Audit Report: Workflow cf1d92ff869046f39b7319007e9a7835

Overview

- Source JSON: reports\workflow-report-cf1d92ff869046f39b7319007e9a7835.json
- Run ID: cf1d92ff869046f39b7319007e9a7835
- Papers input: papers\NetworksElectoralCompetition.pdf
- Started at: 2026-02-15T19:29:20.818045+00:00
- Finished at: 2026-02-15T19:40:51.819956+00:00
- Duration: 0:11:31.001911

Effective Configuration

- Chat model: gpt-5
- 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: skipped (reason: db_unreachable)
- report_store: skipped (reason: db_unreachable)

Agentic Summary

- Status: completed
- Main question: What is the key contribution?
- Report question set: both
- Structured questions generated: 84
- Confidence mean/median: 0.2921069186783523 / 0.30376113152112244
- Confidence labels: low=77, medium=6, high=0

Final Answer

- Core contribution: Formalizes and empirically tests how district homophily governs the diffusion of information about elected officials—modeled as a network statistic that determines the equilibrium share of informed voters—and links this mechanism to electoral behavior. Empirically, it identifies homophily's effects using a within-state border-pairs

design and operationalizes participation via a precise roll-off measure. (page 28 words 0-349; page 18 words 275-496)

- Theory to measurement: Develops a model of news diffusion within districts, summarizing district homophily as a network statistic shaping the informed share; operationalizes abstention conditional on showing up via roll-off = (VotesTop − VotesHouse)/VotesTop. (page 28 words 0-349; page 18 words 275-496)
- Identification: Implements a neighboring-county border-pairs design with county and pair-by-year fixed effects and time-varying controls to isolate district-assignment-induced differences in homophily; documents robustness and the sensitivity of "Selected Party" to adding DMA-by-year FE. (page 28 words 0-349)
- Data infrastructure: Assembles multi-source county and county-by-CD vote counts (precinct-level EDA/EDS merged with Dave Leip's and Daily Kos), CES-based knowledge/preference/turnout measures, and geocoded FEC contributions identifying in-district contributors—enabling comprehensive tests of the theory. (page 47 words 0-207; page 48 words 0-333; page 18 words 275-496)
- Empirical grounding: Provides descriptive baselines for roll-off and turnout (mean roll-off 4.42%; top-of-ticket turnout 51.34%; House turnout 49.07%) and for voter knowledge/behavior (e.g., Heard of Rep 93.2%; Validated General Turnout 57.5%), anchoring the analysis. (page 49 words 0-276)

Sub-Answers

Sub-answer 1

- Question: ResponseTextConfig(format=ResponseFormatText(type='text'), verbosity='medium')
- Question tokens estimate: `19`
- Answer:

No specific question was posed. Below is a concise synthesis of key definitions, data, and identification from the provided context, with citations.

- Roll-off (House turnout relative to top-of-ticket)
  - Definition: Share of top-of-ticket voters who abstain from the House race: (VotesTop − VotesHouse) / VotesTop; captures abstention after incurring the cost of voting; related to "roll-off" literature (Feddersen & Pesendorfer 1996; Miller 2022; Snyder & Strömberg 2010) (page 18 words 275-496).
  - Variable construction and scope: Main analysis at county level from Dave Leip's; robustness at county-by-CD level from Harvard EDA,

Daily Kos, Dave Leip's, and MIT EDS Lab; exclude elections without a top-of-ticket race (page 49 words 0-276).
  - Summary stats: Mean 4.42%, SD 12.22 pp; also report turnout in top-of-ticket (mean 51.34%) and House (mean 49.07%) (page 49 words 0-276).

- Campaign contributions data
  - Source: Kuziemko et al. 2023 (October), constructed from FEC data in Bonica 2014; geocoding used to identify whether contributor resides in the same congressional district as the candidate (page 18 words 275-496).

- Border pairs identification design
  - Setup: Neighboring county pairs that straddle a district boundary; counties are similar except district assignment, yielding different district homophily; restrict to counties fully within one district; collapse outcomes to county level; include one observation per county–pair; compare within-state pairs; precision decreases due to restricted sample (page 28 words 0-349).
  - Specification:
    * $y_{ct} = \alpha_c + \mu_{pt} + \beta\,\bar{\pi}_{c,t} + X'_{ct}\delta + \varepsilon_{ct}$, where y_{ct} is the outcome, μ_{pt} pair-by-year FE, X_{ct} time-varying county controls; use state-by-year FE (insufficient data for district-by-year FE) (page 28 words 0-349).
  - Results note: Qualitatively similar to redistricting design; estimates on "Selected Party" become insignificant after adding DMA-by-year FE (page 28 words 0-349).

- Conceptual framework
  - Develops a theoretical model of information diffusion within districts; district homophily summarized as a network statistic; considers news about elected officials diffusing through networks to determine the equilibrium share of informed voters (page 28 words 0-349).

- Vote count data construction
  - County-by-CD measures built from precinct-level vote counts (Harvard EDA 2000–2010; MIT EDS Lab 2016–2020) combined with county-by-CD vote counts from Dave Leip's (House) and Daily Kos (President, Senator, Governor) (page 47 words 0-207).

- CES variables and summary statistics (knowledge, preferences, voting)
  - Knowledge outcomes (binary): Heard of Representative; Selected Party; Selected Correct Party; defined via pre-survey items on recognizing and assigning party to the House representative (page 47 words 0-207).

  - Voting/preference outcomes (binary): Prefer Incumbent/Opponent/Neither (pre-survey); Voted for Incumbent/Opponent/Neither (post-survey); Validated turnout in general/primary; self-reported general turnout; handling of missing when no incumbent or unmatched rolls (page 48 words 0-333).
  - Summary (sample means): Heard of Rep 93.2%; Selected Party 68.6%; Selected Correct Party 61.7%; Prefer Incumbent 40.1%; Voted for Incumbent 41.0%; Validated General Turnout 57.5%; Self-Reported General Turnout 87.8% (page 49 words 0-276).
- Additional pointers
  - Figures showing effects of homophily on self-reported voting using commuting- and zip-code-based measures (Figures C11, C14); a figure on changes in district homophily with progressively finer bins (Figure C5) (page 61 words 0-15; page 64 words 0-45; page 54 words 0-18).

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 (0.31323401739665313)`
- Retrieval method: `local`
- Evidence type: `direct statement (introduction)`
- Data source: Introduction
- Assumption flag: `False`
- Answer:

How the alignment between social networks and political boundaries (district homophily) affects voters' political knowledge and behavior, including turnout.

- Quote snippet:

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

- Citation anchors:
  - page=2, words=0-349, section=Introduction, note=States the research question explicitly.
  - page=5, words=0-349, section=Contribution, note=Reiterates focus on how the match between networks and boundaries affects knowledge and behavior.

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.3201369538893216)`
- Retrieval method: `local`
- Evidence type: `direct statement in text`
- Assumption flag: `False`
- Answer:

The paper's primary contribution is to provide causal, near-national estimates of how social networks (district homophily) affect voter knowledge—shifting the focus beyond traditional media by leveraging geographic variation in U.S. social ties (and, secondarily, linking social learning to political geography/redistricting).

- 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

- Citation anchors:
  - page=3, words=0-349, section=Introduction, note=States two primary contributions; first is causal, near-national estimates of networks' impact on voter knowledge.
  - page=4, words=275-492, section=Introduction, note=Clarifies leveraging geographic variation in the U.S. social network; relevance for redistricting.
  - page=5, words=275-475, section=Introduction, note=Connects social learning to political geography and implications for gerrymandering measures.

A03: What is the central hypothesis being tested?

- Category: `A) Research question / contribution`
- Confidence: `low (0.28653574361407586)`
- Retrieval method: `local`
- Evidence type: `textual evidence`
- Data source: Paper text in provided context
- Assumption flag: `False`
- Answer:

Greater alignment between social ties and congressional district boundaries (district homophily) facilitates the diffusion of representative-specific political information, making voters more informed and influencing voting behavior.

- Quote snippet:

Assuming that information about representatives spreads more readily between friends in the same district, the way district boundaries are drawn can shape how political information flows through social networks.

- Citation anchors:
  - page=6, words=0-349, section=main, note=Defines district homophily and posits that information spreads more readily between friends in the same district, implying boundaries shape political information flow.
  - page=28, words=275-405, section=main, note=States empirical link and provides a diffusion framework showing how district homophily determines informed share.
  - page=20, words=0-349, section=main, note=Reports event-study evidence that changes in district homophily change voter knowledge, consistent with the hypothesis.

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.34403041836800285)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: CES; Dave Leip's Election Atlas; Harvard Election Data Archive; MIT Election Data and Science Lab; Daily Kos; Census (VAP)
- Table/Figure: Appendix Tables B1, B2, B4–B5
- Assumption flag: `False`
- Answer:

Dependent variables include: (A) Voter information measures from CES: Heard of Representative, Selected Party, Selected Correct Party. (B) Candidate preference (pre-survey): Prefer Incumbent, Prefer Opponent, Prefer Neither. (C) Vote choice (post-survey): Voted for Incumbent, Voted for Opponent, Voted for Neither. (D) Turnout measures from CES: Voted in General Election (Validated), Voted in Primary Election (Validated), Voted in General Election (Self-Report). (E) Vote-count outcomes: House roll-off (share of top-of-ticket voters abstaining in House race), Turnout in Top-of-Ticket Election (as share of VAP), and Turnout in House Election (as share of VAP).

- Quote snippet:

  I study the impact of district homophily on voters' knowledge and political behavior... I then incorporate vote count data to reveal actual voting behavior

- Citation anchors:
  - page=16, words=0-349, note=States study examines voter knowledge, self-reported choices, and actual voting via vote counts.

  - page=47, words=0-207, section=appendix, note=Appendix Table B1 defines voter knowledge outcomes.
  - page=48, words=0-333, section=appendix, note=Appendix Table B2 defines CES voting outcomes (preferences, vote choices, turnout variables).
  - page=49, words=0-276, section=appendix, note=Table B4/B5 define vote-count outcomes: roll-off and turnout measures.
  - page=22, words=0-349, note=Text describes roll-off as share of top-of-ticket voters who abstain in House election.

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.24776601700919515)`
- Retrieval method: `local`
- Evidence type: `textual`
- Table/Figure: Table C2
- Assumption flag: `False`
- Answer:

District homophily—both its level (used directly in regressions) and, for the redistricting event study, the change in district homophily a county experiences between 2012 and 2013 ($\Delta^{-}c$).

- Quote snippet:

  $\Delta^{-}c$ is the change in district homophily experienced by county c between 2012 and 2013

- Citation anchors:
  - page=15, words=0-349, section=main, note=Defines $\Delta^{-}c$ as the change in district homophily between 2012 and 2013 used in event studies.
  - page=65, words=0-288, section=appendix, note=Table C2 shows 'District Homophily' as the independent variable in regressions.

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.31368693464548203)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: national data on social ties
- Assumption flag: `False`
- Answer:

United States (continental U.S.); U.S. congressional districts/elections and voter information diffusion via social networks; centered on the 2010 Census redistricting with treatment in 2012–2013 and an event-study window spanning roughly 2006–2022.

- 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=4, words=275-492, note=Scope across the continental U.S. using national data on social ties
  - page=15, words=0-349, note=Focus on post-2010 Census redistricting; new borders for Nov 2012 elections; treatment timing and event-study years (incl. 2006–2022)

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.26822049639980433)`
- Retrieval method: `local`
- Evidence type: `textual`
- Assumption flag: `False`
- Answer:

Information diffuses through social networks and travels faster among same-district friends (district homophily); thus, alignment between social ties and district boundaries drives voter knowledge about representatives.

- Quote snippet:

  > Accordingly, information about representatives is likely to spread more quickly when people are more likely to interact with others from the same district.

- Citation anchors:
  - page=6, words=0-349, note=Defines mechanism: voters learn via networks; same-district ties accelerate spread; introduces district homophily.
  - page=28, words=275-405, note=Conceptual framework of diffusion with homophily; district homophily as summary statistic.
  - page=34, words=0-349, note=Clarifies district homophily arises from intersection of networks and district boundaries.
  - page=4, words=0-349, note=Highlights social networks as key source of political information.

A08: What alternative mechanisms are discussed?

- Category: `A) Research question / contribution`
- Confidence: `low (0.2426677094907878)`
- Retrieval method: `local`
- Evidence type: `textual evidence`
- Assumption flag: `False`
- 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=150-210, note=Peer-effects literature mentions other mechanisms.

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.34582459069433863)`
- Retrieval method: `local`
- Evidence type: `textual`
- Assumption flag: `False`
- Answer:

The paper argues that how district lines are drawn shapes how political information spreads through social networks, which in turn affects voter knowledge and turnout. Thus, redistricting and assessments of partisan bias should incorporate social learning (rather than assume fixed voter behavior). It further notes that newly available social network data (e.g., the SCI) could help policymakers draw fairer districts, but the same data could be exploited by partisan gerrymanderers.

- 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=Empirical Strategy: District Homophily, note=District boundaries can shape political information flow via network–district alignment.
  - page=5, words=275-475, section=Main text (gerrymandering implications), note=Findings imply turnout depends on how borders group one's social network; literature overlooks social learning.

  - page=36, words=0-309, section=Implications preceding References, note=Social network data (SCI) could be used to draw fairer districts but may also be exploited by partisan gerrymanderers.

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.31149671482875635)`
- Retrieval method: `local`
- Evidence type: `textual_evidence`
- Data source: Main text (pages 19, 36)
- Assumption flag: `False`
- Answer:

The paper does not provide a formal welfare analysis. Its results imply that higher district homophily increases voter knowledge and reduces abstention, and it notes policy relevance: such network data could be used to draw fairer districts but also exploited by gerrymanderers, making welfare implications ambiguous.

- Quote snippet:

  district homophily increases voters' knowledge about their representatives, and accordingly decreases abstention in House elections.

- Citation anchors:
  - page=19, words=40-140, note=States main results: knowledge increases and abstention decreases.
  - page=36, words=150-250, note=Policy relevance highlighting potential for fairer districts and for exploitation by gerrymanderers.

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.26450904786970586)`
- Retrieval method: `local`
- Evidence type: `textual`
- Assumption flag: `True`
- Assumption notes: Identification requires: (1) limited fragmentation with a dominant district per county; (2) simple boundaries so each county spans 2 districts and nearby maps preserve this; (3) leakage to out-of-state nodes ((Πs)<1); (4) non-degenerate network (independent rows).
- Answer:

They note two broad limitations. Substantively, the paper does not yet estimate the diffusion model's key parameters or run counterfactual map simulations—those are deferred to future work. Methodologically, their identification result relies on restrictive map/network assumptions: limited fragmentation with a dominant district per county, simple boundaries where a county intersects at most two districts (and perturbations preserve this), non-trivial out-of-state

network leakage (spectral radius < 1), and network non-degeneracy (linearly independent rows). They acknowledge the two-district split condition is restrictive, though it holds for most U.S. counties.

- 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=200-320, note=Future work to estimate diffusion parameters and simulate counterfactual maps
  - page=44, words=300-491, section=appendix, note=Assumptions 1–2: Limited fragmentation and simple boundaries; note on two-district split prevalence
  - page=45, words=0-180, section=appendix, note=Assumptions 3–4: Non-trivial out-of-state connections and network non-degeneracy
  - page=46, words=175-320, section=appendix, note=Identification depends on leakage and non-degeneracy to ensure invertibility

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

- Category: `A) Research question / contribution`
- Confidence: `low (0.31282513555240943)`
- Retrieval method: `local`
- Evidence type: `textual evidence from main text`
- Data source: Facebook friendship graph (national social ties)
- Assumption flag: `False`
- Answer:

The paper claims novelty in using nationwide social-network data (the Facebook friendship graph) to construct a district homophily measure that aligns social ties with congressional districts, and in identifying causal effects by exploiting geographic variation and redistricting-induced changes in homophily over time.

- Quote snippet:

  By employing national data on social ties, my analysis comprehensively captures social networks

- Citation anchors:
  - page=4, words=275-492, note=National data on social ties; leveraging geographic variation across the U.S.
  - page=6, words=0-349, note=Defines district homophily and states use of Facebook friendship graph
  - page=6, words=275-415, note=Identification leverages redistricting-driven changes over time

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.3420695296675894)`
- Retrieval method: `local`
- Evidence type: `textual description of methods`
- Data source: 2010 U.S. congressional redistricting; SCI-based district homophily
- Answer:

Exploit changes in county-level district homophily induced by the 2010 redistricting in an event-study design (with fixed effects), with a district-border county-pairs comparison as an alternative robustness strategy.

- Quote snippet:

  Focusing on a single redistricting event allows me to avoid concerns related to staggered treatment events

- Citation anchors:
  - page=15, words=0-45, note=Focus on single 2010 redistricting event to avoid staggered treatment and permit pre-trend tests.
  - page=15, words=120-260, note=Event-study setup around treatment timing (2012 as last pre-treatment year) and specification.
  - page=28, words=0-90, note=Alternative identification via border county pairs across district lines.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.2960513458866162)`
- Retrieval method: `local`
- Evidence type: `textual`
- Assumption flag: `False`
- Answer:

Quasi-experimental

- Quote snippet:

  changes in district homophily due to redistricting largely are not. Consequently, plausibly causal identification

- Citation anchors:
  - page=6, words=0-349, section=2 Empirical Strategy and Networks Data, note=Uses redistricting-induced changes in district homophily for plausibly causal identification (natural experiment).
  - page=15, words=0-349, section=Event study setup around 2010 redistricting, note=Event-study design leveraging a single redistricting

event to estimate causal effects.
  - page=20, words=0-349, section=Voters' Choices event studies, note=Event studies treat 2010 as base year with new district boundaries affecting behavior from 2012.
  - page=28, words=0-349, section=Border pairs design, note=Border-pairs comparison across district lines is a quasi-experimental spatial design.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.3452658951371074)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: 2010 Census redistricting (new congressional district borders for 2012 elections)
- Table/Figure: Appendix Figure C4
- Assumption flag: `False`
- Answer:

The 2010 Census–driven congressional redistricting (implemented for the 2012 elections), which induced changes in district homophily between 2012 and 2013; additionally, a border-pairs design comparing adjacent counties across district lines serves as an alternative source of variation.

- Quote snippet:

  ¯c is the change in district homophily experienced by county c between 2012 and 2013

- Citation anchors:
  - page=15, words=0-60, note=Focus on the single redistricting event following the 2010 Census
  - page=15, words=220-270, note=Defines the treatment as the change in district homophily between 2012 and 2013
  - page=28, words=0-120, note=Border-pairs identification comparing counties across district borders

B04: What is the treatment definition and timing?

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.11053936640375273)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: 2010 Census redistricting
- Table/Figure: Appendix Figure C4
- Assumption flag: `False`

- Answer:

Treatment is the change in district homophily caused by the 2010 redistricting, measured as $\Delta^{-}c$—the change a county experiences between 2012 and 2013. Timing: Census in Apr 2010; new borders used for Nov 2012 elections; representatives under new borders seated Jan 2013. The last pre-treatment year is 2012 for outcomes tied to the current representative, and 2011 (or 2010 for even-year outcomes) for upcoming-election outcomes; event studies assume 2012 as the last pre-treatment year.

- Quote snippet:

  "$^{-}$c is the change in district homophily experienced by county c between 2012 and 2013."

- Citation anchors:
  - page=15, words=0-349, note=Defines $\Delta^{-}c$ as the change in district homophily between 2012 and 2013; details timing of Census, election, seating, and last pre-treatment year(s).

B05: What is the control/comparison group definition?

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.21389700792283337)`
- Retrieval method: `local`
- Evidence type: `Study design description (event study and border-pairs)`
- Table/Figure: Appendix Table C2 (Border pairs results)
- Assumption flag: `False`
- Answer:

In the event-study, the comparison is counties with little or no change in district homophily around the 2012 redistricting, using the pre-treatment year as baseline (2012 for current-representative outcomes; 2011/2010 for election-related outcomes). In the border-pairs design, the comparison group is the neighboring county across the district border in the same pair-year (via pair-by-year fixed effects).

- Quote snippet:

  we can identify the impact of district homophily by comparing deviations from the county-pair's mean in one county to deviations from the county-pair's mean in the neighboring county.

- Citation anchors:
  - page=15, words=0-349, note=Defines last pre-treatment year(s) and event-study setup
  - page=28, words=0-349, note=Border-pairs comparison defined (neighboring county within pair-year)

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.3288873548233163)`
- Retrieval method: `local`
- Evidence type: `direct text (equations)`
- Data source: Equations (8) and (11) in text
- Assumption flag: `False`
- Answer:

Baseline specifications given in the text are:

1) Event-study (redistricting) specification: yict = t + Σ Δ¯c I(= t) + Xct + Zict + ict, with year fixed effects, errors clustered at the county level, and optional additions of district-by-year fixed effects, DMA-by-year fixed effects, and partisan exposure controls.
2) Border-pairs specification: yct = c + µpt + ¯c,t + Xct + ct, which includes county fixed effects and pair-by-year fixed effects (and, in this restricted sample, state-by-year fixed effects rather than district-by-year fixed effects).

- Quote snippet:

  The specification for this design is yct = c + µpt + ¯c,t + Xct + ct (11)

- Citation anchors:
  - page=15, words=0-349, note=Event-study estimating equation (Eq. 8) and variable definitions.
  - page=15, words=275-450, note=Optional district-by-year and DMA-by-year fixed effects; partisan exposure control.
  - page=28, words=0-349, note=Border-pairs design estimating equation (Eq. 11).

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.3400683494503729)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Appendix Table C2
- Table/Figure: Appendix Table C2
- Assumption flag: `False`
- Answer:

Included fixed effects: (1) Time: year fixed effects. (2) Unit: county fixed effects; congressional district fixed effects; media market (DMA) fixed effects.

(3) Two-way/higher-dimensional: district-by-year fixed effects; DMA-by-year fixed effects; pair-by-year fixed effects; state-by-year fixed effects. Note: in the border-pairs design, district-by-year FEs are not used; state-by-year and pair-by-year (and, in some specs, DMA-by-year) FEs are included.

- Quote snippet:

  μpt is the pair-by-year fixed effect

- Citation anchors:
  - page=page 15, words=0-349, note=Event study includes year fixed effects (t).
  - page=page 15, words=275-450, note=Mentions adding district-by-year and DMA-by-year fixed effects.
  - page=page 28, words=0-349, note=Border pairs spec: county FE (c), pair-by-year FE (pt); uses state-by-year FEs; omits district-by-year FEs.
  - page=page 14, words=275-429, note=Regression includes media market and congressional district fixed effects.
  - page=page 19, words=275-406, note=Notes results similar when county fixed effects are included.
  - page=page 65, words=0-288, section=appendix, note=Appendix Table C2 shows FEs: County, Pair x Year, State x Year, DMA x Year.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.3215341803039984)`
- Retrieval method: `local`
- Evidence type: `textual`
- Table/Figure: Table C2
- Assumption flag: `False`
- Answer:

Clustered standard errors at the county level.

- Quote snippet:

  Standard errors clustered at the county level in parentheses.

- Citation anchors:
  - page=15, words=0-349, note=Model specification states errors are clustered at the county level.
  - page=65, words=0-288, section=appendix, note=Table C2 note specifies standard errors clustered at the county level.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.33852113349453294)`
- Retrieval method: `local`
- Evidence type: `textual (methods: event-study design and pre-trend checks)`
- Assumption flag: `False`
- Answer:

Parallel trends.

- Quote snippet:

  allows for a visual test of pre-trends in changes in district homophily.

- Citation anchors:
  - page=15, words=24-46, note=Mentions a visual test of pre-trends around the 2012 redistricting
  - page=15, words=120-200, note=Introduces event-study specification (Eq. 8), which relies on parallel trends

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.26440040229632683)`
- Retrieval method: `local`
- Evidence type: `robustness checks, placebo tests, and alternative identification design`
- Data source: Appendix Section C.4 (CES + SCI zip-code networks) and Appendix Table C2
- Table/Figure: Appendix Table C2
- Assumption flag: `False`
- Answer:

The paper supports the identifying assumption with (i) robustness to constructing district homophily using zip-code–level networks—showing similar patterns, slightly larger effects, and no effects on placebo outcomes—and (ii) an alternative border-pairs specification that yields qualitatively similar results to the redistricting design (reported in Appendix Table C2). It also notes institutional/empirical constraints (e.g., most counties lie wholly in one district) that make district homophily informative about the assignment.

- Quote snippet:

  With the border pairs design, I find qualitatively similar results as in the redistricting design

- Citation anchors:
  - page=27, words=275-439, note=Zip-code–level network robustness and placebo outcomes show similar patterns and no placebo effects.
  - page=28, words=0-349, note=Border-pairs specification yields qualitatively similar results; results in Appendix Table C2.
  - page=44, words=0-349, section=appendix, note=Institutional constraints (many counties entirely in one district; ~90%) supporting informativeness of homophily.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.3265706672763601)`
- Retrieval method: `local`
- Evidence type: `textual_evidence`
- Data source: CES (Cooperative Election Study)
- Answer:

Yes. The study uses event-study designs (which also enable a visual pre-trend test). The event studies show that effects on voter knowledge emerge right after redistricting—most strongly in the 2014 survey—and remain relatively stable over time.

- Quote snippet:

  most strongly takes effect in the first survey after redistricting (2014). Impacts are relatively stable over time.

- Citation anchors:
  - page=15, words=0-349, note=States that focusing on a single redistricting allows a visual pre-trend test and lays out the event-study framework.
  - page=20, words=0-349, note=Reports event-study findings: effects appear in 2014 and are stable over time.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.13210708535614082)`
- Retrieval method: `local`
- Evidence type: `inference_from_text`
- Data source: CES surveys and Dave Leip's Election Atlas data (as described)
- Answer:

No instrumental variables are described in the provided text. Identification relies on model structure (invertibility of the Jacobian via the Inverse Function Theorem), not an IV strategy; thus no instrument or first stage is defined here.

- Quote snippet:

  By the Inverse Function Theorem, local injectivity is determined by the invertibility of the Jacobian DF (Q0s).

- Citation anchors:
  - page=45, words=275-370, section=appendix, note=Identification via Inverse Function Theorem and Jacobian invertibility, not IV.
  - page=46, words=275-372, section=appendix, note=States the map is locally identified; no mention of instruments or first-stage regressions.
  - page=32, words=0-349, note=Presents structural dynamics and steady-state expressions; no IV setup.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.30081851757688044)`
- Retrieval method: `local`
- Evidence type: `regression_table_and_placebo_text`
- Data source: CES within-border-pairs regressions (appendix)
- Table/Figure: Table C2
- Answer:

Using the within–border-pairs regressions (Table C2), approximate single-instrument first-stage F-stats (t^2 from coef/SE) in the most saturated specification (col 6) are: Heard of Incumbent F 7.9 (0.256/0.091); Selected Correct Party F 4.6 (0.343/0.160); Selected Party F 0.8 (0.136/0.151). Across specifications, F ranges are 7.4–10.6 (Heard), 4.4–14.3 (Selected Correct Party), and 0.8–6.8 (Selected Party). Partial R^2 is not reported in the provided context. Relevance evidence: statistically significant positive effects for key familiarity measures (p=0.005 for Heard; p=0.032 for Selected Correct Party in col 6), and placebo tests show no significant relationships for governor/senators, supporting relevance specific to House-representative information.

- Quote snippet:

  District Homophily 0.256*** (0.091) ... 0.136 (0.151) ... 0.343** (0.160)

- Citation anchors:
  - page=65, words=0-288, section=appendix, note=Table C2 coefficients, standard errors, and p-values for effects of District Homophily on familiarity outcomes within border pairs.
  - page=55, words=0-152, section=appendix, note=Placebo summary: district homophily does not significantly predict statewide-office outcomes.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.2982535532267215)`
- Retrieval method: `local`
- Evidence type: `textual`
- Assumption flag: `False`
- Answer:

Not applicable—the study does not employ a regression discontinuity design. It uses event-study analyses (with 2010 as the base year) and a border-pairs specification; thus no running variable, cutoff, or bandwidth is defined.

- Quote snippet:

  I run event studies analogous to equation 8 to examine the impact of district homophily on voting-related outcomes

- Citation anchors:
  - page=20, words=275-461, note=States use of event-study design (2010 base year) for voting-related outcomes.
  - page=19, words=275-406, note=Describes event studies and focus on even years for interpretation.
  - page=28, words=0-349, note=Describes border-pairs specification as an alternative identification strategy.

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

- Category: `B) Identification strategy / causal design`
- Confidence: `low (0.3241949630430817)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: 2010 Census redistricting (Nov 2012 elections)
- Assumption flag: `False`
- Answer:

No staggered adoption. The design centers on a single redistricting shock from the 2010 Census, with effects realized for the Nov 2012 elections and new representatives seated Jan 2013. The estimator is an event-study difference-in-differences that interacts the county's change in district homophily (Δ^{-}c, 2012–2013) with event-time indicators, includes year fixed effects, county-by-year and individual controls, errors clustered at the county level, and can include district-by-year fixed effects.

- Quote snippet:

  Focusing on a single redistricting event allows me to avoid concerns related to staggered treatment events… the congressional representa-

tives first elected under the new borders assumed office in January 2013.

- Citation anchors:
  - page=15, words=0-349, note=Single event avoids staggered adoption; timing (2012 elections, Jan 2013 seating) and event-study specification with FE and clustering.
  - page=19, words=275-406, note=Event-study interpretation context (even-year focus; outcome scaling).

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.3256958101891945)`
- Retrieval method: `local`
- Evidence type: `textual evidence from methods and appendix`
- Data source: Appendix B.1.1; Sections 3.1–3.2; Section 5.3; Page 8 (SCI details)
- Assumption flag: `False`
- Answer:

Cooperative Election Study (CES, formerly CCES); Meta's Social Connectedness Index (SCI); Dave Leip's Atlas of U.S. Presidential Elections (Election Atlas); precinct-level vote counts from the Harvard Election Data Archive (2000–2010) and the MIT Election Data and Science Lab (2016–2020); county-by-congressional-district vote counts from Daily Kos.

- Quote snippet:

  precinct-level vote count data from the Harvard Election Data Archive (for 2000-2010) and the MIT Election Data and Science Lab (for 2016-2020), combined with county-by-congressional-district vote count data from Dave Leip's Election Atlas (for House elections) and Daily Kos

- Citation anchors:
  - page=16, words=0-349, note=Use of Cooperative Election Study (CES) for voter information and behavior
  - page=8, words=275-624, note=Description of the Social Connectedness Index (SCI) from Meta apps
  - page=47, words=0-207, section=appendix, note=Vote count data sources: Harvard EDA, MIT Election Data and Science Lab, Dave Leip's Election Atlas, and Daily Kos
  - page=17, words=275-542, note=Use of Dave Leip's Election Atlas in turnout/vote share analysis
  - page=27, words=0-349, note=Zip-code-level construction using SCI

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.33652792509374285)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: Cooperative Election Study (CES)
- Table/Figure: Table B3: CES Data: Summary Statistics
- Assumption flag: `False`
- Answer:

individual

- Quote snippet:

  yict is the outcome for a given individual i in county c in year t

- Citation anchors:
  - page=15, words=180-260, note=Defines the outcome at the individual level: yict is for a given individual i in county c in year t.
  - page=17, words=300-370, note=States the analysis uses CES survey responses with hundreds of thousands of respondents (individuals).
  - page=49, words=0-60, section=appendix, note=Summary statistics list large numbers of observations for respondent-level variables.

C03: What is the sample period and geographic coverage?

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.2497455184946808)`
- Retrieval method: `local`
- Evidence type: `text`
- Assumption flag: `False`
- Answer:

2006–2022; the 48 contiguous U.S. states (continental U.S.), analyzed at the county level (3,136 counties).

- Quote snippet:

  I only include respondents in the 48 contiguous states... across all 17 years of the CES

- Citation anchors:
  - page=17, words=0-349, note=Geographic restriction to 48 contiguous states.
  - page=17, words=275-542, note=Coverage across 17 years of CES.
  - page=15, words=0-349, note=Event-study window includes 2006–2022.
  - page=8, words=275-624, note=County coverage: 3,136 counties.

C04: What are the sample restrictions / inclusion criteria?

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.25791566656218423)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: CES (Cooperative Election Study); Dave Leip's Election Atlas; Meta Social Connectedness Index (SCI)
- Assumption flag: `False`
- Answer:

CES analyses include only respondents in the 48 contiguous states. Exclusions: (1) respondents with missing answers to the candidate party recognition question; (2) respondents missing required demographic answers when demographic controls are used; (3) a small number in 2006–2007 assigned to counties outside their state; (4) 925 North Carolina respondents in 2020 assigned to incorrect congressional districts (shown wrong candidate names). For voting/event-study analyses, the sample focuses on even years (odd years are smaller and noisier; most voting outcomes are only available in even years). Coverage note: in even years at least one respondent is present in 80–90% of counties, about two-thirds in odd years. For SCI-based measures, the SCI includes only users who interacted with a Meta app in the prior 30 days, and areas with very small populations are removed.

- Quote snippet:

  I only include respondents in the 48 contiguous states.

- Citation anchors:
  - page=17, words=0-349, note=Includes 48 contiguous states and county coverage notes.
  - page=17, words=275-542, note=Footnote detailing exclusions (missing responses, demographic nonresponse, misassigned counties, NC 2020 district error).
  - page=19, words=275-406, note=Restriction to even years for consistency and data availability.
  - page=8, words=275-624, note=SCI inclusion criterion (recent Meta app use) and small-population area removal.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.2534699615932913)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: CES (Cooperative Election Study)
- Assumption flag: `True`
- Assumption notes: Interprets 'main analysis' as the CES-based analysis

sample; the paper also cites SCI coverage (3,136 counties; 9,834,496 county pairs) for network data.
- Answer:

612,085 respondents

- Quote snippet:

  In these 48 states and across all 17 years of the CES, there are 612,085 respondents

- Citation anchors:
  - page=17, words=275-542, note=States total CES respondents used across 17 years in 48 states

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.3102898319611432)`
- Retrieval method: `local`
- Evidence type: `Appendix variable descriptions and methods text`
- Data source: CES (Cooperative Election Study); Dave Leip's Election Atlas; Harvard Election Data Archive; MIT Election Data and Science Lab; Daily Kos
- Table/Figure: Table B1; Table B2; Table B4
- Assumption flag: `True`
- Assumption notes: Interprets the paper's primary outcomes as the CES voter knowledge/voting measures and county-level turnout/roll-off; no evidence of additional transformations provided.
- Answer:

The key outcomes are measured as binary survey indicators (e.g., Heard of Representative, Selected Party, Selected Correct Party) from the CES, and as voting outcomes that are either binary choices or turnout shares of the Voting Age Population (VAP) and a roll-off difference. No log, z-score, or index transformations are applied; estimates are interpreted as changes in probability since outcomes are binary and district homophily is scaled 0–1.

- 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, note=States outcomes are binary and district homophily is 0–1, so estimates are probability changes.
  - page=47, words=0-207, section=appendix, note=Defines voter knowledge outcomes as Binary from pre-survey.

- page=48, words=0-333, section=appendix, note=Defines voting preference/choice and validated turnout variables as Binary from pre/post surveys.
- page=49, words=0-276, section=appendix, note=Defines roll-off and turnout measures; turnout is share of VAP.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.3034837296648951)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: CES surveys; Dave Leip's Election Atlas; Harvard Election Data Archive; MIT Election Data and Science Lab; Daily Kos
- Assumption flag: `False`
- Answer:

Treatment/exposure is the change in district homophily at the county level induced by the 2010 redistricting, operationalized as $\Delta^{-}c$—the difference in a county's district homophily between 2012 (pre) and 2013 (post). Constructed variables include: (1) county exposure to Democrats (friendship-share–weighted Democratic vote shares across connected counties); (2) aggregated network measures connecting the model to data—the county–county friendship matrix Π (c,c), district population shares q(c,d), and aggregated county steady-state informed share oc(t); (3) CES voter knowledge binaries (Heard of Representative, Selected Party, Selected Correct Party); and (4) voting outcomes such as House roll-off (top-of-ticket votes minus House votes) and turnout rates as shares of VAP. Models also use district-by-year and DMA-by-year fixed effects to net out election- and media-market shocks.

- Quote snippet:

  $^{-}$c is the change in district homophily experienced by county c between 2012 and 2013

- Citation anchors:
  - page=15, words=0-349, note=Defines $\Delta^{-}c$ as the change in district homophily between 2012 and 2013 and frames treatment timing.
  - page=15, words=275-450, note=Constructed exposure to Democrats and DMA-by-year fixed effects.
  - page=33, words=0-349, note=Constructed network aggregation: Π (c,c), q(c,d), and aggregated oc(t).
  - page=47, words=0-207, section=appendix, note=Constructed CES knowledge variables: definitions for Heard/Selected/Selected Correct Party.
  - page=49, words=0-276, section=appendix, note=Constructed voting outcomes: Roll-Off and turnout measures; data sources.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.34956408360977403)`
- Retrieval method: `local`
- Evidence type: `direct text`
- Data source: Main text (p. 15) and Appendix Table C2 (p. 65)
- Table/Figure: Table C2
- Assumption flag: `False`
- Answer:

Yes. Key covariates/controls include: (a) year fixed effects (always in the event-study specs); (b) county-by-year demographic controls; and (c) individual-level controls. Additional controls used in some specifications are district-by-year fixed effects, DMA-by-year fixed effects, and a county-level Democratic exposure measure. In the within-border-pair regressions (Table C2), County and Pair×Year fixed effects are included in every column, with State×Year and DMA×Year added in some columns.

- Quote snippet:

  > t are year fixed effects, Xct is a vector of county-by-year controls, and Zict is a vector of individual controls. I can additionally include district-by-year fixed effects… include DMA-by-year fixed effects.

- Citation anchors:
  - page=15, words=0-349, note=Event-study spec: year FEs, county-by-year controls (Xct), individual controls (Zict); clustering.
  - page=15, words=275-450, note=Optional additions: district-by-year FEs, DMA-by-year FEs, Democratic exposure control.
  - page=65, words=0-288, section=appendix, note=Within border-pair regressions: FE structure across columns; definitions of individual and county-by-year controls.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.31647121514709664)`
- Retrieval method: `local`
- Evidence type: `text_from_document`
- Data source: CES; Dave Leip's Election Atlas; Harvard Election Data Archive; MIT Election Data and Science Lab; Daily Kos
- Table/Figure: Table B2
- Assumption flag: `False`
- Answer:

Missing/invalid observations are excluded rather than imputed. Specifically:

(1) survey cases with missing candidate-party recognition responses are dropped; respondents missing demographic answers are also excluded when those controls are used; misassigned county/district cases (including 925 NC respondents in 2020) are removed. (2) Several voting variables are coded missing when there is no incumbent running, or when both validated voting indicators are missing. (3) In vote-count data, elections with no top-of-ticket race and cases where turnout exceeds VAP are excluded. Weighting: CES cumulative weights are used to re-weight observations across years; weights do not adjust for pre/post attrition. No imputation is described.

- Quote snippet:

  I exclude missing responses to the candidate party recognition question (<2% of respondents in each year...). When including individual demographic controls, I similarly exclude respondents who did not answer...

- Citation anchors:
  - page=17, words=275-542, note=Drops missing candidate-party recognition responses; excludes respondents missing demographics; removes misassigned county/district cases; weights do not account for attrition.
  - page=16, words=0-349, note=Uses CES cumulative weights to re-weight observations across years.
  - page=48, words=0-333, section=appendix, note=Voting variables set to missing if no incumbent; post-survey voting variables missing if both validated indicators are missing.
  - page=49, words=0-276, section=appendix, note=Excludes elections with no top-of-ticket race and cases where turnout exceeds VAP.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `medium (0.3628265093281664)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: CES (Cooperative Election Study) survey weights
- Assumption flag: `False`
- Answer:

Yes. CES survey weights are used so the respondent sample represents people living in the U.S.; however, the weights do not adjust for attrition between the pre- and post-election surveys.

- Quote snippet:

  Because the weighted sample is representative of people living in the U.S. (rather than of U.S. counties)

- Citation anchors:
  - page=17, words=270-349, note=States that the weighted sample is representative of U.S. residents.
  - page=20, words=430-461, note=Notes that weights do not account for attrition between survey waves.

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.30299321978219823)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Facebook Social Connectedness Index (SCI); county–district population shares q(c,d); commuting flows; CES respondent zip codes; county-level vote/contribution data.
- Assumption flag: `False`
- Answer:

Yes. The paper links Facebook SCI network data to congressional districts by aggregating county–county friendship shares and weighting by county–district population shares. Specifically, it builds c,k from SCI, then computes district homophily as Σd D(c) Σk c,k × q(c,d) × q(k,d), where q(c,d) is the share of county c's population in district d. For robustness, the same aggregation is done using commuting flows. At the zip level, SCI is linked to CES respondents via reported zip codes. County-level SCI is used to match with county-level vote counts and campaign contributions.

- Quote snippet:

  > I do this by using the SCI to construct the Π matrix of county-county friendship shares, and then for each county summing friendship shares across same-district counties (adjusting for counties that intersect multiple districts).

- Citation anchors:
  - page=7, words=0-349, section=2.1.1 Definition of District Homophily, note=Defines district homophily and introduces population-weighted linkage via q(c,d).
  - page=9, words=0-257, section=2.1.3 Construction of District Homophily from SCI, note=Explains aggregating SCI to Π and summing within-district shares; constructs c,k from SCI.
  - page=8, words=0-349, section=2.1.2 Proxy for Social Networks: Facebook Social Connectedness Index, note=Defines SCI and notes county focus to facilitate matching to county-level outcomes.
  - page=27, words=0-349, section=5.3 Zip-Code-Level Social Network Data, note=Links zip-level SCI to CES via respondents' reported zip codes.

- page=33, words=0-349, section=6.7 Aggregating to County-Level, note=Details aggregation using county–district population shares and county-county friendship probabilities.

C12: What summary statistics are reported for main variables?

- Category: `C) Data, sample, and measurement`
- Confidence: `medium (0.3757044005327591)`
- Retrieval method: `local`
- Evidence type: `text and tables`
- Table/Figure: Table B3: CES Data: Summary Statistics; Table B5: Voting Outcomes: Summary Statistics; Table C1: Placebo Outcomes Summary Statistics
- Assumption flag: `False`
- Answer:

Summary statistics reported include: (1) for district homophily—mean, standard deviation, minimum, maximum, 1st and 99th percentiles, and ranges for the middle 50% and 80% of counties; (2) for CES knowledge/voting and county voting outcomes—number of observations, mean (percent), and standard deviation (percentage points).

- Quote snippet:

  > mean district homophily is 41% with a standard deviation of 14pp; minimum district homophily is 2% and maximum is 87%, while the 1st percentile is 8% and the 99th percentile is 67%

- Citation anchors:
  - page=13, words=0-349, note=District homophily summary statistics: mean, SD, min, max, percentiles, middle ranges.
  - page=49, words=0-276, section=appendix, note=Tables B3 and B5 show Observations, Mean (%), SD (pp) for CES and voting outcomes.
  - page=55, words=0-152, section=appendix, note=Table C1 (placebo outcomes) also uses Observations, Mean (%), SD (pp).

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

- Category: `C) Data, sample, and measurement`
- Confidence: `low (0.33953190147297013)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: Paper appendix (figures cited)
- Table/Figure: Appendix Figures C1–C2; Figures C4–C5
- Assumption flag: `False`
- Answer:

Yes. Appendix Figures C1–C2 provide descriptive figures on how geographic and demographic features correlate with district homophily (2012 and 2020), and Appendix Figures C4–C5 map the nationwide distribution of changes.

- 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=13, words=0-349, note=Mentions Appendix Figures C1–C2 summarizing correlations in 2012 and 2020.
  - page=15, words=0-349, note=States Appendix Figure C4 provides a map of changes.
  - page=53, words=0-40, section=appendix, note=Notes that Figures C4–C5 illustrate how changes are distributed around the country.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `medium (0.3668964869451069)`
- Retrieval method: `local`
- Evidence type: `text (main text)`
- Table/Figure: Figure 8
- Answer:

Positive: a 10pp increase in district homophily raises the share of contributions to in-district candidates by 7.4pp.

- Quote snippet:

  In particular, a 10pp increase in district homophily is associated with a 7.4pp increase in the share of contributions to in- district candidates, from a mean of 51%.

- Citation anchors:
  - page=24, words=0-229, note=Campaign contributions main effect

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.19272777411442193)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Facebook Social Connectedness Index (SCI)
- Table/Figure: Figure 1 (page 9)
- Assumption flag: `False`
- Answer:

The preferred specification constructs district homophily using the Facebook Social Connectedness Index (SCI), because SCI is a strong proxy for real-world social networks, whereas commuting flows are more geographically concentrated and tend to overstate homophily.

- Quote snippet:

  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, note=States the use of SCI as the data source for social networks and justifies it as a top proxy.
  - page=9, words=0-257, note=Explains that commuting flows are more concentrated, leading to higher commuting-based homophily than SCI.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.33059979194916095)`
- Retrieval method: `local`
- Evidence type: `Reported quantitative estimates from event-study analyses in the paper`
- Data source: CES survey; Dave Leip's Election Atlas (county-level votes)
- Table/Figure: Figure 6 (Roll-Off); Figure 8 (Contributions); Table B3 (CES summary stats)
- Assumption flag: `True`
- Assumption notes: Interprets effects per 10pp assuming linear scaling and computes relative percent changes using reported means; adopts paper's stated 10% roll-off reduction and 2pp VAP-based roll-off change.
- Answer:

A 10pp rise in district homophily produces modest but nontrivial shifts: voter knowledge rises by 0.7pp (0.75% relative) for name recognition, 3.2pp (4.7%) for picking a party, and 3.3pp (5.3%) for picking the correct party. It reduces ballot roll-off by about 10% (and by about 2pp when roll-off is constructed using votes/VAP). For money, it reallocates giving toward local races: the in-district share of House contributions rises by 7.4pp from a 51% mean (~14.5% relative), with no change in total donations.

- Quote snippet:

  a 10pp increase in district homophily is associated with a 7.4pp increase in the share of contributions to in-district candidates, from a mean of 51%.

- Citation anchors:

  - page=20, words=0-349, note=Knowledge effects per 10pp increase and baseline means
  - page=24, words=0-229, note=Roll-off reduction (~10%) and contributions shift (+7.4pp from 51%)
  - page=67, words=0-258, section=appendix, note=Alternative roll-off measure shows ~2pp decrease; no change in turnout levels
  - page=49, words=0-276, section=appendix, note=Baseline means for knowledge outcomes (Table B3)

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.3283513759686765)`
- Retrieval method: `local`
- Evidence type: `mixed`
- Data source: CES (Cooperative Election Study)
- Table/Figure: Table C2: Effect of District Homophily on Voter Familiarity with Representative, within Border Pairs
- Assumption flag: `False`
- Answer:

Key checks include: (1) adding district-by-year fixed effects to absorb election-specific shocks; (2) adding DMA-by-year fixed effects to rule out media-market confounds; (3) controlling for partisan bias in network ties via each county's exposure to Democrats; (4) using state-by-year fixed effects; (5) exploiting a within–border-pair design with county and pair-by-year fixed effects; (6) adding individual demographic controls (gender, race, education, age, co-partisanship) and (7) adding county-by-year demographics (population, race/age/gender shares, urban share); and (8) placebo tests on governors and senators. Results largely survive: the effects on 'Heard of Incumbent' and 'Selected Correct Party' remain positive and statistically significant across specifications with these controls, while the weaker 'Selected Party' measure loses significance. Placebo outcomes show no significant relationship, supporting robustness.

- Quote snippet:

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

- Citation anchors:
  - page=15, words=275-450, note=Describes adding district-by-year and DMA-by-year fixed effects; controls for partisan exposure to Democrats.
  - page=65, words=0-288, section=appendix, note=Table C2 shows robustness across County & Pair×Year, State×Year, DMA×Year FEs; adds Democratic exposure, individual and county-year controls; significance for Heard of Incumbent and Selected Correct Party.

  - page=55, words=0-152, section=appendix, note=Placebo tests summary: district homophily does not significantly predict governor/senator outcomes.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.18274002630391975)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Cooperative Election Study (CES)
- Table/Figure: Appendix Table C1
- Assumption flag: `False`
- Answer:

Placebo tests replicate the knowledge measures for statewide offices unaffected by congressional district borders: for the governor and each U.S. senator, indicators for heard of, selected party, and selected correct party (nine outcomes). These tests show no significant effect of district homophily on any of the placebo outcomes.

- Quote snippet:

  I do not find evidence that district homophily increases voters' knowledge on placebo outcomes (i.e., the same three outcomes but for the respondent's governor and senators).

- Citation anchors:
  - page=26, words=0-170, note=Describes placebo outcomes for governor and senators and states no significant impact.
  - page=55, words=0-152, section=appendix, note=Appendix summary reiterating that district homophily does not significantly predict placebo outcomes.
  - page=20, words=0-120, note=Mentions lack of evidence for impacts on placebo outcomes (same three outcomes for governor and senators).

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.275338042494289)`
- Retrieval method: `local`
- Evidence type: `textual evidence`
- Data source: Cooperative Election Study (CES)
- Table/Figure: Table C1: CES Data: Summary Statistics for Placebo Outcomes
- Answer:

Placebo (falsification) outcomes are voters' knowledge measures for statewide offices: for the governor and for each of the two U.S. senators—whether the respondent has heard of them, selects a party, and selects the correct party (nine outcomes total).

- Quote snippet:

  > I test whether district homophily impacts voters' knowledge of their governor and senators... I find no significant impact of district homophily on these nine outcomes.

- Citation anchors:
  - page=26, words=0-220, note=Defines placebo tests: governor and senators; outcomes are heard of, select party, select correct party; reports no significant impact.
  - page=55, words=0-120, section=appendix, note=Appendix C.2 summary and Table C1 list the nine placebo outcomes and note no significant prediction by district homophily.
  - page=20, words=120-260, note=States they do not find evidence that district homophily increases knowledge on placebo outcomes for governor and senators.
  - page=16, words=170-320, note=CES asks about House representative, both senators, and governor—basis for constructing analogous placebo outcomes.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `medium (0.37521798394600775)`
- Retrieval method: `local`
- Evidence type: `textual`
- Assumption flag: `False`
- Answer:

Reported heterogeneity pertains to district homophily levels rather than treatment-effect splits: (1) Income: With media-market and district fixed effects, higher white non-Hispanic share correlates positively with changes in homophily, while higher poverty share correlates negatively. (2) Size: Larger counties (by population) have lower homophily—each 1% population increase predicts a 0.05pp decrease; urban areas are more often split. (3) Baseline exposure: Homophily is higher where residents are farther from district borders/areas with geographically larger districts; authors also control for social-network partisanship exposure in outcome models. (4) Region: Counties in single-district states have higher average homophily (53%), consistent with networks following state boundaries; rural areas show higher homophily than urban areas.

- Quote snippet:

  Counties in single district states have higher district homophily on average (53%)… a one percent increase in county population is associated with a 0.05pp decrease in district homophily.

- Citation anchors:
  - page=13, words=0-349, note=Size, urban/rural patterns; single-district states; population-population effect on homophily; distance/border logic.
  - page=14, words=275-429, note=Income and demographic correlations with changes in district homophily; controls used.
  - page=19, words=0-200, note=Models include control for partisanship of the social network (baseline exposure control).

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.1494383940132082)`
- Retrieval method: `local`
- Evidence type: `Textual summary of figure/caption results and appendix placebo tests`
- Data source: CES (roll-off); county donations to House candidates (contribution shares)
- Table/Figure: Figure 6 and Figure 8
- Assumption flag: `False`
- Answer:

Two mechanism tests are reported: (1) Ballot roll-off: higher district homophily reduces House roll-off (e.g., a 10pp increase cuts roll-off by about 0.04pp, 10%), indicating greater down-ballot participation. (2) Campaign contributions: a 10pp increase in district homophily raises the share of county donations going to in-district House candidates by 7.4pp (from a 51% mean) with no change in total House donations, implying a reallocation toward local candidates rather than more overall giving. Placebo tests for governors and senators show no significant effects, consistent with a district-specific information/salience mechanism.

- Quote snippet:

  district homophily reduces roll-off; a 10pp increase raises the in-district contribution share by 7.4pp, with no impact on total donations.

- Citation anchors:
  - page=24, words=0-229, note=Roll-off reduction and in-district contribution share increase; no change in total donations
  - page=55, words=0-152, section=appendix, note=Placebo outcomes show no significant effects for governor/senators

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.3485816104887248)`
- Retrieval method: `local`
- Evidence type: `Appendix table and narrative robustness text`
- Data source: CES; Nielsen DMA boundaries; 2016 5-Year ACS County-County Commuting Flows; Placebo tests (Appendix C.2)
- Table/Figure: Appendix Table C2
- Assumption flag: `True`
- Assumption notes: Bandwidth sensitivity (if any) is not discussed in the provided excerpts.
- Answer:

The main findings are robust across alternative samples and control sets. Estimates are similar with county fixed effects and when including odd years (though odd-year coefficients are noisier). Results remain after adding district-by-year and DMA-by-year fixed effects and controlling for Democratic exposure, demographics, and county-year covariates; the 'heard of incumbent' and 'correct party' effects stay positive and significant, while the 'selected party' (any party) effect attenuates and loses precision in the within-border-pairs design. Placebo tests for governors/senators show no effects, and using commuting flows as an alternative network measure yields qualitatively similar effects at about half the magnitude. Effects appear quickly post-redistricting and are relatively stable over time.

- Quote snippet:

  Impacts are relatively stable over time.

- Citation anchors:
  - page=19, words=275-406, note=Similar results with county fixed effects; odd years yield similar results but noisier estimates.
  - page=15, words=275-450, note=Controls: district-by-year FEs, DMA-by-year FEs, and Democratic exposure.
  - page=65, words=0-288, section=appendix, note=Table C2 shows robustness across added controls; 'Heard of Incumbent' and 'Selected Correct Party' remain significant; 'Selected Party' attenuates.
  - page=26, words=0-349, note=Placebo outcomes show no significant impacts; commuting flows robustness with similar direction, ~half the magnitude.
  - page=55, words=0-152, section=appendix, note=Placebo tests for governor/senators: generally no significant prediction by district homophily.
  - page=20, words=0-349, note=Effects emerge in first post-redistricting survey and are stable over time.

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

- Category: `D) Results, magnitudes, heterogeneity, robustness`
- Confidence: `low (0.2241474231421388)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: Social Connectedness Index (SCI)
- Assumption flag: `False`
- Answer:
- Evidence points toward effects concentrating support towards in-district House candidates and away from out-of-district candidates.
- Calls for future research on how candidates adjust campaign strategies to district homophily, aiming for a general-equilibrium view of social-learning effects from district boundaries.
- Policy implication: newly available social network data (e.g., SCI) can help draw fairer districts but could also be exploited for partisan gerrymandering.
- Quote snippet:

  > towards House candidates in their own districts, away from House candidates running in other districts... enabling its use by policymakers to draw fairer districts, but also by partisan gerrymanderers

- Citation anchors:
  - page=36, words=0-309, section=Conclusion/Implications, note=Concluding implications about in-district tilt, research agenda, and policy risks/uses of SCI.

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

- Category: `E) Citations and related literature`
- Confidence: `low (0.33105038096722883)`
- Retrieval method: `local`
- Evidence type: `textual synthesis from introduction, literature review, methods, and references`
- Assumption flag: `False`
- Answer:

Key prior work falls into three pillars that the paper builds on: (1) Redistricting fairness measures that largely ignore turnout/network effects—Stephanopoulos & McGhee (2015) and McCartan & Imai (2023)—which the paper critiques

by incorporating social-network-driven information and turnout responses; (2) Social learning and political networks showing information spreads through ties and calling for causal estimates—foundationally Lazarsfeld et al. (1944); evidence on diffusion in networks (Conley & Udry 2010; Banerjee et al. 2013; Beaman et al. 2021); and explicit calls for causal work (Fowler et al. 2011)—motivating the paper's causal designs and its focus on district homophily; (3) Empirical and experimental studies of peer influence and information transmission—lab/field evidence (Klar & Shmargad 2017; Druckman et al. 2018; Fafchamps et al. 2019; Arias et al. 2019) and turnout contagion/peer effects (Gerber et al. 2008; Sinclair et al. 2012; Klofstad 2007; Nickerson 2008; Pons 2018), plus place effects (Cantoni & Pons 2022; Brown et al. 2023)—which ground the mechanisms linking networks to knowledge and participation. Methodologically, the paper's border-pairs identification follows Spenkuch & Toniatti (2018). It also leverages national social-tie data in the spirit of Alt et al. (2022) and demonstrates large-scale social influence akin to Bond et al. (2012). Conceptually, its diffusion model aligns with network diffusion under homophily (Jackson & López-Pintado 2013) and segregation measures (Echenique & Fryer 2007). Finally, the knowledge–turnout link (Snyder & Strömberg 2010) underpins why increased information from higher district homophily can shift participation.

- Quote snippet:

  I build on this work by highlighting the role of social networks as a key source of information for voters.

- Citation anchors:
  - page=2, words=0-349, note=Frames critique of redistricting fairness measures; positions social networks and knowledge-turnout link.
  - page=4, words=0-349, note=Literature on social networks, calls for causal estimates, media avoidance; experimental/peer-effects studies.
  - page=4, words=275-492, note=Use of national social-tie data; links to Alt et al. (2022) and Bond et al. (2012).
  - page=28, words=0-349, note=Border-pairs design referencing Spenkuch & Toniatti (2018) and conceptual diffusion framework.
  - page=37, words=0-331, note=References for Bond (2012), Cantoni & Pons (2022), Brown et al. (2023), and other peer-effects works.
  - page=38, words=0-326, note=References including Echenique & Fryer (2007) and several diffusion/political info studies.
  - page=39, words=0-327, note=References including Jackson & López-Pintado (2013) and McCartan & Imai (2023).
  - page=40, words=0-303, note=References including Snyder & Strömberg (2010) and additional turnout/peer influence works.

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

- Category: `E) Citations and related literature`

- Confidence: `low (0.30810576278123053)`
- Retrieval method: `local`
- Evidence type: `textual citations from the paper's framing`
- Assumption flag: `False`
- Answer:

Most directly, it builds on work using large-scale social network data to study political behavior and diffusion (Alt et al. 2022; Bond et al. 2012) and on place/peer-effects studies (Cantoni & Pons 2022; Brown et al. 2023). It also extends strategic gerrymandering models by incorporating social learning effects (Owen & Grofman 1988; Friedman & Holden 2008, 2020; Gul & Pesendorfer 2010; Kolotilin & Wolitzky 2020; Bouton et al. 2023).

- Quote snippet:

  I build on these studies by estimating how social network structure impacts voter knowledge… Second, I bridge the literature on social learning with the literature on models of political geography

- Citation anchors:
  - page=4, words=275-492, section=Main text, note=Author states they build on prior studies, citing Alt et al. (2022), Bond et al. (2012), and positioning relative to Cantoni & Pons (2022) and Brown et al. (2023).
  - page=5, words=0-349, section=Main text, note=Author bridges social learning with political geography/redistricting models and lists canonical gerrymandering papers.
  - page=5, words=275-475, section=Main text, note=Notes implications once social learning is considered and references broader gerrymandering measurement work.

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

- Category: `E) Citations and related literature`
- Confidence: `low (0.2745280798867526)`
- Retrieval method: `local`
- Evidence type: `textual citation`
- Data source: McCartan et al. (2021) 50-State Redistricting Simulations
- Assumption flag: `False`
- Answer:

The results benchmark against the 50-State Redistricting Simulations by McCartan, Kenny, Simko, Kuriwaki, et al. (2021), comparing features of actual maps to the distribution across 5,000 simulated maps per state.

- Quote snippet:

  McCartan, Kenny, Simko, Kuriwaki, et al. 2021 simulate 5,000 con-

gressional district maps for each of the 50 states... enabling... comparing a feature of interest of a given map against the distribution

- Citation anchors:
  - page=34, words=275-379, section=main text, note=Describes using McCartan et al. (2021) simulated maps to compare features against a distribution.
  - page=39, words=0-327, section=references, note=Bibliographic entry for McCartan et al. (2021) 50-State Redistricting Simulations.

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

- Category: `E) Citations and related literature`
- Confidence: `low (0.3308023899105018)`
- Retrieval method: `local`
- Evidence type: `direct textual evidence from methods/data descriptions`
- Data source: Facebook Social Connectedness Index (Oct 2021, county–county and zip–zip); CES (2006–2022); Commuting flows; Dave Leip's Atlas of U.S. Presidential Elections; Pew Research Center usage surveys
- Assumption flag: `False`
- Answer:

Cited datasets and uses: (1) Facebook Social Connectedness Index (SCI): used as the main proxy for real-world social networks to construct district homophily, primarily with U.S. county–county pairs from the Oct 2021 snapshot; also constructed at the zip code (ZCTA) level for robustness. (2) Commuting flows: used as an alternative proxy to construct district homophily; strongly correlated with SCI-based measures and yields qualitatively similar results. (3) Cooperative Election Study (CES): nationally representative surveys (2006–2022) used to measure voters' information about representatives and self-reported voting/preferences; CES county and district identifiers enable linkage to homophily; CES zip codes allow zip-level robustness checks. (4) Dave Leip's Atlas of U.S. Presidential Elections: used for vote count data to study actual voting behavior. (5) Pew Research Center surveys on Facebook usage (Auxier & Anderson 2021; Vogels et al. 2021): cited to document that Facebook usage rates are relatively even across demographic groups and parties, supporting SCI's suitability.

- Quote snippet:

  I use the SCI for U.S. county-county pairs from the October 2021 snapshot.

- Citation anchors:
  - page=7, words=275-379, section=2.1.2 Proxy for Social Networks: Facebook Social Connectedness Index, note=Introduces SCI as proxy for social networks aggregated from Facebook friendships.
  - page=8, words=0-349, section=2.1.2 Proxy for Social Networks:

Facebook Social Connectedness Index, note=Defines SCI, states use of Oct 2021 county–county pairs; mentions commuting flows as alternative and correlation.
  - page=8, words=275-624, section=2.1.2 Proxy for Social Networks: Facebook Social Connectedness Index, note=Details SCI snapshot construction and cites Pew evidence on even Facebook usage across demographics and parties.
  - page=16, words=0-349, section=3 Outcomes Data; 3.1 Voters' Information, note=Describes CES data, years, content, and linkage to county/district for voter knowledge outcomes.
  - page=17, words=275-542, section=Outcomes and vote data linkage, note=Cites Dave Leip's Atlas; explains use of CES pre/post election surveys for voting outcomes.
  - page=27, words=0-349, section=5.3 Zip-Code-Level Social Network Data, note=Explains construction of district homophily using SCI at the zip (ZCTA) level and use of CES zip codes.

E05: What methodological or econometric references are cited (e.g., DiD, IV, RDD methods)?

- Category: `E) Citations and related literature`
- Confidence: `medium (0.3997880642088801)`
- Retrieval method: `local`
- Evidence type: `references/citations`
- Assumption flag: `False`
- Answer:

Cited methodological/econometric references include: Gourieroux, Monfort, & Renault (1993) on indirect inference; Sinclair, McConnell, & Green (2012) on detecting spillover effects in multilevel experiments; McCartan & Imai (2023) on sequential Monte Carlo sampling for redistricting plans; McCartan et al. (2021) on 50-state redistricting simulations; McCormick, Salganik, & Zheng (2010) on estimating personal network size; Echenique & Fryer (2007) on a segregation measure based on social interactions; Snyder & Strömberg (2010) on press coverage and political accountability (used for identification strategies); and Spenkuch & Toniatti (2018) on political advertising and election results (informing a border-pairs/border-discontinuity design).

- Quote snippet:

  Gourieroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. Journal of Applied Econometrics, 8 (S1), S85–S118.

- Citation anchors:
  - page=38, words=0-326, section=References, note=Gourieroux, Monfort, & Renault (1993). Indirect inference (econometric method).
  - page=40, words=0-303, section=References, note=Sinclair, McConnell, & Green (2012). Detecting Spillover Effects: design and

analysis of multilevel experiments.
  - page=39, words=0-327, section=References, note=McCartan & Imai (2023). Sequential Monte Carlo for sampling balanced and compact redistricting plans.
  - page=39, words=0-327, section=References, note=McCartan et al. (2021). 50-State Redistricting Simulations (simulation methodology).
  - page=40, words=0-303, section=References, note=McCormick, Salganik, & Zheng (2010). Estimating personal network size (network measurement method).
  - page=38, words=0-326, section=References, note=Echenique & Fryer (2007). A Measure of Segregation Based on Social Interactions (measurement/statistic).
  - page=40, words=0-303, section=References, note=Snyder & Strömberg (2010). Press Coverage and Political Accountability (identification using media-market congruence).
  - page=28, words=0-349, section=Border pairs specification, note=Explicit use of Spenkuch & Toniatti (2018) to justify border-pairs design.
  - page=40, words=0-303, section=References, note=Spenkuch & Toniatti (2018). Political Advertising and Election Results* (empirical design at media-market borders).

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

- Category: `E) Citations and related literature`
- Confidence: `low (0.27354932674976956)`
- Retrieval method: `local`
- Evidence type: `text`
- Assumption flag: `False`
- Answer:

Yes. The paper challenges classic strategic gerrymandering models and measures that assume voter distributions and decisions are independent of district maps, citing Owen & Grofman (1988), Friedman & Holden (2008, 2020), Gul & Pesendorfer (2010), Kolotilin & Wolitzky (2020), and Efficiency Gap work (Stephanopoulos & McGhee, 2015).

- Quote snippet:

  > Existing models of strategic partisan redistricting … assume that changes to district boundaries do not affect the distribution of partisans… Voters' decisions are independent of the district map.

- Citation anchors:
  - page=5, words=0-349, section=main text, note=States existing redistricting models assume voter distributions and decisions are independent of the map; lists classic references and contrasts with findings.

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

- Category: `E) Citations and related literature`
- Confidence: `low (0.3061721700291295)`
- Retrieval method: `local`
- Evidence type: `appendix and data repository citations`
- Data source: Appendix B; Dave Leip's Election Atlas; Harvard Election Data Archive; MIT Election Data and Science Lab; Daily Kos; U.S. Census; CES datasets (DOIs in references).
- Table/Figure: Table B4: Descriptions for Voting Outcome Variables
- Assumption flag: `False`
- Answer:

Yes. The paper cites essential appendices (Appendix B) detailing data construction and variables, and references multiple data repositories used in the analysis (Dave Leip's Election Atlas, Harvard Election Data Archive, MIT Election Data and Science Lab, Daily Kos, Census, and CES datasets with DOIs). No explicit code repository is cited in the provided text.

- Quote snippet:

  > combined with county-by-congressional-district vote count data from Dave Leip's Election Atlas (for House elections) and Daily Kos (for President, Senator, and Governor elections).

- Citation anchors:
  - page=47, words=0-207, section=appendix, note=Appendix B describes construction of vote count measures using Harvard Election Data Archive, MIT Election Data and Science Lab, Dave Leip's Atlas, and Daily Kos.
  - page=49, words=0-276, section=appendix, note=Appendix B tables and variable descriptions cite data sources for roll-off and turnout (Dave Leip's Atlas, Harvard Election Data Archive, Daily Kos, MIT Election Data and Science Lab, Census).
  - page=17, words=275-542, note=Main text cites Dave Leip's Election Atlas and CES survey data; references CES methodology and exclusions.
  - page=40, words=0-303, section=references, note=References include CES 2020 and 2022 datasets with DOIs (data repository citations).
  - page=39, words=0-327, section=references, note=References include Kuriwaki (2018) Cumulative CES Common Content (Dataverse DOI) and 50-State Redistricting Simulations (data resource DOI).

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

- Category: `E) Citations and related literature`

- Confidence: `low (0.30376113152112244)`
- Retrieval method: `local`
- Evidence type: `Direct textual evidence from introduction and literature review`
- Data source: Provided context excerpts (pages 2, 4–5)
- Assumption flag: `False`
- Answer:

They highlight three main gaps: (1) a lack of causal estimates on how social networks affect political knowledge, especially the causal role of district boundaries aligning (or not) with social ties; (2) existing peer-effects and mobilization studies often focus on turnout or persuasion and many are lab-based or outside the U.S., leaving information transmission within networks and nationwide U.S. networks underexplored; and (3) gerrymandering models and fairness metrics assume voter distributions and turnout are invariant to district maps, overlooking social learning and the possibility that borders change participation by grouping social ties.

- Quote snippet:

  > 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=0-349, section=Introduction, note=States causal estimates of network effects on political knowledge are limited and the role of political boundaries in social learning has not been causally explored.
  - page=4, words=0-349, section=Literature review, note=Notes calls for more causal estimates in social-network political information literature; discusses lab and field studies.
  - page=4, words=275-492, section=Literature review, note=Says related studies focus on turnout/peer effects and not directly on information transmission; emphasizes leveraging national U.S. social ties.
  - page=5, words=0-349, section=Contribution/Models of gerrymandering, note=Argues existing gerrymandering models and measures assume fixed partisan distributions and ignore social networks' effects on turnout.

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

- Category: `F) Replication and transparency`
- Confidence: `low (0.21170377402826074)`
- Retrieval method: `local`
- Evidence type: `textual`
- Assumption flag: `False`

- Answer:

No. The provided context does not mention any replication files or code or where to access them.

- Quote snippet:

  B Data Descriptions

- Citation anchors:
  - page=47, words=0-25, section=appendix, note=Appendix lists data and variable descriptions; no replication files or code referenced.
  - page=1, words=372-388, note=Acknowledgments section ends without mentioning replication files or code.

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

- Category: `F) Replication and transparency`
- Confidence: `low (0.25252882149704253)`
- Retrieval method: `local`
- Evidence type: `textual (no mention in provided excerpts)`
- Answer:

No. The provided excerpts do not mention a pre-analysis plan or any trial/registry; no registration details are given.

- 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, note=Methods and event-study setup discussed; no pre-analysis plan or registration mentioned.
  - page=48, words=0-333, section=appendix, note=Appendix defines variables and outcomes; no pre-registration details provided.

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

- Category: `F) Replication and transparency`
- Confidence: `low (0.2639892199729601)`
- Retrieval method: `local`
- Evidence type: `textual evidence (privacy-related data limitations)`
- Data source: Meta Social Connectedness Index (SCI)
- Assumption flag: `True`
- Assumption notes: Conclusion based only on provided excerpts; other parts of the document may discuss access terms not shown here.

- Answer:

No explicit restricted-access, proprietary, or NDA requirements are disclosed. The text only notes privacy-related limitations to the SCI (small-population areas removed, noise added; some zip codes omitted for privacy).

- Quote snippet:

  due to privacy concerns, zip codes with small populations do not appear in the SCI data.

- Citation anchors:
  - page=8, words=275-624, note=SCI removes small-population areas and adds noise to preserve privacy.
  - page=27, words=0-349, note=SCI omits some zip codes due to privacy concerns.

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

- Category: `F) Replication and transparency`
- Confidence: `low (0.2328718636910073)`
- Retrieval method: `local`
- Evidence type: `methods and appendix text`
- Data source: Harvard Election Data Archive; MIT Election Data and Science Lab; Dave Leip's Election Atlas; Daily Kos; Meta Social Connectedness Index
- Table/Figure: Tables B1–B5; Figure 1 reference (methods text)
- Assumption flag: `False`
- Answer:

Yes. The document explicitly details how key datasets are constructed (e.g., district homophily from SCI; county-by-CD vote measures) and records key cleaning steps and exclusions for survey data and voting outcomes.

- Quote snippet:

  I exclude missing responses to the candidate party recognition question (<2% of respondents in each year; ...). Lastly, in the 2020 survey, 925 respondents in North Carolina were assigned to incorrect congressional districts...

- Citation anchors:
  - page=9, words=0-257, note=Construction of district homophily from SCI with aggregation procedure and formulas.
  - page=47, words=0-207, section=appendix, note=Construction of vote count measures and CES variable descriptions.
  - page=48, words=0-333, section=appendix, note=Definitions of voting preference and vote-choice variables.
  - page=49, words=0-276, section=appendix, note=Summary statistics and precise definitions for turnout and roll-off with data sources.

  - page=17, words=275-542, note=Survey data cleaning: explicit exclusions for missing responses, demographic nonresponse, misassignments.
  - page=8, words=275-624, note=SCI data construction choices (scaling, privacy noise, user inclusion, location assignment).

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

- Category: `F) Replication and transparency`
- Confidence: `low (0.271322823909785)`
- Retrieval method: `local`
- Evidence type: `narrative description in main text and appendix`
- Table/Figure: Table C1 (Appendix C.2)
- Assumption flag: `False`
- Answer:

Fully reported.

- 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=5 Robustness, note=Dedicated Robustness section outlines multiple checks (placebo outcomes, alternative homophily measure).
  - page=55, words=0-152, section=appendix, note=Appendix C.2 reports placebo outcome tests and notes generally null effects, indicating reported robustness results.

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

- Category: `G) External validity and generalization`
- Confidence: `low (0.2854722064950393)`
- Retrieval method: `local`
- Evidence type: `Study design and data coverage details`
- Data source: Cooperative Election Study (CES); Meta Social Connectedness Index (SCI)
- Assumption flag: `False`
- Answer:

Adults in the continental United States, across counties and congressional districts, especially in U.S. House election and post-redistricting settings. Generalizability is strongest to U.S. voters because the study links a nationally representative CES sample to counties/districts and uses nationwide social-network

measures (SCI) that cover all 3,136 counties, with Facebook usage relatively even across most demographic and geographic groups. Findings also likely extend to settings where non–social-media proxies for ties (e.g., commuting flows) capture similar network structure.

- Quote snippet:

  The CES is a nationally representative survey that has run annually from 2006 to 2022

- Citation anchors:
  - page=4, words=275-492, section=main, note=Uses data from across the continental U.S.; leverages geographic variation in U.S. social networks; relevance to redistricting.
  - page=8, words=275-624, section=main, note=SCI coverage (3,136 counties; nationwide pairs) and relatively even Facebook usage across demographic and geographic groups.
  - page=16, words=0-349, section=main, note=CES is nationally representative; large samples; linked to county and congressional district.
  - page=20, words=0-349, section=main, note=Similar results when using commuting flows as an alternative proxy for social networks.

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

- Category: `G) External validity and generalization`
- Confidence: `low (0.26154170789138104)`
- Retrieval method: `local`
- Evidence type: `Textual evidence from data/methods and appendix assumptions`
- Data source: Meta Social Connectedness Index (SCI) and U.S. Facebook usage rates cited (Pew, 2021)
- Assumption flag: `False`
- Answer:

Least likely to generalize are: (1) geographies excluded from results (Alaska, Hawaii, Washington, D.C., and territories); (2) very small-population areas removed from the SCI; (3) groups with low Meta/Facebook use—especially older adults (65+); and (4) counties with highly fragmented or complex district splits that violate the limited-fragmentation/simple-boundaries assumptions.

- Quote snippet:

  areas with particularly small populations are removed

- Citation anchors:
  - page=13, words=275-555, note=Results exclude Alaska, Hawaii, Washington, D.C., and territories
  - page=8, words=275-624, note=SCI removes very small-population areas; includes only users active in last 30 days

  - page=8, words=275-624, note=Facebook usage lowest among 65+ (50%)
  - page=44, words=275-491, section=appendix, note=Local mapping assumes limited fragmentation/simple boundaries; some counties are split across many districts (rare)

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

- Category: `G) External validity and generalization`
- Confidence: `low (0.2842624617857855)`
- Retrieval method: `local`
- Evidence type: `text`
- Assumption flag: `True`
- Assumption notes: Appendix imposes: (1) Limited fragmentation with a dominant district per county; (2) Simple boundaries (each county intersects 2 districts); (3) Non-trivial out-of-state connections (network leakage, spectral radius < 1); (4) Network non-degeneracy (linearly independent rows), ensuring local identification.
- Answer:

Yes. They explicitly note scope limits (e.g., findings specific to the U.S.; analysis over the continental 48 states) and impose boundary/identification conditions in the appendix (limited fragmentation, simple boundaries, out-of-state leakage, and network non-degeneracy). They also restrict the border-pairs design to counties fully within one district and within-state pairs due to data/precision limits.

- Quote snippet:

  Though my findings are specific to the U.S., this approach may hold relevance in any context where political boundaries are drawn.

- Citation anchors:
  - page=page 5, words=275-475, note=Scope statement that findings are specific to the U.S.
  - page=page 13, words=0-349, note=Scope over continental 48 states and determinants of district homophily
  - page=page 28, words=0-349, note=Border-pairs design restrictions and precision limits
  - page=page 44, words=275-491, section=appendix, note=Assumptions: Limited fragmentation and Simple boundaries
  - page=page 45, words=0-349, section=appendix, note=Assumptions: Non-trivial out-of-state connections; Network non-degeneracy
  - page=page 46, words=0-349, section=appendix, note=Identification result relying on assumptions (invertibility argument)

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

- Category: `G) External validity and generalization`
- Confidence: `low (0.31940451009186777)`
- Retrieval method: `local`
- Evidence type: `Event-study estimates over time; fixed-effects controls (DMA-by-year, district-by-year); border-pairs robustness`
- Assumption flag: `False`
- Answer:

Over time, effects appear quickly and persist: voter-knowledge impacts take hold in the first post-redistricting survey (2014) and remain relatively stable. Using odd-year data yields similar patterns but with noisier estimates due to much smaller samples. Across media markets, controlling for DMA-by-year fixed effects generally leaves results qualitatively similar—suggesting they are not driven by market-level media—but in the stricter border-pairs design the 'Selected Party' effect becomes insignificant after adding DMA-by-year fixed effects. Moreover, correlations between county traits and changes in homophily largely vanish once media market and district fixed effects are included.

- Quote snippet:

  most strongly takes effect in the first survey after redistricting (2014). Impacts are relatively stable over time.

- Citation anchors:
  - page=20, words=0-349, section=Event studies timing, note=Impacts strongest in 2014 and relatively stable over time
  - page=19, words=275-406, section=Sample timing, note=Odd years have smaller samples; similar results but noisier estimates
  - page=15, words=275-450, section=Media market controls, note=Includes DMA-by-year fixed effects to address media market confounding
  - page=28, words=0-349, section=Border-pairs robustness, note='Selected Party' becomes insignificant after adding DMA-by-year fixed effects
  - page=14, words=0-349, section=Predictors and fixed effects, note=County-level correlations disappear once media market and district FEs are included

H01: Are key variables measured directly or via proxies?

- Category: `H) Measurement validity`
- Confidence: `low (0.3328462292312862)`
- Retrieval method: `local`
- Evidence type: `text`
- Data source: Facebook Social Connectedness Index; CES surveys; Dave

Leip's Election Atlas; Harvard Election Data Archive; MIT Election Data and Science Lab; Daily Kos
- Assumption flag: `False`
- Answer:

Both. Social networks are proxied using Facebook's Social Connectedness Index and a constructed partisan-exposure measure, while knowledge and voting outcomes are measured directly from CES surveys and administrative vote counts/validated rolls.

- Quote snippet:

  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, note=SCI used as proxy for real-world social networks
  - page=15, words=275-450, note=Constructed proxy for partisan exposure from friendships and vote shares
  - page=48, words=0-333, section=appendix, note=Voting outcomes measured from pre/post CES surveys and validated rolls
  - page=47, words=0-207, section=appendix, note=Vote counts built from Harvard EDA, MIT/MedSL, Dave Leip, Daily Kos

H02: What measurement error risks are acknowledged or likely?

- Category: `H) Measurement validity`
- Confidence: `low (0.27088332269027793)`
- Retrieval method: `local`
- Evidence type: `quoted_and_paraphrased`
- Data source: Social Connectedness Index (SCI); CES; Dave Leip's Election Atlas; Harvard Election Data Archive; MIT Election Data and Science Lab; Daily Kos
- Assumption flag: `True`
- Assumption notes: Coverage heterogeneity in SCI by age and potential inconsistencies from merging multiple vote data sources are inferred risks based on the described data construction and usage-rate differences.
- Answer:

Acknowledged and likely measurement error risks include: (1) SCI measurement noise and suppression: small-population areas are removed and noise is added; (2) SCI coverage/assignment issues: only active Meta users in last 30 days are included, usage varies by age, and locations are assigned via self-reported and device data; (3) CES response error: respondents may guess parties and lucky guesses cannot be ruled out; (4) Survey assignment errors: some respondents were assigned to incorrect districts (e.g., 2020 NC), though excluded; (5) Voting data quality/construct issues: turnout metrics exclude cases where turnout ex-

ceeds VAP; (6) Potential harmonization inconsistencies from combining multiple vote data sources.

- Quote snippet:

  Lastly, the third dummy variable, "Selected Correct Party", is coded as 1 if the respondent selected the correct party for the incumbent and 0 otherwise. While lucky guesses cannot be ruled out

- Citation anchors:
  - page=8, words=275-315, note=SCI adds privacy noise and removes small-population areas
  - page=8, words=315-380, note=SCI only includes users active in last 30 days
  - page=8, words=380-430, note=SCI location assignment based on self-reported info and device data
  - page=8, words=520-624, note=Facebook usage rates vary by age (coverage heterogeneity)
  - page=17, words=60-140, note=Selected Party may involve guessing
  - page=17, words=140-220, note=Lucky guesses cannot be ruled out for Selected Correct Party
  - page=17, words=420-520, note=2020 NC respondents assigned to incorrect districts; excluded
  - page=49, words=180-230, section=appendix, note=Elections where turnout exceeds VAP are excluded
  - page=47, words=0-120, section=appendix, note=Voting outcomes constructed by combining multiple data sources

H03: Are there validation checks for key measures?

- Category: `H) Measurement validity`
- Confidence: `low (0.26064539994033026)`
- Retrieval method: `local`
- Evidence type: `Placebo tests and external stability checks`
- Data source: CES; Social Connectedness Index (Meta apps); Dave Leip's Election Atlas
- Table/Figure: Table C1
- Assumption flag: `False`
- Answer:

Yes. The study validates key measures via placebo tests for voter information (district homophily does not predict placebo outcomes), external stability/representativeness checks for the SCI (e.g., >0.99 year-to-year correlation and broadly even usage rates), and robustness noting similar results with county fixed effects.

- Quote snippet:

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

- Citation anchors:
  - page=55, words=0-152, section=appendix, note=Placebo outcomes show no significant prediction by district homophily.
  - page=8, words=275-624, note=SCI validation: >0.99 cross-year correlation; usage rates relatively even across demographics.
  - page=19, words=275-406, note=Robustness: results similar with county fixed effects.

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

- Category: `H) Measurement validity`
- Confidence: `low (0.22703125944891173)`
- Retrieval method: `local`
- Evidence type: `placebo tests and robustness checks`
- Data source: Cooperative Election Study (CES)
- Table/Figure: Table C1
- Assumption flag: `False`
- Answer:

Yes. They assess construct validity via placebo outcome tests—showing district homophily does not affect knowledge of statewide officials (governor, senators) where no effect is expected—and corroborate this with robustness using alternative network measures.

- Quote snippet:

  I find no significant impact of district homophily on these nine outcomes.

- Citation anchors:
  - page=26, words=0-349, note=Placebo outcomes for governor and senators; no significant impact reported.
  - page=27, words=0-349, note=No significant impact on placebo outcomes when using commuting-flow-based homophily.
  - page=55, words=0-152, section=appendix, note=Appendix C.2 summary: district homophily does not significantly predict placebo outcomes; Table C1.

I01: What policy counterfactuals are considered or implied?

- Category: `I) Policy counterfactuals and welfare`
- Confidence: `low (0.34996576379917405)`
- Retrieval method: `local`
- Evidence type: `textual`
- Data source: McCartan et al. (2021) 50-State Redistricting Simulations
- Assumption flag: `False`

- Answer:

The implied policy counterfactuals are alternative congressional district maps: (1) simulating counterfactual district maps via the information-diffusion model to forecast voter information under different boundaries, and (2) comparing outcomes across many legally compliant simulated redistricting plans (5,000 per state).

- Quote snippet:

  allow for the simulation of counterfactual district maps. … simulate 5,000 congressional district maps for each of the 50 states… constrained to follow the given state's redistricting laws.

- Citation anchors:
  - page=34, words=150-215, section=Conceptual Framework: Information Diffusion within Districts, note=Model enables simulation of counterfactual district maps.
  - page=34, words=250-349, section=Comparisons Across Many Simulated Maps, note=5,000 simulated congressional district maps per state, constrained by state laws.

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

- Category: `I) Policy counterfactuals and welfare`
- Confidence: `low (0.30339278782441154)`
- Retrieval method: `local`
- Evidence type: `Textual evidence from results and discussion`
- Data source: CES survey; campaign contributions data; Facebook friendship graph/SCI; commuting flows
- Table/Figure: Figure 6 and Figure 8
- Answer:

The paper highlights distributional impacts from how social networks align with district lines: (1) Higher district homophily raises voters' information about their House representative and reduces roll-off, concentrating these gains in places where networks and districts align; (2) It reallocates campaign money toward in-district candidates and away from out-of-district ones without changing total donations; (3) Using social-network data to draw districts could improve fairness but also risks partisan exploitation.

- 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=20, words=0-349, section=Section 4.1.1–4.1.2, note=Increases in district homophily raise voter knowledge; timing and magnitudes

reported.
  - page=24, words=0-229, section=Section 4.2, note=Roll-off reduced; donations shift toward in-district candidates with no change in totals.
  - page=36, words=0-309, section=Conclusion/Implications, note=Policy tradeoff: social network data could aid fair maps but also be exploited by gerrymanderers.

I03: Are cost-benefit or incidence analyses provided?

- Category: `I) Policy counterfactuals and welfare`
- Confidence: `low (0.24479389793772252)`
- Retrieval method: `local`
- Evidence type: `textual review of provided excerpts (absence of mention)`
- Table/Figure: Table B4; Table C1
- Assumption flag: `True`
- Assumption notes: Conclusion is limited to the provided excerpts; other parts of the full document may exist but are not included here.
- Answer:

No. The provided excerpts contain event studies, placebo tests, and variable descriptions but do not present cost-benefit or incidence analyses.

- Quote snippet:

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

- Citation anchors:
  - page=55, words=0-152, section=appendix, note=Describes placebo outcomes; no cost-benefit or incidence analysis indicated.
  - page=49, words=0-276, section=appendix, note=Defines voting outcome variables and summary stats; not cost-benefit or incidence.
  - page=15, words=0-349, section=main, note=Event study specification and design; no cost-benefit or incidence framing.

I04: What policy recommendations are stated or implied?

- Category: `I) Policy counterfactuals and welfare`
- Confidence: `low (0.3071000189687213)`
- Retrieval method: `local`
- Evidence type: `textual inference from main text`
- Data source: Social Connectedness Index (SCI)
- Assumption flag: `False`
- Answer:

Use social-network data when drawing districts to make them fairer and improve voter information, while instituting safeguards because the same data

could be exploited for partisan gerrymandering. The findings are especially relevant for redistricting, and the framework implies policymakers can evaluate or simulate alternative district maps with network-based measures (e.g., district homophily).

- 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=Main text, note=Policymakers can use SCI to draw fairer districts; caution about partisan exploitation.
  - page=4, words=275-492, section=Main text, note=Findings especially relevant for redistricting.
  - page=34, words=0-349, section=Main text, note=Framework enables simulating counterfactual district maps based on information diffusion.

J01: What missingness or attrition patterns are reported?

- Category: `J) Data quality and integrity`
- Confidence: `low (0.31242748280810845)`
- Retrieval method: `local`
- Evidence type: `document text`
- Data source: CES (Cooperative Election Study)
- Assumption flag: `False`
- Answer:

Reported patterns include: (1) attrition between the CES pre- and post-election surveys; (2) incomplete county coverage each year (even years have at least one respondent from 80–90% of counties; odd years about two-thirds); (3) much smaller samples in odd years (about one-fifth the size of even years); and (4) exclusions for missing/incorrect data: <2% missing party-recognition responses, respondents missing demographic answers (when such controls are used), a small number in 2006–2007 misassigned to counties outside their state, and 925 North Carolina respondents in 2020 misassigned to the wrong district.

- Quote snippet:

  …utilize the same sample as the information outcomes (because there is some attrition between surveys).

- Citation anchors:
  - page=17, words=275-542, section=main, note=Attrition between pre- and post-election surveys; exclusions for missing party-recognition (<2%), missing demographics, 2006–2007 county/state misassignment, and 2020 NC wrong-district misassignment.

  - page=17, words=0-349, section=main, note=County coverage by year: even years 80–90% of counties have a respondent; odd years about two-thirds.
  - page=19, words=275-406, section=main, note=Odd-year samples are about one-fifth the size of even years.

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

- Category: `J) Data quality and integrity`
- Confidence: `low (0.3092116149025004)`
- Retrieval method: `local`
- Evidence type: `document_text`
- Data source: CES; Dave Leip's Election Atlas; Census (VAP)
- Table/Figure: Table B4
- Assumption flag: `False`
- Answer:

No winsorization or trimming is described. Outliers are not specially transformed; instead, observations are excluded in specific cases: (1) CES respondents with missing candidate-party recognition, missing demographic answers, or erroneous county/district assignments (including 925 NC respondents in 2020); (2) elections with no top-of-ticket race; and (3) cases where turnout exceeds the voting-age population. Some variables are set missing when no incumbent runs.

- 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, note=CES exclusions: missing responses, missing demographics, misassigned counties/districts; 2020 NC misassignment excluded.
  - page=49, words=0-276, section=appendix, note=Voting outcomes data exclusions: no top-of-ticket race; turnout exceeds VAP.
  - page=48, words=0-333, section=appendix, note=Variables marked missing when no incumbent runs or when validation variables are missing.

J03: Are there data audits or validation steps described?

- Category: `J) Data quality and integrity`
- Confidence: `low (0.24185108042339895)`
- Retrieval method: `local`
- Evidence type: `Data cleaning rules, validated measures, and placebo tests documented in appendix`

- Data source: CES linked to state voter rolls; Dave Leip's Election Atlas
- Table/Figure: Table B2
- Assumption flag: `False`
- Answer:

Yes. The study describes several validation and data-cleaning steps: use of validated voting records by linking CES respondents to state voter rolls, exclusion of misassigned or missing-response CES records, dropping elections where turnout exceeds the VAP (and those without a top-of-ticket race), and placebo outcome tests to validate findings.

- Quote snippet:

  > Voted in General Election (Validated) Respondent can be linked to state voter rolls, and there is a record of the (Validated) respondent voting in the general election.

- Citation anchors:
  - page=48, words=0-333, section=appendix, note=Validated voting variables via linkage to state voter rolls (CES)
  - page=17, words=275-542, note=Exclusions for missing responses, demographic nonresponse, county/state mismatches, and misassigned districts
  - page=49, words=0-276, section=appendix, note=Excluding elections where turnout exceeds VAP and without top-of-ticket race
  - page=55, words=0-152, section=appendix, note=Placebo outcome tests show no significant effects for non-House offices

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

- Category: `J) Data quality and integrity`
- Confidence: `low (0.30692777890252254)`
- Retrieval method: `local`
- Evidence type: `Direct textual evidence from methods and results descriptions`
- Data source: CES (Cooperative Election Study); Dave Leip's Atlas of U.S. Presidential Elections
- Table/Figure: Table C1 (Appendix)
- Assumption flag: `False`
- Answer:

No. The study transparently reports sample restrictions and weighting, justifies focusing on even years, includes placebo tests and null findings (e.g., no turnout effects), and notes robustness checks—showing no clear signs of reporting bias or selective inclusion.

- Quote snippet:

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

- Citation anchors:
  - page=16, words=0-349, note=Uses CES cumulative weights; describes large, nationally representative samples and linkage to counties/districts.
  - page=17, words=0-349, note=States inclusion of 48 contiguous states; coverage variation across years; notes representativeness differences and transparency on exclusions.
  - page=19, words=275-406, note=Explains focusing on even years due to smaller odd-year samples; similar results with more noise—justifies, not cherry-picks.
  - page=55, words=0-152, section=appendix, note=Reports placebo outcomes with generally no significant effects—transparent reporting of null results.
  - page=67, words=0-258, section=appendix, note=Vote count analysis: reports no turnout effect; shows roll-off reduction and discusses precision/pre-trend—balanced reporting.
  - page=20, words=0-349, note=Notes robustness (commuting flows proxy) and that placebo outcomes show no effect—supports lack of selective reporting.

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

- Category: `K) Model fit and diagnostics`
- Confidence: `low (0.28805989117751185)`
- Retrieval method: `local`
- Evidence type: `appendix regression table`
- Table/Figure: Table C2
- Assumption flag: `False`
- Answer:

Reported diagnostics include R-squared (R2), the number of observations (Obs), clustered standard errors (clustered at the county level) in parentheses, p-values in square brackets, and significance stars.

- Quote snippet:

  Standard errors clustered at the county level in parentheses. P-values in square brackets.

- Citation anchors:
  - page=65, words=0-288, section=appendix, note=Table C2 lists Obs and R2 for each specification; notes state SEs are clustered at county level and p-values in brackets.

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

- Category: `K) Model fit and diagnostics`
- Confidence: `low (0.2771974868223541)`
- Retrieval method: `local`
- Evidence type: `textual (methods) and table (specification robustness)`
- Table/Figure: Table C2
- Answer:

Not specified in the provided context. The excerpts indicate linear probability interpretations and robustness via added fixed effects and controls, but do not report tests of alternative functional forms (logs or nonlinearities).

- Quote snippet:

  > District homophily is measured on a scale from 0 to 1, and outcome variables are binary. As such, reported estimates give the change in probability of the outcome

- Citation anchors:
  - page=19, words=275-406, note=Outcomes are binary and estimates interpreted as changes in probability, implying a linear specification; no mention of functional form tests.
  - page=65, words=0-288, section=appendix, note=Robustness shown by adding fixed effects and controls (Table C2), not by changing functional form.
  - page=15, words=275-450, note=Discussion focuses on additional fixed effects and controls (DMA-by-year, partisan exposure), not functional form variations.

K03: Are residual checks or specification tests reported?

- Category: `K) Model fit and diagnostics`
- Confidence: `low (0.22376402325218922)`
- Retrieval method: `local`
- Evidence type: `appendix text`
- Data source: Appendix C.2
- Assumption flag: `False`
- Answer:

Yes. Specification tests are reported via placebo outcomes; district homophily does not significantly predict these placebo outcomes. No residual diagnostics are described in the provided text.

- 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=55, words=0-152, section=appendix, note=Appendix C.2 summarizes placebo tests and their null results.
  - page=27, words=275-439, note=Notes additional placebo tests with zip-code networks also showing no significant impact.

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

- Category: `K) Model fit and diagnostics`
- Confidence: `medium (0.3655497023341587)`
- Retrieval method: `local`
- Evidence type: `text_and_table`
- Data source: CES survey; 2016 5-Year ACS County-County Commuting Flows
- Table/Figure: Appendix Table C2
- Assumption flag: `False`
- Answer:

Results are generally robust across alternative specifications and estimators. They remain similar when adding county fixed effects and when including odd-year surveys (though odd years are noisier). Using commuting flows as an alternative network measure yields qualitatively similar positive effects on voter familiarity and incumbent support, with magnitudes about half as large and shifts coming from reduced no preference/not voting rather than reduced opponent support. Placebo tests for governors and senators show no significant effects. An alternative border-pairs design delivers qualitatively similar findings; however, the 'Selected Party' outcome becomes insignificant after adding DMA-by-year fixed effects, while 'Heard of Incumbent' and 'Selected Correct Party' remain significant in most specifications (as shown in Appendix Table C2), with reduced precision due to the restricted sample.

- Quote snippet:

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

- Citation anchors:
  - page=19, words=275-406, note=Similar results with county fixed effects; odd years similar but noisier.
  - page=20, words=0-349, note=Stability over time; similar results with commuting flows; placebo outcomes null.
  - page=26, words=0-349, note=Placebo outcomes show no significant impact; describes commuting flows measure.
  - page=26, words=275-375, note=Commuting homophily effects about half as large; reductions from no preference/not voting.

  - page=28, words=0-349, note=Border-pairs design: qualitatively similar; 'Selected Party' becomes insignificant with DMA×year FEs; precision decreases.
  - page=65, words=0-288, section=appendix, note=Appendix Table C2 shows robustness across FE sets; 'Selected Party' loses significance with richer FEs; other outcomes remain significant.
  - page=55, words=0-152, section=appendix, note=Placebo outcomes summary: generally no significant prediction by district homophily.

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

- Category: `P) Previous questions`
- Confidence: `n/a`
- Retrieval method: `n/a`
- Evidence type: `n/a`
- Answer:

No specific question was posed. Below is a concise synthesis of key definitions, data, and identification from the provided context, with citations.

- Roll-off (House turnout relative to top-of-ticket)
  - Definition: Share of top-of-ticket voters who abstain from the House race: (VotesTop − VotesHouse) / VotesTop; captures abstention after incurring the cost of voting; related to "roll-off" literature (Feddersen & Pesendorfer 1996; Miller 2022; Snyder & Strömberg 2010) (page 18 words 275-496).
  - Variable construction and scope: Main analysis at county level from Dave Leip's; robustness at county-by-CD level from Harvard EDA, Daily Kos, Dave Leip's, and MIT EDS Lab; exclude elections without a top-of-ticket race (page 49 words 0-276).
  - Summary stats: Mean 4.42%, SD 12.22 pp; also report turnout in top-of-ticket (mean 51.34%) and House (mean 49.07%) (page 49 words 0-276).
- Campaign contributions data
  - Source: Kuziemko et al. 2023 (October), constructed from FEC data in Bonica 2014; geocoding used to identify whether contributor resides in the same congressional district as the candidate (page 18 words 275-496).
- Border pairs identification design
  - Setup: Neighboring county pairs that straddle a district boundary; counties are similar except district assignment, yielding different district homophily; restrict to counties fully within one district; collapse outcomes to county level; include one observation per county–pair;

compare within-state pairs; precision decreases due to restricted sample (page 28 words 0-349).
  - Specification:
    * $y_{ct} = \alpha_c + \mu_{pt} + \beta\,\bar{\pi}_{c,t} + X'_{ct}\delta + \varepsilon_{ct}$, where y_{ct} is the outcome, μ_{pt} pair-by-year FE, X_{ct} time-varying county controls; use state-by-year FE (insufficient data for district-by-year FE) (page 28 words 0-349).
  - Results note: Qualitatively similar to redistricting design; estimates on "Selected Party" become insignificant after adding DMA-by-year FE (page 28 words 0-349).

- Conceptual framework
  - Develops a theoretical model of information diffusion within districts; district homophily summarized as a network statistic; considers news about elected officials diffusing through networks to determine the equilibrium share of informed voters (page 28 words 0-349).

- Vote count data construction
  - County-by-CD measures built from precinct-level vote counts (Harvard EDA 2000–2010; MIT EDS Lab 2016–2020) combined with county-by-CD vote counts from Dave Leip's (House) and Daily Kos (President, Senator, Governor) (page 47 words 0-207).

- CES variables and summary statistics (knowledge, preferences, voting)
  - Knowledge outcomes (binary): Heard of Representative; Selected Party; Selected Correct Party; defined via pre-survey items on recognizing and assigning party to the House representative (page 47 words 0-207).
  - Voting/preference outcomes (binary): Prefer Incumbent/Opponent/Neither (pre-survey); Voted for Incumbent/Opponent/Neither (post-survey); Validated turnout in general/primary; self-reported general turnout; handling of missing when no incumbent or unmatched rolls (page 48 words 0-333).
  - Summary (sample means): Heard of Rep 93.2%; Selected Party 68.6%; Selected Correct Party 61.7%; Prefer Incumbent 40.1%; Voted for Incumbent 41.0%; Validated General Turnout 57.5%; Self-Reported General Turnout 87.8% (page 49 words 0-276).

- Additional pointers
  - Figures showing effects of homophily on self-reported voting using commuting- and zip-code-based measures (Figures C11, C14); a figure on changes in district homophily with progressively finer bins (Figure C5) (page 61 words 0-15; page 64 words 0-45; page 54 words 0-18).