on a scale from 0 to 1, and outcome variables are binary.

- Citation anchors:
 - page=19, words=District-binary., section=3.2 Voter Turnout, note=Outcomes are binary and interpreted as probability changes; district homophily is 0–1.

C07: How is treatment/exposure measured? Any constructed variables?

- Category: C) Data, sample, and measurement
- Confidence: low (0.3035829814992705)
- Retrieval method: local
- Evidence type: Methodology description from the study
- Data source: CES survey data and Dave Leip’s Election Atlas (with Nielsen DMA boundaries for fixed effects analysis)
- Table/Figure: Appendix Figure C4
- Assumption flag: True
- Assumption notes: Assumes friendships are uniformly distributed within a county; uses $(c,d)(c,d) = c,c \times q(c,d)$ and population shares to aggregate.
- Answer:

Exposure (treatment) is measured primarily with district homophily, operationalized as the change in district homophily experienced by each county (Δ^c) between 2012 and 2013 after redistricting. District homophily itself is on a 0–1 scale. A constructed exposure proxy aggregates county social ties by weighting each connected county’s share of friends by that county’s Democratic vote share, summing across counties to approximate the share of a county’s friends who voted Democratic. The analysis also uses district-by-year and DMA-by-year fixed effects to control for election- and media-market-level exposure.

- Quote snippet:

District homophily is measured on a scale from 0 to 1, and outcome variables are binary.
- Citation anchors:
 - page=15, words=0-40, section=Main text, note=Event-study form and Δ^c as change in district homophily
 - page=15, words=275-350, section=Data and Controls, note=DMA-by-year fixed effects; Democratic exposure construction
 - page=17, words=0-30, section=Measurement, note=District homophily on a 0–1 scale
 - page=33, words=0-60, section=Appendix C, note=Aggregation to county-level; network probabilities and q
 - page=65, words=0-40, section=Appendix C2, note=Table with District Homophily effects on voter familiarity

C08: Are there key covariates/controls? Which ones are always included?

- Category: C) Data, sample, and measurement
- Confidence: low (0.3496017264877621)
- Retrieval method: local
- Evidence type: table note
- Answer:

Yes. The analysis uses key covariates/controls described in Table C2: (1) Individual demographic controls (gender, race, education, age categories, and whether the respondent is affiliated with the same party as their representative); and (2) County-by-year controls (population and shares by race, age categories, gender, and county urban population share). The six specifications vary by which fixed effects and control sets are included (e.g., county-pair, state-by-year, DMA-by-year FEs, and whether demographic/exposure controls are included). Consequently, there is no single covariate that is guaranteed to be included in every specification; the models differ in the fixed effects and the inclusion of the demographic/ county-year control sets.

- Quote snippet:

Individual controls include gender, race, education, age categories, and whether the respondent is affiliated with the same party as their representative.

- Citation anchors:
 - page=65, words=0-24, section=Table C2, note=Individual controls include gender, race, education, age categories, and whether the respondent is affiliated with the same party as their representative.
 - page=65, words=25-60, section=Table C2, note=County-by-year controls include population and shares by race, age categories, gender, and county urban population share.

C09: How are missing data handled (dropping, imputation, weighting)?

- Category: C) Data, sample, and measurement
- Confidence: low (0.3162877908919168)
- Retrieval method: local
- Evidence type: Methods/data handling description
- Data source: Cooperative Election Study (CES); Dave Leip's Election Atlas (for turnout/outcomes data) as referenced in the methods
- Answer:

Missing data are handled by dropping/excluding cases with missing key responses (e.g., missing candidate party recognition, missing demographic questions, and respondents misassigned to counties/districts). The analysis uses the

CES's cumulative weights to re-weight observations so samples are comparable across years, and these weights do not account for attrition.

- Quote snippet:
I use the CES's cumulative weights, which re-weight observations to make sample sizes comparable across years (see Kuriwaki 2018).
- Citation anchors:
 - page=16, words=0-349, section=3 Outcomes Data, note=Describes use of CES cumulative weights to adjust sample sizes across years.
 - page=17, words=0-349, section=3 Outcomes Data, note=Specifies dropping missing responses to candidate party recognition (<2%), excluding respondents with missing demographic questions, and excluding misallocated district respondents; mentions attrition and weighting context.

C10: Are weights used (survey weights, propensity weights)? How?

- Category: C) Data, sample, and measurement
- Confidence: medium (0.3629425432808147)
- Retrieval method: local
- Evidence type: methodology description
- Data source: CES (Congressional Election Study)
- Answer:

Yes. The study uses weights, but they do not account for attrition between the pre- and post-surveys; the specific weighting scheme (e.g., survey vs propensity weights) is not specified in the provided context.

- Quote snippet:
Weights do not account for this attrition.
- Citation anchors:
 - page=19, words=275-406, section=main text, note=Mentions weights and attrition in CES data

C11: Are data linked/merged across sources? How is linkage performed?

- Category: C) Data, sample, and measurement
- Confidence: low (0.30301259861776314)
- Retrieval method: local
- Evidence type: methodology description
- Data source: Facebook Social Connectedness Index (SCI) for U.S. counties; county-level outcomes data for vote counts and campaign contributions; Zip-code level data via SCI as robustness check (CES zip codes mentioned).
- Answer:

Yes. Data are linked across sources primarily by geography; county-level social-network data from the Facebook SCI are matched to county-level outcomes data (vote counts and campaign contributions). For districts, district homophily is constructed by population-weighted averages across counties that intersect districts. Zip-code level data are used for robustness checks, but the main linkage relies on county-level networks.

- Quote snippet:
I focus on county-county pairs to facilitate matching to county-level outcomes data for vote counts and campaign contributions.
- Citation anchors:
 - page=8, words=0-349, section=2.1.2 Proxy for Social Networks: Facebook Social Connectedness Index, note=Linkage of SCI to county-level outcomes data; county-level matching.
 - page=27, words=0-349, section=5.3 Zip-Code-Level Social Network Data, note=Zip-code level data used; main analysis uses county-level networks; data privacy restrictions.

C12: What summary statistics are reported for main variables?

- Category: C) Data, sample, and measurement
- Confidence: medium (0.3756314526873376)
- Retrieval method: local
- Evidence type: table
- Data source: CES dataset (pre- and post-surveys; Tables B3 and B5)
- Answer:

They report, for each main CES variable, the number of observations, the mean percentage, and the standard deviation in percentage points (Observed as: Heard of Representative, Selected Party, Selected Correct Party, Prefer Incumbent, Prefer Opponent, Prefer Neither, Voted for Incumbent, Voted for Opponent, Voted for Neither, Voted in General Election (Validated), Voted in Primary Election (Validated), Voted in General Election (Self-Report)).

- Quote snippet:
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
- Citation anchors:
 - page=49, words=0-276, section=appendix, note=Table B3: CES Data: Summary Statistics (main variables show Observations, Mean (%), SD (pp))
 - page=49, words=0-276, section=appendix, note=Table B5: Voting Outcomes: Summary Statistics

C13: Are there descriptive figures/maps that establish baseline patterns?

- Category: C) Data, sample, and measurement
- Confidence: low (0.3392542027170478)
- Retrieval method: local
- Evidence type: descriptive figures/maps
- Data source: Appendix Figures C1–C2 and Figure C4 (appendix)
- Answer:

Yes. The document includes descriptive figures/maps establishing baseline patterns: Appendix Figures C1–C2 show how geographic and demographic features correlate with district homophily (2012 and 2020), and Figure C4 maps changes in district homophily after redistricting.

- Quote snippet:

Appendix Figures C1–C2 summarize how various geographic and demographic features correlate with district homophily, separately in 2012 and 2020.

- Citation anchors:
 - page=page 13, words=0-349, section=2.1.5 Summary Statistics and Predictors of District Homophily, note=Appendix Figures C1–C2 summarize how geographic and demographic features correlate with district homophily, separately in 2012 and 2020.
 - page=page 44, words=275-491, section=appendix, note=Figure C4: Changes in District Homophily: Equal Number of Counties Per Bin; maps illustrating changes in district homophily.

D01: What is the headline main effect estimate (sign and magnitude)?

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: medium (0.3669439595432873)
- Retrieval method: local
- Evidence type: main effect estimate from regression table
- Data source: CES data
- Table/Figure: Table C2
- Answer:

0.254 (positive)

- Quote snippet:

Heard of Incumbent District Homophily 0.254***

- Citation anchors:
 - page=65, words=Heard-Incumbent, section=section appendix, note=Main effect of district homophily on voter familiarity with incumbent

D02: What is the preferred specification and why is it preferred?

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.19304479958656237)
- Retrieval method: local
- Evidence type: Theoretical derivation / identifiability result
- Assumption flag: True
- Assumption notes: Relies on: (1) Limited fragmentation, (2) Simple Boundaries (each county intersects at most two districts), (3) Non-trivial out-of-state connections, (4) Network non-degeneracy. Under these, the mapping from Q_s to \bar{s} is locally injective around Q_0 s.
- Answer:

The preferred specification is the state-level social-network specification under limited fragmentation and simple boundaries, i.e., assuming each county intersects at most two districts (dominant district with possible secondary district), plus non-trivial out-of-state connections and network non-degeneracy. Under these conditions, there exists an $\bar{s} > 0$ such that the mapping from county-district assignment Q_s to district homophily \bar{s} is locally injective (i.e., identifiable) around the observed map Q_0 s, making district homophily informative for the assignment and enabling identification.

- Quote snippet:

There exists $\bar{s} > 0$ such that if $s < \bar{s}$, the mapping $Q_s \mapsto \bar{s}$ is locally injective at Q_0 s within the set of maps satisfying Simple Boundaries.

- Citation anchors:
 - page=page 44, words=275-491, section=A.2 Local Uniqueness of District Homophily, note=Assumptions 1-4 (Limited fragmentation, Simple Boundaries, Non-trivial out-of-state connections, Network non-degeneracy) lead to local injectivity of $Q_s \mapsto \bar{s}$ under $s < \bar{s}$.

D03: How economically meaningful is the effect (percent change, elasticity, dollars)?

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.3307902963772008)
- Retrieval method: local
- Evidence type: Quantitative estimates from the main results (even-year event studies) and Appendix B3.
- Data source: Dave Leip's Election Atlas (county-level) and CES survey data
- Assumption flag: True
- Assumption notes: Interpretation relies on the paper's stated linearity assumption (effects scale linearly with a 10pp change in district homophily). Elasticities are computed relative to stated means; results pertain to prob-

ability outcomes, not monetary values.

- Answer:

The economic meaning is modest in absolute terms but meaningful as policy-relevant information. A 10-percentage-point increase in district homophily raises the probability a respondent has heard of their representative by 0.7 percentage points (from a mean of 93.2%), and raises the probability of selecting a party by about 3.2 percentage points and selecting the correct party by about 3.3 percentage points (from means of 68.6% and 61.7%). Relative to those baselines, these correspond to roughly a 0.75% increase for hearing and about a 4–5% relative increase for party-choice outcomes.

- Quote snippet:

an increase in county's district homophily by 10pp would increase the probability that a respondent in that county has heard of their representative by 0.7pp

- Citation anchors:

– page=page 20, words=0-349, section=4.1.1 Voters' Knowledge about Representatives, note=Contains the linear-impact translation: a 10pp increase in district homophily -> 0.7pp more likely to have heard of their representative; 3.2pp more likely to select a party; 3.3pp more likely to select the correct party.

D04: What are the key robustness checks and do results survive them?

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.3282898239655706)
- Retrieval method: local
- Evidence type: robustness checks and placebo tests; specification and controls robustness
- Data source: Cooperative Election Study (CES) data and Dave Leip's Election Atlas (county-level)
- Answer:

Key robustness checks include placebo tests and specification/controls variations. The placebo outcomes show district homophily does not significantly predict placebo variables, supporting that the main findings are not driven by spurious correlations. Moreover, the district homophily effects on voter familiarity with the incumbent remain statistically significant across multiple fixed-effects specifications and control sets (e.g., County & Add FE, Add FE, Pair x Year FE, State x Year FE, DMA x Year FE, plus Demographic Controls and County-Year/DMA/State interactions), indicating results are robust to these alternative specifications.

- Quote snippet:

In general, district homophily does not significantly predict the placebo outcomes.

- Citation anchors:
 - page=55, words=0-152, section=section appendix, note=Placebo outcomes found to be generally not predicted by district homophily; supports robustness of main results
 - page=65, words=0-288, section=section appendix, note=District homophily effects persist across multiple FE/control specifications (e.g., County & Add FE, Pair x Year FE, State x Year FE, DMA x Year FE) with significant coefficients

D05: What placebo tests are run and what do they show?

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.18274930424665023)
- Retrieval method: local
- Evidence type: placebo_outcome_and_robustness_test
- Data source: Cooperative Election Study (CES) data
- Table/Figure: Table C1: CES Data: Summary Statistics for Placebo Outcomes
- Answer:

They run placebo tests using governor and senator knowledge as placebo outcomes, showing district homophily has no significant effect on these outcomes; they also perform a robustness check using commuting flows as an alternative network measure, which yields results similar in direction to the main findings but with smaller magnitudes.

- Quote snippet:

I find no significant impact of district homophily on these nine outcomes.
- Citation anchors:
 - page=26, words=0-349, section=5.1 Placebo Outcomes, note=Governor and Senator placebo outcomes; no significant impact reported.

D06: What falsification outcomes are tested (unaffected outcomes)?

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.27516588037445294)
- Retrieval method: local
- Evidence type: placebo_outcomes / falsification_tests
- Data source: Cooperative Election Study (CES) data; placebo outcomes described in Appendix C1
- Answer:

The falsification (placebo) outcomes tested are the governor and two senators knowledge/party-outcome measures: Heard of Governor; Selected Governor Party; Selected Correct Gov. Party; Heard of Senator 1; Selected Senator 1 Party; Selected Correct Sen. 1 Party; Heard of Senator 2; Selected Senator 2 Party; Selected Correct Sen. 2 Party.

- Quote snippet:
I test whether district homophily impacts voters' knowledge of their governor and senators because these offices are elected through statewide elections.
- Citation anchors:
 - page=26, words=0-349, section=5.1 Placebo Outcomes, note=Governor and Senator placebo-outcome variables described

D07: What heterogeneity results are reported (by income, size, baseline exposure, region)?

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: medium (0.375266028916009)
- Retrieval method: local
- Evidence type: textual
- Answer:

No heterogeneity results by income, county size, baseline exposure, or region are reported in the provided context. The analysis uses various fixed effects and exposure-related controls, but it does not present explicit heterogeneous effects by those dimensions.

- Quote snippet:
Table C2: Effect of District Homophily on Voter Familiarity with Representative, within Border Pairs
- Citation anchors:
 - page=65, words=0-288, section=Appendix Table C2: Effect of District Homophily on Voter Familiarity with Representative, within Border Pairs, note=Contains border-pairs results; no explicit income/size/region heterogeneity reported.
 - page=14, words=275-429, section=2.2 Redistricting, note=Mentions control variables (race, poverty) and fixed effects; indicates analyses rely on controls rather than heterogeneity estimates.

D08: What mechanism tests are performed and what do they imply?

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.1492305857441584)
- Retrieval method: local

- Evidence type: Placebo-outcome robustness tests and alternative homophily measure; diffusion mechanism validation
- Data source: CES data (*placebo outcomes*) and robustness discussion
- Answer:

The authors perform mechanism tests by (i) running placebo-outcome tests to see if district homophily spuriously predicts outcomes it should not affect, and (ii) constructing an alternative district homophily measure based on commuting flows to check robustness of the diffusion mechanism. The placebo tests show district homophily does not significantly predict placebo outcomes, supporting a causal mechanism via information diffusion within districts rather than spurious correlations; the alternative homophily measure yields qualitatively similar results, reinforcing the proposed diffusion mechanism that increases voters' knowledge about their House representatives specifically (not about governors or senators).

- Quote snippet:
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
- Citation anchors:
 - page=24, words=0-229, section=Robustness checks, note=Robustness: placebo outcomes and alternative homophily measure using commuting flows
 - page=55, words=0-152, section=Appendix C.2 Placebo Outcomes, note=Placebo outcomes summary; reports null relation between district homophily and placebo measures

D09: How sensitive are results to alternative samples/bandwidths/controls?

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.3486691202305757)
- Retrieval method: local
- Evidence type: robustness checks / sensitivity analyses
- Data source: Appendix C (Tables C1, C2) and related text; CES data; commuting flows data
- Table/Figure: Table C2
- Answer:

Results are robust to alternative samples, bandwidths, and controls. Including odd years (a smaller sample) yields similar results though with noisier estimates; using commuting flows as an alternative network measure produces effects of smaller magnitude but with the same qualitative direction; placebo tests show no significant effects; and results remain consistent across models with different fixed effects (county-year, DMA-year, etc.).

- Quote snippet:

including odd years yields similar results with noisy estimates on the odd-year coefficients.

- Citation anchors:
 - page=19, words=275-406, section=Sensitivity to sample (odd years), note=The odd years have a sample about one-fifth the size of even years; including odd years yields similar results with noisy estimates.
 - page=26, words=0-349, section=Alternative network measure: commuting flows, note=Estimates are of smaller magnitudes (about half as large) but otherwise are similar.
 - page=46, words=0-152, section=Placebo outcomes robustness, note=I find no significant impact of district homophily on these nine placebo outcomes.
 - page=65, words=0-288, section=Robustness to fixed effects (Table C2), note=Results are shown under multiple FE configurations (county, pair, state, DMA) with favorable significance patterns.
 - page=15, words=275-450, section=DMA-by-year fixed effects, note=DMA-by-year FE are added to account for media-market confounding.

D10: What are the main takeaways in the conclusion (bullet summary)?

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.22416777293153328)
- Retrieval method: local
- Evidence type: not_available
- Answer:

Conclusion not present in provided excerpts; can't extract main takeaways.

- Quote snippet:

This paper makes two primary contributions: First, I contribute to the literature on how voters learn about politics by providing causal estimates.
- Citation anchors:
 - page=1, words=275-388
 - page=2, words=0-349
 - page=3, words=275-478
 - page=4, words=0-349
 - page=13, words=0-349
 - page=36, words=0-309
 - page=37, words=0-331
 - page=38, words=0-326
 - page=39, words=0-327
 - page=40, words=0-303

E01: What are the most important prior papers cited and why are they central here?

- Category: E) Citations and related literature
- Confidence: low (0.3311747965932579)
- Retrieval method: local
- Evidence type: Literature foundation and key causal studies on social network diffusion of political information and peer effects relevant to political behavior and redistricting.
- Answer:

The most important prior papers are those establishing how social networks disseminate political information and influence political behavior, and those providing causal inference in network contexts. Key citations include Lazarsfeld, Berelson, and Gaudet (1944) for foundational ideas about information flow through social ties; Fowler et al. (2011) for causal considerations in political networks; Bond et al. (2012) for large-scale social-influence experiments; Nickerson (2008) for voting contagion and peer effects; Pons (2018) for countrywide experiments on discussion and choice; Cantoni & Pons (2022) and Brown et al. (2023) for contextual/place effects on voting behavior; Snyder & Strömberg (2010) on media coverage and accountability; and Alt et al. (2022) and related network diffuse literature. These papers are central because they provide the mechanisms (diffusion through social networks, peer effects) and the methodological tools (causal inference in networks) that the current work builds upon to study how district boundaries interact with social networks to shape political knowledge and turnout.

- Quote snippet:

voters primarily learn political information through their friends, families, neighbors, and coworkers— i.e., their social networks (Lazarsfeld et al. 1944).
- Citation anchors:
 - page=2, words=100-140, section=Introduction, note=Cites Lazarsfeld et al. (1944) as foundational on learning via social networks.
 - page=2, words=170-210, section=Introduction, note=Notes need for causal estimates in network effects; cites Fowler et al. (2011).
 - page=4, words=275-492, section=Introduction, note=Cites Cantoni & Pons (2022) and Brown et al. (2023) on context/place effects in political behavior.
 - page=37, words=260-300, section=References, note=Bond et al. (2012): large-scale social influence experiment on mobilization.
 - page=40, words=150-190, section=References, note=Nickerson (2008): voting contagion and peer effects.
 - page=36, words=0-60, section=References, note=Alt et al. (2022): diffusion of information through social networks; broad data.

E02: Which papers does this work most directly build on or extend?

- Category: E) Citations and related literature
- Confidence: low (0.31118492709752804)
- Retrieval method: local
- Evidence type: in-text_citations
- Data source: National data on social ties across the continental U.S.
- Answer:

This work directly builds on and extends the social networks and political information diffusion literature, including Baybeck & Huckfeldt (2002); Fowler et al. (2011); Sokhey & Djupe (2011); Sokhey & McClurg (2012); Campbell (2013); Klar & Shmargad (2017); Druckman et al. (2018); Fafchamps et al. (2019); Arias et al. (2019); Alt et al. (2022); Bond et al. (2012); Cantoni & Pons (2022); Brown et al. (2023); Nickerson (2008); Pons (2018).

- Quote snippet:
I build on these studies by estimating how social network structure impacts voter knowledge
- Citation anchors:
 - page=4, words=0-349, section=Introduction, note=Intro cites foundational social-network literature and key papers calling for causal estimates.
 - page=4, words=275-492, section=Introduction, note=Statement of building on these studies by estimating how social network structure impacts voter knowledge.
 - page=5, words=0-349, section=Introduction, note=List of related empirical work and lab/field experiments that the study builds on.
 - page=5, words=275-475, section=Introduction, note=Continuation listing papers and the claim to extend the literature.

E03: Which papers are used as benchmarks or comparisons in the results?

- Category: E) Citations and related literature
- Confidence: low (0.2744867434732311)
- Retrieval method: local
- Evidence type: textual evidence of benchmarking against McCartan et al. 2021
- Data source: McCartan, Kenny, Simko, Kuriwaki, et al. 2021; 50-State Redistricting Simulations
- Answer:

McCartan, Kenny, Simko, Kuriwaki, et al. (2021) — 50-State Redistricting Simulations.

- Quote snippet:

McCartan, Kenny, Simko, Kuriwaki, et al. 2021 simulate 5,000 congressional district maps for each of the 50 states. They use Monte Carlo simulation.

- Citation anchors:
 - page=34, words=275-379, section=7 Comparisons Across Many Simulated Maps, note=Mentions McCartan et al. (2021) as a source for simulated district maps used for comparison.

E04: What data sources or datasets are cited and how are they used?

- Category: E) Citations and related literature
- Confidence: low (0.3307139068230466)
- Retrieval method: local
- Evidence type: textual
- Data source: Facebook Social Connectedness Index (SCI) data; Cooperative Election Study (CES); Dave Leip's Election Atlas; ZIP-code-level SCI data; commuting-flow data as an alternative proxy.
- Answer:

Cited data sources include the Facebook Social Connectedness Index (SCI) used as a proxy for social networks to construct district and zip-code homophily; the Cooperative Election Study (CES) data for voters' knowledge and voting behavior (pre-election and post-election surveys, 2006–2022, linked to county/district); and Dave Leip's Election Atlas for actual vote data. The analysis also uses ZIP-code-level SCI data and uses commuting-flows as an alternative proxy for social networks to test robustness.

- Quote snippet:

For data on social networks, I use the Facebook Social Connectedness Index (SCI)—one of the best existing proxies for real-world social networks.
- Citation anchors:
 - page=7, words=275-379, section=Proxy for Social Networks: Facebook Social Connectedness Index (SCI), note=Introduction and use of SCI as a proxy for real-world social networks.
 - page=8, words=0-349, section=SCI construction and use for county- and zip-code pairs, note=SCI for U.S. county-county pairs and its availability for zip-code pairs; baseline description.
 - page=8, words=275-624, section=SCI data details and scale, note=Details on SCI scale, population considerations, and privacy measures.
 - page=16, words=0-349, section=Outcomes Data and CES data, note=3. Outcomes Data; CES data description, sample size, years, and linking to county/district homophily.
 - page=17, words=275-542, section=Dave Leip's Election Atlas data,

E06: Are there any seminal or classic references the paper positions itself against?

- Category: E) Citations and related literature
- Confidence: low (0.27366860789598946)
- Retrieval method: local
- Evidence type: textual evidence (reference to prior literature)
- Answer:

Yes. The paper positions itself against classic gerrymandering models that treat partisans and turnout as fixed; it contrasts with this tradition by emphasizing social learning and networks. Classic references cited as the fixed-distribution baseline include Owen and Grofman (1988); Friedman and Holden (2008, 2020); Gul and Pesendorfer (2010).

- Quote snippet:

Existing models of strategic partisan redistricting generally assume that changes to district boundaries do not affect the distribution of partisans in a given area. Owen and Grofman 1988; Friedman and Holden 2008; Gul and Pesendorfer 2010.
- Citation anchors:
 - page=page 5, words=0-349, section=Gerrymandering models (classic baseline), note=Cites traditional models Owen & Grofman 1988; Friedman & Holden 2008, 2020; Gul & Pesendorfer 2010; contrasts with social learning.
 - page=page 36, words=0-309, section=Conclusion on literature, note=States that literature on gerrymandering has largely overlooked social learning; reinforces against fixed-partisan view.

E07: Are there citations to code, data repositories, or appendices that are essential to the claims?

- Category: E) Citations and related literature
- Confidence: low (0.3061055953112889)
- Retrieval method: local
- Evidence type: data_repositories_and_appendices
- Data source: Harvard Election Data Archive; MIT Election Data and Science Lab; Dave Leip's Election Atlas; Daily Kos; CES Common Content
- Answer:

Yes. The claims cite multiple data repositories and appendices as essential sources, including Dave Leip's Election Atlas, the Harvard Election Data Archive, the MIT Election Data and Science Lab, Daily Kos, and the Cooperative Election Study (CES) Common Content. Appendix B provides data descriptions and variable definitions (e.g., vote-count construction and CES data tables).

- Quote snippet:
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)
- Citation anchors:
 - page=47, words=0-207, section=appendix, note=Appendix B: Data Descriptions; B.1 Construction of Vote Count Measures.
 - page=49, words=0-276, section=appendix, note=Appendix B: CES data tables (Table B1, B3-B5) and variable descriptions.

E08: What gaps in the literature do the authors say these citations leave open?

- Category: E) Citations and related literature
- Confidence: low (0.3038366165053296)
- Retrieval method: local
- Evidence type: literature gaps identified by authors
- Answer:

They point to a lack of causal estimates on how social networks affect political knowledge and turnout, and to the literature’s failure to causally examine how district boundaries align with social networks in shaping information diffusion and participation; furthermore, the gerrymandering literature largely treats turnout and partisans as fixed and does not incorporate information or social networks.

- Quote snippet:
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.
- Citation anchors:
 - page=2, words=0-60, section=Introduction, note=States causal estimates of social networks on political knowledge are limited; boundaries in social learning not explored.
 - page=4, words=0-25, section=Literature review, note=Notes calls for more causal estimates; aims to provide first causal estimates relating social networks to political knowledge/behavior.
 - page=5, words=0-20, section=Gerrymandering literature, note=Existing models assume turnout distribution fixed and largely overlook social networks.

F01: Are replication files or code provided? If so, where?

- Category: F) Replication and transparency
- Confidence: low (0.21164554089205545)
- Retrieval method: local
- Evidence type: not_mentioned_in_context
- Data source: Harvard Election Data Archive; MIT Election Data and Science Lab; Dave Leip's Election Atlas; Daily Kos
- Answer:

The provided context does not indicate whether replication files or code are provided, or where they would be located.

- Quote snippet:

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).

- Citation anchors:
 - page=1, words=275-388
 - page=8, words=0-349
 - page=17, words=275-542
 - page=34, words=0-349
 - page=34, words=275-379
 - page=36, words=0-309
 - page=37, words=0-331
 - page=39, words=0-327
 - page=40, words=0-303
 - page=47, words=0-207, section=appendix

F02: Is there a pre-analysis plan or registered trial? Provide details if mentioned.

- Category: F) Replication and transparency
- Confidence: low (0.25282481628114784)
- Retrieval method: local
- Evidence type: not_mentioned
- Data source: CES data (Cooperative Election Study); Dave Leip's Election Atlas (county-level) data
- Answer:

No. The provided context does not mention a pre-analysis plan or a registered trial.

- Quote snippet:

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.

- Citation anchors:
 - page=15, words=0-349, section=Methods/empirical design, note=No reference to a pre-analysis plan or registered trial in excerpt.
 - page=20, words=0-461, section=Methods/empirical design, note=Describes event-study approach; no pre-registration mentioned.

F03: Are data access constraints disclosed (restricted access, proprietary data, NDAs)?

- Category: F) Replication and transparency
- Confidence: low (0.263893269852198)
- Retrieval method: local
- Evidence type: privacy_constraint_disclosure
- Data source: Meta's Social Connectedness Index (SCI) data
- Assumption flag: False
- Answer:

Yes.

- Quote snippet:

due to privacy concerns, zip codes with small populations do not appear in the SCI data.
- Citation anchors:
 - page=8, words=320-340, section=Data constraints/Privacy, note=Privacy-related constraint: zip codes with small populations do not appear in the SCI data.

F04: Are key steps in data cleaning and construction documented?

- Category: F) Replication and transparency
- Confidence: low (0.233004065736592)
- Retrieval method: local
- Evidence type: Appendix documentation (B.1–B.2) and methods sections describing data cleaning and construction
- Data source: CES survey data; Harvard Election Data Archive; MIT Election Data and Science Lab; Dave Leip's Election Atlas; Daily Kos
- Answer:

Yes—the data cleaning and construction steps are documented, including how vote counts are constructed and cleaned (appendix B.1–B.2) and the exclusions/processing applied to CES survey data.

- Quote snippet:

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).

- Citation anchors:
 - page=page 46, words=0-349, section=section appendix, note=Aggregation framework and data descriptions tied to vote-count construction and district-level aggregation.
 - page=page 47, words=0-207, section=section appendix, note=Appendix B.1–B.2 detailing data descriptions and voting outcome construction.
 - page=page 48, words=0-333, section=section appendix, note=Variable descriptions and outcomes variables for CES data, i.e., data cleaning steps for survey data.
 - page=page 17, words=0-349, section=main text, note=Exclusion of missing responses to the candidate party recognition question in CES data cleaning.

F05: Are robustness and sensitivity analyses fully reported or partially omitted?

- Category: F) Replication and transparency
- Confidence: low (0.2713375293344792)
- Retrieval method: local
- Evidence type: Robustness and sensitivity analyses described in the robustness section; placebo tests and robustness checks are documented (e.g., alternative measures, placebo outcomes).
- Data source: CES survey data (Cooperative Election Study) and supplementary electoral data (Dave Leip's Election Atlas)
- Table/Figure: Figure 6
- Answer:

Fully reported.

- Quote snippet:

5 Robustness I find in my main specification that district homophily has a positive effect on voters' knowledge of their representatives.
- Citation anchors:
 - page=page 24, words=0-229, section=5 Robustness, note=Discusses robustness checks, placebo outcomes, and an alternative measure (commuting flows).

G01: What populations or settings are most likely to generalize from this study?

- Category: G) External validity and generalization
- Confidence: low (0.2853960652719921)

- Retrieval method: local
- Evidence type: Scope and data generalizability based on national data and redistricting focus
- Data source: Cooperative Election Study (CES)
- Answer:

The study generalizes to adult voters across the continental United States, particularly in counties affected by congressional redistricting, because it uses nationally representative CES data linked to county-level district homophily and emphasizes redistricting contexts.

- Quote snippet:

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.

- Citation anchors:
 - page=4, words=275-492, section=Intro, note=Introduces national data across the continental U.S. and relevance to redistricting.
 - page=19, words=275-406, section=3.1 Voters' Information, note=Describes CES as a nationally representative survey linking to county-level district homophily.
 - page=16, words=0-349, section=Data / Methods, note=Details CES data coverage and its use in linking to political information outcomes.
 - page=8, words=275-624, section=Data / SCI, note=Shows national-scale, county-pair data supporting broad geographic applicability.
 - page=15, words=0-349, section=Redistricting Context, note=Discusses redistricting-focused analysis in the U.S., supporting generalization to redistricting settings.

G02: What populations or settings are least likely to generalize?

- Category: G) External validity and generalization
- Confidence: low (0.2614942591602127)
- Retrieval method: local
- Evidence type: methodology limitation (explicit focus on 48 contiguous states).
- Data source: Meta's Social Connectedness Index (SCI) data from Facebook/Instagram/WhatsApp.
- Answer:

Populations/settings outside the 48 contiguous U.S. states are least likely to generalize (i.e., Alaska, Hawaii, Washington, D.C., and U.S. territories).

- Quote snippet:

Throughout this paper, I focus on the 48 contiguous states—i.e., in results I exclude Alaska, Hawaii, Washington, D.C., and territories.

- Citation anchors:
 - page=13, words=0-349, section=2.1.5 Summary Statistics and Predictors of District Homophily, note=Explicitly states focus on 48 contiguous states and exclusion of Alaska, Hawaii, DC, and territories.

G03: Do the authors discuss boundary conditions or scope limits?

- Category: G) External validity and generalization
- Confidence: low (0.28429339368151213)
- Retrieval method: local
- Evidence type: boundary_conditions_and_scope
- Data source: Appendix section (section appendix)
- Assumption flag: True
- Assumption notes: Key boundary-related assumptions include: 1) Limited fragmentation (dominant district per county), 2) Simple boundaries (each county intersects at most two districts; 95% of counties satisfy this), 3) Non-trivial out-of-state connections ($(\Pi_s) < 1$), 4) Network non-degeneracy (rows of Π s are linearly independent).
- Answer:

Yes. The authors explicitly discuss boundary conditions and scope limits, outlining constraints such as limited fragmentation and simple boundaries (no county intersects more than two districts in the observed map), openness to out-of-state connections ($(\Pi_s) < 1$), and network non-degeneracy.

- Quote snippet:
 2. Simple boundaries: The map configuration is locally simple. For the observed map Q_0 s, every county intersects at most two districts ($|\{d : q(c,d) > 0\}| \leq 2$).
- Citation anchors:
 - page=44, words=275-491, section=appendix, note=2. Simple boundaries; 3. Non-trivial out-of-state connections described; boundary scope.
 - page=45, words=0-349, section=appendix, note=Assumptions 3 and 4; open-system with leakage to out-of-state counties.
 - page=46, words=0-349, section=appendix, note=Detailed identification under boundary constraint assumptions; network leakage.

G04: How might the results change in different time periods or markets?

- Category: G) External validity and generalization
- Confidence: low (0.3194082133476395)

- Retrieval method: `local`
- Evidence type: `empirical_event_study`
- Answer:

Results may differ across time periods and markets because the size and timing of changes in district homophily depend on redistricting events, and the observable effects on voter information and choices vary by year and market. After redistricting, changes in district homophily are larger and have the strongest impact on voter knowledge in the first post-redistricting survey (2014), with effects stabilizing over time. Odd-year data are noisier due to smaller samples, reducing precision. In markets, controlling for media-market structure (DMA-by-year fixed effects) can attenuate or eliminate some effects (e.g., the Selected Party outcome in border-pairs designs), indicating variability in results across markets depending on diffusion through media and local networks.

- Quote snippet:

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).

- Citation anchors:
 - page=20, words=0-349, section=4.1.2 Voters’ Choices, note=Event studies show the strongest effect on voter knowledge in the first post-redistricting survey (2014) and stability over time.
 - page=15, words=275-450, section=2.2 Redistricting, note=DMA-by-year fixed effects are used to address media-market correlations; results can change when these fixed effects are included.
 - page=28, words=0-349, section=Appendix/Border-pairs design, note=Border-pairs results are similar to redistricting designs, but some estimates (Selected Party) become insignificant after DMA-by-year fixed effects.
 - page=14, words=0-349, section=2.1.6 Changes in District Homophily over Time, note=Describes how changes in district homophily vary with redistricting and between years.

H01: Are key variables measured directly or via proxies?

- Category: `H) Measurement validity`
- Confidence: `low (0.3328604927197932)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Facebook SCI proxy for social networks; CES survey responses for direct measures.
- Answer:

Key variables are measured via proxies in this work. Social connectedness is proxied by the Facebook Social Connectedness Index (SCI), and information-

diffusion constructs are built from county-level friendship shares and population shares. In contrast, core survey-based outcomes (Heard of Representative, Selected Party, and voting outcomes) are measured directly from CES survey responses.

- Quote snippet:

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

- Citation anchors:
 - page=7, words=275-379, section=2.1.2 Proxy for Social Networks, note=SCI used as proxy for social networks.
 - page=17, words=0-349, section=3.2 Voter Knowledge / Voter Turnout, note=Discussion of using CES survey responses to study voting outcomes.
 - page=49, words=0-276, section=Table B3: CES Data: Summary Statistics, note=Describes survey-based variables (Heard of Representative, Selected Party, etc.)

H02: What measurement error risks are acknowledged or likely?

- Category: H) Measurement validity
- Confidence: low (0.2709235671458542)
- Retrieval method: local
- Evidence type: data limitations / measurement error risks in survey and administrative data
- Data source: Cooperative Election Study (CES); Dave Leip's Election Atlas; Harvard Election Data Archive; MIT Election Data and Science Lab; Daily Kos
- Answer:

Measurement error risks acknowledged or likely include: (1) reliance on self-reported survey data for hearing of representatives, party identification, and vote choices, which can be misreported or misunderstood; (2) attrition between pre- and post-election CES waves leading to potential attrition bias in longitudinal measurements; (3) misassignment of respondents to incorrect congressional districts in 2020, with some responses excluded to mitigate this; (4) potential misalignment when linking respondents to county-level district homophily measures; (5) weighting schemes that adjust to the population of people (not counties) may introduce weighting-related measurement error; (6) use of multiple vote-count data sources with possible data quality issues and cross-source inconsistencies.

- Quote snippet:

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.

- Citation anchors:
 - page=17, words=0-349, section=3.2 Voter Turnout, note=Attrition between surveys; longitudinal measurement risk.
 - page=17, words=275-542, section=3.2 Voter Turnout, note=In 2020, 925 NC respondents were assigned to incorrect districts; excluded to mitigate misassignment.
 - page=16, words=0-349, section=3 Outcomes Data, note=Use of CES cumulative weights; potential weighting-related measurement error.
 - page=49, words=0-276, section=Appendix B, note=Variable descriptions rely on self-reported survey responses (e.g., Heard of Representative, Selected Party); potential measurement error.
 - page=47, words=0-207, section=B Data Descriptions, note=Vote-count measures drawn from multiple sources (Harvard Archive, MIT Lab, Dave Leip's, Daily Kos); potential data quality issues.

H03: Are there validation checks for key measures?

- Category: H) Measurement validity
- Confidence: low (0.26058062390360937)
- Retrieval method: local
- Evidence type: textual_evidence
- Data source: CES Voting Outcomes data (validated) with placebo outcome tests in Appendix tables
- Table/Figure: Table B2: Descriptions for CES Voting Outcomes (and related Table B3) and Table C1: Placebo Outcomes
- Answer:

Yes. The study includes validation checks for key measures, including using validated voting data (e.g., 'Voted in General Election (Validated)') and placebo-outcome tests to validate the measures.

- Quote snippet:

"Voted in General Election (Validated) 417,421 57.5 49.4"
- Citation anchors:
 - page=page 49, words=0-276, section=section appendix, note=Table B2/B3 describe CES Voting Outcomes and include a 'Validated' voting measure.
 - page=page 55, words=0-152, section=section appendix, note=Placebo-outcome tests (Table C1) used to validate key measures.

H04: Do the authors discuss construct validity for core outcomes?

- Category: H) Measurement validity
- Confidence: low (0.22677342498415975)
- Retrieval method: local

- Evidence type: Textual evidence: measurement construction and placebo validation described; no explicit construct validity section
- Data source: Cooperative Election Study (CES)
- Answer:

The text does not explicitly discuss construct validity for core outcomes. It describes how core outcomes are constructed from CES responses (e.g., Heard of Representative, Selected Party, Selected Correct Party) and notes Appendix B contains detailed variable descriptions; it also reports placebo-outcome analyses as validity checks, but uses the phrase 'construct validity' nowhere in the excerpts.

- Quote snippet:
I construct three binary variables to assess how familiar respondents are with their current representative.
- Citation anchors:
 - page=16, words=0-60, section=3 Outcomes Data / 3.1 Voters' Information, note=Describes construction of core information outcomes and references Appendix Table B1
 - page=47, words=0-40, section=Appendix B / B.1 Construction of Vote Count Measures, note=Variable descriptions and measurement construction
 - page=55, words=0-40, section=Appendix C.2 Placebo Outcomes for Voter Information, note=Placebo outcomes used to validate measurement
 - page=16, words=60-100, section=3 Outcomes Data / 3.1 Voters' Information, note=Appendix reference to Appendix Table B1
 - page=34, words=0-20, section=6 Conceptual Framework, note=Discussion of district homophily as a summary statistic (indirectly relates to construct validity)

I01: What policy counterfactuals are considered or implied?

- Category: I) Policy counterfactuals and welfare
- Confidence: low (0.3499960543662869)
- Retrieval method: local
- Evidence type: policy-counterfactuals (redistricting maps)
- Data source: McCartan et al. 2021; 50-State Redistricting Simulations; related literature on counterfactual district maps
- Answer:

Alternative redistricting maps (district boundary changes) are the policy counterfactuals considered; the work discusses simulating counterfactual district maps to study diffusion, including thousands of maps per state.

- Quote snippet:

Future work will focus on estimating the parameters of the diffusion process, which would allow for the simulation of counterfactual district maps.

- Citation anchors:
 - page=34, words=0-349, section=main, note=Future work: simulate counterfactual district maps by estimating diffusion parameters and simulating before/after redistricting.
 - page=38, words=0-326, section=main, note=McCartan et al. 2021 simulate 5,000 congressional district maps for each state.
 - page=40, words=0-303, section=main, note=50-State Redistricting Simulations.
 - page=28, words=0-349, section=main, note=Identification of district effects and redistricting boundary design used to study district homophily.

I02: What are the main welfare tradeoffs or distributional impacts discussed?

- Category: I) Policy counterfactuals and welfare
- Confidence: low (0.3034664343756288)
- Retrieval method: local
- Evidence type: Empirical findings on district homophily, knowledge, turnout, and campaign contributions
- Data source: Facebook friendship network data; commuting flows; CES survey data
- Table/Figure: Figure 6: Roll-Off effects
- Answer:

The paper discusses welfare and distributional implications arising when social networks align or misalign with district boundaries. Key findings suggest that increasing district homophily (alignment of county networks with districts) improves voter knowledge, reduces roll-off, and shifts campaign contributions toward in-district candidates, implying welfare gains from better information and more district-focused political participation. However, these effects are heterogeneous across counties, and redistricting often does not drastically change district homophily, leading to uneven benefits. Moreover, once social learning is considered, turnout decisions become dependent on how boundaries are drawn, raising fairness concerns: districts with highly aligned networks gain more informational and turnout advantages, while others may gain less. The results also imply potential policy trade-offs: using social-network data to draw fairer districts could enhance information diffusion and representation, but could also be exploited by partisan actors seeking to manipulate outcomes.

- Quote snippet:

an increase in district homophily by 10pp would increase the probability that a respondent in that county has heard of their represen-

tative by 0.7pp.

- Citation anchors:
 - page=20, words=0-349, section=4.1.2 Voters' Choices, note=Event studies: district homophily changes affect voter knowledge and choices (first survey after redistricting)
 - page=24, words=0-229, section=4.1.2 Voters' Choices, note=Roll-off: higher district homophily reduces roll-off; 10pp increase -> ~0.04pp drop in roll-off
 - page=24, words=0-229, section=4.2 Campaign Contributions, note=10pp increase in district homophily raises in-district share of contributions by ~7.4pp
 - page=5, words=0-349, section=2 Empirical Strategy and Networks Data, note=Turnout decisions depend on how district borders are drawn due to social learning
 - page=38, words=0-326, section=Robustness/Policy Implications, note=Networks data can help policymakers draw fairer districts, but can be exploited by partisan actors

I03: Are cost-benefit or incidence analyses provided?

- Category: I) Policy counterfactuals and welfare
- Confidence: low (0.24491668942730976)
- Retrieval method: local
- Evidence type: absence of cost-benefit/incidence analysis in the provided context
- Answer:

No cost-benefit or incidence analyses are provided in the excerpt.

- Quote snippet:

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.
- Citation anchors:
 - page=page 18, words=275-496, note=Discusses 'cost of voting' and turnout but does not present cost-benefit or incidence analyses.

I04: What policy recommendations are stated or implied?

- Category: I) Policy counterfactuals and welfare
- Confidence: low (0.306898121324069)
- Retrieval method: local
- Evidence type: policy_recommendation
- Data source: Social Connectedness Index (SCI) data
- Assumption flag: True

- Assumption notes: Assumes policymakers will incorporate network data when drawing electoral districts to improve fairness; acknowledges possible misuse by gerrymandering actors.
- Answer:

Policy recommendations include using detailed social network data (e.g., the Social Connectedness Index) to inform redistricting in order to draw fairer districts, while recognizing that such data could be exploited by partisan gerrymandering.

- Quote snippet:
enabling its use by policymakers to draw fairer districts, but also by partisan gerrymanderers who may seek to exploit it.
- Citation anchors:
 - page=36, words=0-309, section=36, note=Policy relevance: using SCI data to produce fairer districts; warns about misuse by gerrymanderers.

J01: What missingness or attrition patterns are reported?

- Category: J) Data quality and integrity
- Confidence: low (0.31231345719318343)
- Retrieval method: local
- Evidence type: text
- Data source: CES data (Congressional Election Survey)
- Answer:

Attrition and missingness patterns reported include: (1) attrition between pre-election and post-election CES surveys; (2) uneven county representation across years (even years 80–90% of counties represented; odd years ~ two-thirds); (3) missing responses to key questions (candidate party recognition <2% missing; in many cases House candidate name missing; demographic questions often unanswered), and (4) certain respondents are excluded (e.g., 2020 NC respondents assigned to the wrong district).

- Quote snippet:
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).
- Citation anchors:
 - page=17, words=0-349, section=Methods / CES data, note=Attrition between pre- and post-election CES surveys; uneven county representation across years.
 - page=17, words=275-542, section=Methods / Data cleaning, note=NC 2020 misassignment and exclusion; missing data handling for candidate party recognition and demographics.

J02: How are outliers handled (winsorization, trimming, exclusions)?

- Category: J) Data quality and integrity
- Confidence: low (0.30922656251558667)
- Retrieval method: local
- Evidence type: Methods: data exclusions (sample selection)
- Answer:

The text does not describe any outlier handling such as winsorization or trimming. It only mentions exclusions of certain observations due to missing responses and mis-assigned districts.

- Quote snippet:

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).
- Citation anchors:
 - page=17, words=275-542, section=Data Cleaning / Exclusions, note=Excludes missing responses and misassigned respondents; describes exclusions for 2006-07 counties and 2020 NC misassignment.

J03: Are there data audits or validation steps described?

- Category: J) Data quality and integrity
- Confidence: low (0.24188171286533028)
- Retrieval method: local
- Evidence type: data cleaning and validation; placebo validation
- Data source: Cooperative Election Study (CES); Dave Leip's Election Atlas; Harvard Election Data Archive; MIT Election Data and Science Lab; Daily Kos
- Answer:

Yes. The context describes data audits/validation steps, including cleaning and exclusion criteria for survey responses (e.g., removing missing responses, misassigned counties/districts, and incorrect congressional district assignments in 2020), and uses placebo tests to validate that district homophily does not predict placebo outcomes.

- Quote snippet:

The nine figures that follow show the results of the placebo tests. In general, district homophily does not significantly predict the placebo outcomes.
- Citation anchors:
 - page=page 17, words=275-542, section=3.2 Voter Turnout, note=Exclusion of missing/incorrect responses and misassignments; 2020 NC misassignment — data cleaning/validation.

- page=page 55, words=0-152, section=C.2 Placebo Outcomes for Voter Information, note=Placebo tests validation: district homophily not predicting placebo outcomes.
- page=page 16, words=0-349, section=3 Outcomes Data, note=Use of weights for comparability across years; data structure.

J04: Is there evidence of reporting bias or selective sample inclusion?

- Category: J) Data quality and integrity
- Confidence: low (0.307008179298014)
- Retrieval method: local
- Evidence type: Methodology and robustness checks
- Data source: Cooperative Election Study (CES) and Dave Leip's Election Atlas
- Answer:

No explicit evidence of reporting bias or selective sample inclusion is reported in the provided context. The author notes steps to mitigate bias, including applying CES cumulative weights, excluding missing responses (<2%), and excluding misassigned cases (e.g., 2020 North Carolina respondents). They also acknowledge limitations such as attrition between surveys, smaller odd-year samples, and urban-area concentration. Placebo tests show district homophily does not predict placebo outcomes, and results are robust to county fixed effects, suggesting robustness rather than bias in sampling. Nonetheless, these caveats indicate potential sample composition limitations rather than confirmed reporting bias.

- Quote snippet:
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).
- Citation anchors:
 - page=16, words=0-349, section=3 Outcomes Data / 3.1 Voters' Information, note=Describes use of CES data, cumulative weights, and sample exclusions (<2%).
 - page=17, words=0-349, section=3.2 Voter Turnout, note=Discusses post- vs pre-election surveys, attrition between surveys, and sample size differences (odd vs even years).
 - page=17, words=275-542, section=3.2 Voter Turnout / Data robustness, note=Notes 2020 NC misassignment (925 respondents) and county-fixed-effects robustness; 2010 base year; odd-year sample details.
 - page=55, words=0-152, section=Appendix C. Placebo Outcomes, note=Summary statistics and results for placebo tests; district homophily generally not predictive of placebo outcomes.
 - page=67, words=0-258, section=Appendix C.7 Vote Count Data, note=Describes turnout analysis and lack of top-of-ticket effects; dis-

cussion of roll-off.

K01: What goodness-of-fit or diagnostic metrics are reported?

- Category: K) Model fit and diagnostics
- Confidence: low (0.28800660448158044)
- Retrieval method: local
- Evidence type: Regression table diagnostics (Table C2) showing R2 per specification with clustered standard errors.
- Data source: CES data; county-by-congressional-district voting/outcome data; Appendix C2.
- Table/Figure: Table C2
- Answer:

The reported goodness-of-fit diagnostics are R-squared (R2) values for each specification, along with standard errors in parentheses and significance stars; the table also reports the number of observations (Obs).

- Quote snippet:
R2 0.606 0.606 0.699 0.699 0.718 0.720
- Citation anchors:
 - page=page 65, words=0-288, section=appendix, note=Contains Table C2: Effect of District Homophily on Voter Familiarity with Representative, within Border Pairs; reports Obs and R2 values across specifications.

K02: Are functional form choices tested (logs, levels, nonlinearities)?

- Category: K) Model fit and diagnostics
- Confidence: low (0.27720984198510307)
- Retrieval method: local
- Evidence type: methodology_description
- Answer:

There is no explicit evidence in the provided context that the authors tested alternative functional forms (e.g., logs, levels, nonlinearities). The described analyses reference fixed effects, binary outcomes interpreted as changes in probability, and event studies, but do not report tests of alternative functional forms.

- Quote snippet:
District homophily is measured on a scale from 0 to 1, and outcome variables are binary.
- Citation anchors:
 - page=19, words=275-406, section=main_text, note=Binary outcomes; interpretation as change in probability; no mention of alternate functional forms.

- page=15, words=275-450, section=main_text, note=Discussion of district- and DMA-by-year fixed effects; no mention of testing different functional forms.
- page=65, words=0-288, section=main_text, note=Regression table with fixed effects; showcases model specification but not nonlinear/log form tests.
- page=34, words=0-349, section=main_text, note=Mathematical framework; no explicit tests of alternative functional forms mentioned.

K03: Are residual checks or specification tests reported?

- Category: K) Model fit and diagnostics
- Confidence: low (0.2238758557419299)
- Retrieval method: local
- Evidence type: robustness/specification tests
- Data source: Appendix C2; CES data with border-pair design
- Table/Figure: Table C2
- Answer:

Yes. The paper reports specification tests/robustness checks across multiple model specifications (as shown in Table C2).

- Quote snippet:
Table C2: Effect of District Homophily on Voter Familiarity with Representative, within Border Pairs
- Citation anchors:
 - page=65, words=0-288, section=section appendix, note=Table C2 shows multiple specifications (County & Year FE, State x Year FE, DMA x Year FE, etc.) as robustness checks.

K04: How sensitive are results to alternative specifications or estimators?

- Category: K) Model fit and diagnostics
- Confidence: medium (0.3655676425372999)
- Retrieval method: local
- Evidence type: robustness_checks
- Data source: CES survey data
- Answer:

The results are robust to alternative specifications and estimators. Using commuting flows as an alternative network measure yields similar qualitative effects with smaller magnitudes; a border-pairs identification gives qualitatively similar results to the redistricting design, though some coefficients lose significance when additional fixed effects are included; placebo outcomes show no effect of

district homophily; and results remain consistent across different fixed-effects specifications (county-by-year, pair-by-year, state-by-year, DMA-by-year).

- Quote snippet:
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.
- Citation anchors:
 - page=26, words=Commuting-incumbent., section=Appendix C.3, note=Commuting district homophily estimates are smaller but similar to main results.
 - page=28, words=With-effects., section=Appendix, note=Border-pairs design yields qualitatively similar results; some estimates lose significance with DMA-by-year FE.
 - page=55, words=Placebo-outcomes., section=Appendix, note=Placebo outcomes show district homophily does not significantly predict placebo outcomes.
 - page=65, words=Heard-Incumbent, section=Appendix, note=Coefficients for 'Heard of Incumbent' remain robust across FE specifications.

P01: ResponseTextConfig(format=ResponseFormatText(type='text'), verbosity='medium')

- Category: P) Previous questions
- Confidence: n/a
- Retrieval method: n/a
- Evidence type: n/a
- Answer:
- Roll-off measure definition and purpose
 - Definition: difference between top-of-ticket votes and House votes, as a share of top-of-ticket votes; captures share of voters who paid the voting cost for top-ticket but abstain from House (roll-off). Also linked to Swing Voter's Curse literature. (Source: page 18, words 275–496)
- Campaign contributions data and identification strategy
 - Data: campaign contributions to House candidates from Kuziemko et al. 2023 (Oct), built from FEC data via Bonica 2014.
 - Identification: geocode to flag whether contributor lives in the same congressional district as the candidate. (Source: page 18)

- Border-pairs empirical design and estimation
 - Design: compare county-pair deviations from their pair mean; restrict to counties fully within one district; collapse to county level to keep within-district border variation.
 - Key specification: $y_{ct} = \alpha_c + \alpha_{pt} + \beta_{c,t} + X_{ct} + \epsilon_{ct}$ (11); $\beta_{c,t}$ is the coefficient of interest; include state-by-year fixed effects; no district-by-year fixed effects due to data limits. (Source: page 28, 0–349)
- Conceptual framework: information diffusion within districts
 - Main idea: district homophily summarizes network structure, shaping the diffusion of information and the equilibrium share of informed voters. (Source: page 39, 0–327)
- Appendix B: data construction and variable descriptions
 - B.1: construction of vote-count measures (county-by-CD; sources include Harvard MIT data for various years, Dave Leip’s Atlas, Daily Kos)
 - B.2: variable descriptions for CES and voting outcomes (e.g., Heard of Representative, Selected Party, Voted in General Election, etc.)
 - B.3–B.5: descriptive statistics for CES outcomes and voting measures (Source: page 47–49)
- Appendix C: data and figures (highlights)
 - C2: results for border-pairs design; C5, C11, C14 figures illustrating district homophily effects on self-reported voting outcomes (Source: page 64; figure references on 54, 61, 64)

If you’d like, I can extract specific equations or variable definitions into a compact reference list. (Cited sources: page 18; page 28; page 39; Appendix B pages 47–49; Appendix C pages 54, 61, 64)