Panel Nearest Neighborhoods

A new approach applied to voter files

Taylor J Weidman

February 1, 2024

Abstract. *This paper develops a new time-consistent* k-nearest-neighbors *algorithm to observe the residential network structure of residential neighborhoods in more detail. Applied to a panel of voters, this produces a dataset that measures the changes through time in the demographic composition and political choices of voters' neighbors through every major election in North Carolina between 2010 and 2020. The resulting dataset uncovers a number of new descriptive relationships during a time of great political and economic change that were not observable with past approaches. First, using the neighborhood structure of the data, the paper replicates the well known residential clustering of voters and uncovers a new empirical regularity: voters turnout at higher rates in point-neighborhoods with more copartisans in nearly every election. Second, using the new panel structure of the data, the paper shows how party and turnout clustering is impacted by residential and party choices, and is intensified by changes in turnout in neighborhoods with changing political alignment. This project makes the code publicly available, opening up many further areas of research into individual behavior within residential networks.*

I. Introduction

American voters are more residentially segregated by party than at any time in recent history *[Sussel (2013); Kaplan et al. (2022)]*. The social sciences has widely implicated this type of residential segregation in the nature of social networks, economic mobility, and political representation. Since *Bishop (2009)*, much attention has been given to the relationship between residential clustering and political polarization. However, focusing on election outcomes can obscure the many possible sources of electoral change since the members of the electorate make three primary decisions: where (*address*), how (*party*), and whether (*turnout*) to vote. Voter panels make it possible to separate these changes. But as *Brown and Enos (2021)* show, measuring segregation using tessellations, like precincts, washes out important local variation *[Shertzer et al. (2016)]*. The k-nearest-neighbors (*kNN*) approach substantially improves on past measures of segregation by observing those within a point-neighborhood around each voters' residence *[Brown and Enos (2021)]*. But in exchange, this approach has given up the measurement of neighborhood changes *[Autor et al. (2016); Martin and Webster (2020); McCartney et al. (2021)]*. A technical feature of *kNN* has made it impossible to observe how these point-neighborhoods evolve through time.

This paper solves this measurement tradeoff, building a time-consistent *kNN* algorithm that combines the through-time variation in panels with the highly local residential view in point-neighborhoods. This approach was first developed for panels of voters but can be applied broadly in research on the role of neighborhoods and individual behavior (*eg. segregation, networks*). This paper applies the approach to the North Carolina electorate to produce a new administrative dataset with a voter's neighborhood view though time. The paper then explores this dataset, uncovering a number of previously unobservable relationships between the residential environment and voter behavior, suggesting new research opportunities. First, to highlight the geographic structure of the data, this paper 1) replicates the finding that voters live nearer those of their own party at a finer geographic scale, and 2) uncovers that voters in clusters turnout at higher rates. These findings together describe the *Dual Sort*, defined in more detail later. Second, as *Shertzer et al. (2016)* suggest, combined individual and neighborhood data can be used in models with neighborhoods as either the outcome or the predictor variable. Using neighborhood segregation as the *outcome* variable, voter choices impact this dual sort through 1) the residential sorting of active voters into more aligned neighborhoods, and 2) the party sorting of active voters into the affiliation that prevails in the neighborhood. Neighborhood change can also be used as the *predictor* variable in models to show the relationship between the neighbors and the turnout decision. With this approach and descriptive findings, this project aims to facilitate a program to investigate the role of highly local residential networks and environments on individual behavior, particularly among voters.

II. Constructing the Data

Much past research using changes through time has relied on variation in fixed tiles on the plane (e.g. precincts) as the unit of observation *[Autor et al. (2016); Choi et al. (2021)]*. While appropriate for many settings, when individuals can sort on multiple dimensions like party and neighborhood, this variation cannot directly separate the sources of change. For example, a precinct in which all Republicans switch party and vote exclusively for Democrats, is indistinguishable from a precinct in which all Republicans move out, being replaced with all Democrats. And a precinct in which all Republicans become politically disengaged would look identical to a precinct in which all Republicans move out.[1]

Voter-level variation through time can separate these three sources of change, even placing a voter within a tile on the plane *[Martin and Webster (2020)]*. But as *Brown and Enos (2021)* recently suggested, many relationships are simply not observable with this type of spatial variation. In panels, tessellations measure variation between tiles but wash out the variation in residential environments between nearby voters. For example, two voters within a precinct may have two very different residential environments, but will be assigned to the same tile.

The standard geographic k nearest-neighbors *(kNN)* approach fixes the problem of tessellations by constructing circular point-neighborhoods around each voter, offering a significant improvement on segregation measures *[Brown and Enos (2021)]*. This approach defines the point-neighborhood around the individual i at time t with a radius r_{it} to hold k constant, the number of nearest neighbors. *kNN* not only can measure geographically smaller neighborhoods, the primary improvement is the ability to assign different neighborhoods to different individual within the same tile. For example, variation in voters' point-neighborhoods *within* their precinct makes it newly possible to observe that voters turnout at higher rates in neighborhoods with more copartisans. This relationship is obscured in tessellations, no matter the tile size.

However, because the radius r_{it} can vary through time for multiple reasons, this approach does not observe consistent changes in the neighborhood. To illustrate, new neighbor moving in next door would lead to a smaller point-neighborhood radius, r_{it}, pushing out a neighbor further away, even though this more distant neighbor did not move.[2] In this way, nearest neighbors solves the problem of tessellations by precisely measuring the individual's residential view, but it cannot measure its changes through time, *kNN* has meant settling for point-neighborhoods in repeated cross sections, taking away the ability to separate sources of neighborhood change.

The most straightforward solution to this problem holds r constant, not k, constructing point-neighborhoods of equal size for all voters through time: $r_{it} = r$. This provides

[1] As Shertzer et al. (2016) show, this type of approach is more suitable for settings in which the measure is an unchanging characteristic.

[2] A minor issue is that a voter may not be contained in the point-neighborhood of each of their neighbors.

the promised spatial resolution and within-tile variation through time, but creates point-neighborhoods with higher k in areas with higher residential density. This variation in neighborhood population has the potential to obscure relationships between the closest neighbors in denser areas. This paper's approach fixes these problems by assigning each individual i a point-neighborhood defined by an unchanging radius $r_{it} = r_i$ such that it contains the voter's k nearest neighbors in the first year they appear in the data at their current address. Locating the voter at the center of an unchanging circle with a population of k measures the individual' idiosyncratic residential view consistent through time.

II.a. Computational Complexity

Without optimization this point-neighborhood approach carries infeasible computational costs, since a residential distance must be measured between every voter $i \in N$ in the $|N| \times |N|$ triangular distance matrix. Even worse, finding the closest neighbors requires running a sorting algorithm over this matrix in $|N| \log(|N|)$ time, which is prohibitive with N on the order of millions.

Partitioning the distance matrix makes it feasible to compute these neighborhoods. To illustrate, imagine that the state is divided into four partitions, cut vertically along a horizontal line and horizontally along a vertical line. A voter near the center of a partition will have all their neighbors in the same partition. For this voter, the required distance matrix is much smaller after the partitioning, measuring only the distances between the voters in their partition. But a voter on the boundary between partitions will likely have neighbors in an adjacent partition, introducing mismeasurement. Subpartitioning the partitions into four quadrants of its own both worsens the mismeasurement problem and but further improves computational feasibility by reducing the number of distances to be measured.

This tradeoff between computational feasibility and mismeasurement can be resolved using a superpartition around each partition that is ensured to contain the voter's k nearest neighbors. This is done in three steps. Step One initializes the superpartition as all partitions within a radius equal to one edge of the square partition. Step Two increases the size of this radius until the superpartition contains k voters. Then to ensure that the point-neighborhood around voters on the edge of the partition is contained in the superpartition, Step Three further selects the neighboring partitions around the superpartition created in Step Two. These three steps ensure every voter's nearest neighbors are contained in their superpartition, partitioning the distance matrix into something computationally feasible.

II.b. Voter Files

The time-consistent nearest neighbors approach is used to construct a panel of voters, their turnout decisions, and a vector of voters in a fixed point-neighborhood around the voter's residence. The panel is constructed from a sequence of administrative records of North Carolina voters between 2010 and 2020. North Carolina maintains administrative snapshots of all registered voters beginning in 2005, with the years after 2008 providing a linking ID. A dataset containing individual turnout histories supplement these snapshots for voters in more recent years. These voter files have been maintained by each states' Secretary of State in response to the US Congress adoption of the Help America Vote Act of 2002 over concerns about the consistency and accuracy of voter records raised by the 2000 Gore-Bush election *[Igielnik et al. (2018)]*. Figure 1 presents topline descriptive statistics of the North Carolina electorate, with a summary table in the appendix. Voters registered as Democrats and Republicans compose the two major parties, with Independents categorizing all other non-major party voters.

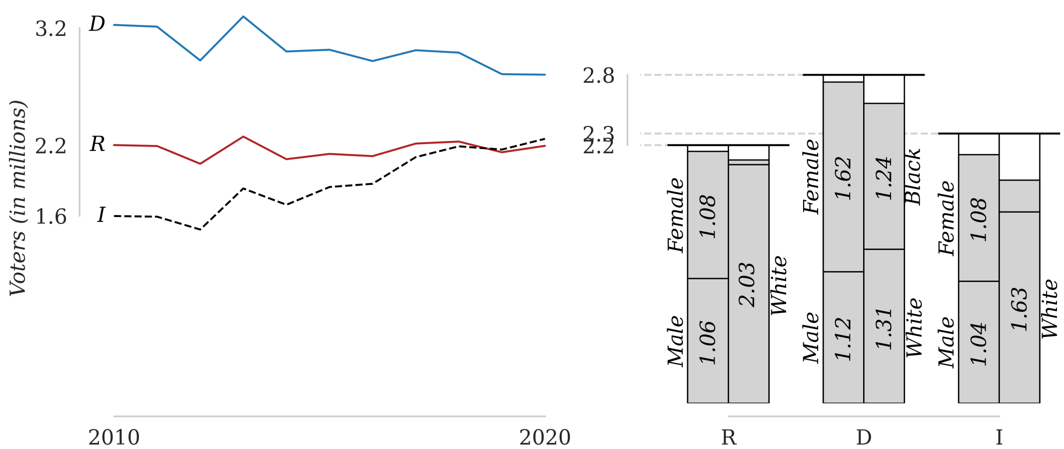

Figure 1. *shows the demographics of the data snapshots. The leftmost panel shows the demographics of each party in 2010. The rightmost panel shows the demographics of each party in 2020. The timeseries shows the evolution of the registration numbers between these years.*

The number of voters registered as Republican has remained nearly constant, the number of voters registered as Democrat has declined, and the number of voters not registered with either major party (labeled Independents) has increased dramatically. Republican and Independent voters are almost all White, while Democrats are nearly equally White and Black, with a slight lean toward women. North Carolina's two largest metropolitan areas, Charlotte and Raleigh, are solidly Democratic, with higher concentrations of Black and

Hispanic voters. In contrast, the state's rural areas are overwhelmingly Republican, with fewer minority voters. Suburban areas tend to be more evenly split between the two major parties.[3]

These administrative records contain each voters' residential address, placing each at a point on the map. The Census Bureau's geocoding service geocodes more than 70% of all registered voters in the panel with no topline systematic relationship between party and success rate. To show the geographic scale of the neighborhood, the left panel of Figure 2 maps a simulated voter and those in their neighborhood for sizes $k \in \{5, 25, 800\}$. The right panel of Figure 2 plots the empirical distribution of neighborhood sizes for $k \in \{5, 25, 800\}$, showing the variability in the radii of the point-neighborhoods. Small k neighborhoods cluster around small radii, and increase in size as k grows.

58

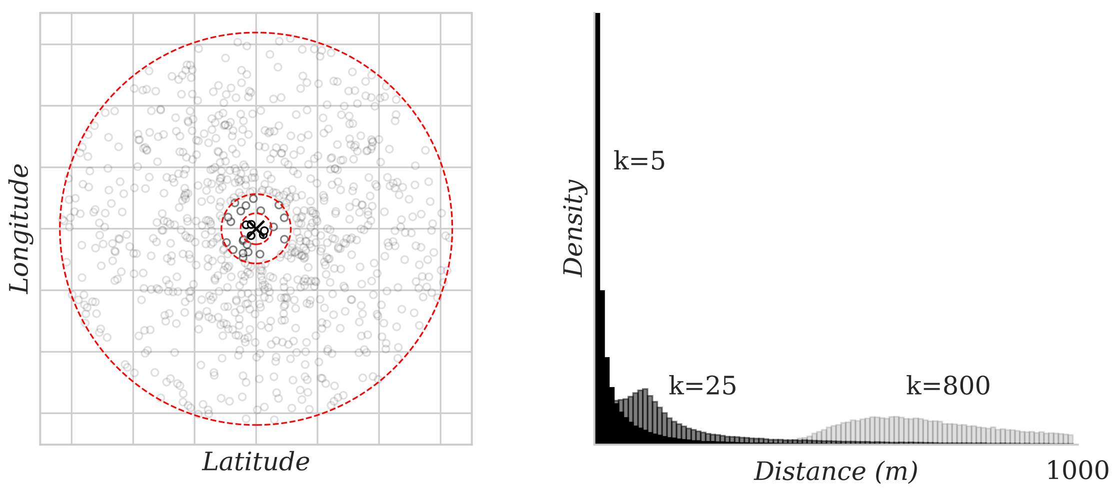

Figure 2. *Simulated point-neighborhood on the left shows the geographic scale of the point neighborhoods. The histograms on the right show the empirical distribution of the size of the radius of these point neighborhoods.*

The data contains the share of the North Carolina electorate's nearest k neighbors and their turnout decisions in every year and election between 2010 and 2020. Time-consistent point-neighborhoods make it possible to both observe a highly local residential view and how it changes through time. The next two sections explore first the highly local view, then how this view changes through time.

[3] See the appendix for turnout maps.

III. Neighborhoods

Point-neighborhoods at this scale can observe relationships simply not visible with tesselations or larger k neighborhoods. This highly local neighborhood view through time can open up new research questions. Before exploring the new panel structure, the point-neighborhood data in cross sections makes it possible to observe two descriptive parallel clustering relationships through time.

II.a Party Clustering

First, the nearest neighbors approach makes it possible to observe the voter's highly local neighborhood view through time, replicating the changing nature of voters' residential and party choices in sharper spatial resolution than has been published to date. *Party Clustering (PC)* defines the phenomenon of voters living nearer others belonging to the same party, relative to the statewide average. To characterize *Party Clustering*, the full measure of voter i's neighborhood composition C_i^k is a vector of the fractions of the neighborhood that belongs to each voter group (D_i^k, I_i^k, R_i^k) divided by the number of voters in the k neighborhood, N_i^k.

$$\hat{C}_i^k = \Big(\frac{|D_i^k|}{|N_i^k|}, \frac{|I_i^k|}{|N_i^k|}, \frac{|R_i^k|}{|N_i^k|}\Big)$$

This measure is then averaged across members of each voter group, g, returning a neighborhood composition $\hat{C}_g^k$, defined for every voter group g and neighborhood size k. In neighborhoods with small k, the in-party component of $\hat{C}_g^k$, $\hat{P}_g^k$, is similar to the isolation score used in *Brown and Enos (2021)* but with equally weighted distances. Further, to provide a meaningful comparison of how neighborhoods have changed within broader statewide changes, the statewide composition is subtracted from the hyper-local neighborhood composition *[Einstein (2011); Sussel (2013)]*.

$$C_g^k = \hat{C}_g^k - \hat{C}_g^{NC} \tag{1}$$

These measures can further be projected into two-party space, $\vec{C}_g^k$, providing a more direct measure of the political clustering of the two major parties.[4] Figure 3 plots these measures on the simplex, with the corners representing fully homogeneous neighborhoods. Neighborhood

[4] This projection normalizes according to the two-party share: $\frac{|N_i^k|}{|D_i^k|+|R_i^k|}$.

composition is represented by a small circle in 2010 and a large circle in 2020. Black circles represent the statewide average. Projections of these compositions into two-party space are shown on the leftmost side, measuring the clustering of major party voters in isolation. A summary table is provided in the appendix.

This replicates the broader clustering of the electorate at the neighborhood level. First, the share of Independents has grown during this time, even in neighborhoods that favor the major parties. Second, at the same time, the major party share of the electorate has shifted toward the Republican party, even in Democrats' neighborhoods. This shift is more easily seen by projecting all six points onto the two-party axis to isolate out the rise of the Independents. The higher projections in 2020 show that, like the statewide shift toward the Republican party, the neighborhood major-party shares have become more Republican for voters of both major parties, although more weakly for Democrats.

Finally, to isolate the changes in neighborhood composition relative to the statewide shifts, the two-party projections can be compared to the two-party statewide averages. Relative to the statewide composition, major party voters' hyper-local neighborhoods ($k = 5$) are roughly 30% more clustered in 2020 than in 2010. Between 2010 and 2020 Republican voters go from having 20.3% to 23.3% more Republican neighborhoods than the statewide share, while Democratic voters go from having 17.3% to 22.6% more Democratic neighbors than the statewide share.

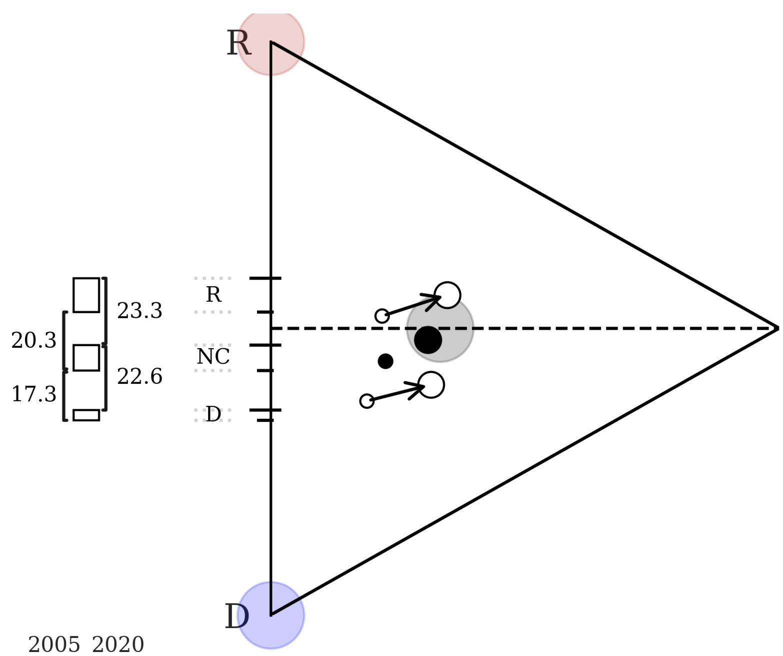

Figure 3. *Statewide registration shares between 2005 and 2020 on the simplex. The horizontal dashed line shows an equal share of both major party voters. The arrows point from the shares in 2005 to the shares in 2020, with the larger points representing the increase in total voter registration during this time. The vertical axis on the left of the simplex shows the two party projection, representing the two party shares of the major parties. The smaller and larger dashes reflect the increase in total registration during this time. The boxes on the left show the change in the two party shares of the major parties relative to the statewide average. This shows how neighborhoods have changed relative to statewide changes.*

These changes are the subject of much research interest: spatial equilibrium models suggest voters move to public goods according to a constrained optimization problem; labor market polarization suggests voters may residentially sort on party; and voters may also directly care about the political party of their neighbors *[Autor (2019); McCartney et al. (2021)]*. While the impact of the complex changes in voter decisions on aggregate changes requires a more detailed investigation, as will be explored later, a time consistent neighborhood can offer a previously unavailable view of its changes.

II.b Turnout Clustering

The approach developed in this paper can be applied to measure a highly local view of other observables. While the clustering of voters is measured primarily using party affiliation, other dimensions of political engagement, like turnout, may also be clustering. Merging voter's turnout histories with their residential environment uncovers that voters indeed turnout at higher rates in political clusters. This *Turnout Clustering (TC)* is measured as a positive relationship between the share of in-party neighbors at time t, P_{it}^k, and the decision to turnout, y_{it}.[5]

$$y_{it} = \beta P_{it}^k + \alpha_{it} + \epsilon_{it} \tag{2}$$

This model measures the relationship between turnout rates and the share of in-party neighbors within the hyper-local neighborhood, with *Turnout Clustering* appearing as $\beta > 0$. To the extent that campaign activities like door knocking are effective, targeting resources at more politically aligned precincts would naturally increase the turnout rates of those in clusters. Precinct-level fixed effects, α_{it}, absorb the factors common across members of precinct like this campaign targeting and instrumental incentives *[Gerber and Green (2017)]*.

[5] P_{it}^k is defined as the component of C_g^k that corresponds to a voter's in-group.

Figure 4 plots β across neighborhood compositions ($k = 5$), election types ($k = 5$), and increasing k distances, highlighting the coefficients for a representative election on November 8th, 2018. Estimates for all elections are listed in the appendix. First, estimating equation (2) separately for each bin over the neighborhood's Republican share shows the consistency of the estimate across neighborhood types. Second, plotting β against the aggregate turnout rate (R_t) for all elections between 2010 and 2020 with at least 1% turnout highlights the consistency of the relationship. Finally, estimating equation (2) for increasingly large k neighborhood sizes highlights the highly geographic structure of the relationship. Error margins reflect 99% confidence intervals.

First, voters turnout at higher rates in politically self-similar neighborhoods across the support of partisan compositions, showing the consistency of *Turnout Clustering*. Binning neighborhoods according to their Republican share and controlling for precinct turnout, the model returns a positive estimate across the support using turnout on 11/06/2018. The relationship is not mechanical. While Republican *turnout* is mechanically lower in neighborhoods with more Democrats simply since there are fewer *Republicans*, Republicans turnout at higher *rates* in neighborhoods with more Republicans, and Democrats turnout at higher *rates* in neighborhoods with more Democrats.

Second, voters turnout at higher rates in neighborhoods with politically self-similar neighbors in every election with at least 1% turnout. These estimates are plotted against the election's turnout rate to allow for comparisons across elections with very different turnout rates. The magnitude of the effect is positively related to turnout rate in small elections, reflecting the fact that the relationship between the neighbors and the voter's turnout rate should be positively related to the election's baseline turnout rate. The declining relationship in large elections reflects the fact that the relationship will be mechanically less meaningful if every voter turns out, and possibly the larger role of a saturated information environment. Highlighting turnout on 11/06/2018, voters in hyper-local neighborhoods with 10% more in-group neighbors have a roughly 7.78% higher turnout rate.

Finally, nearer neighbors are more related to the turnout decision than those further away, a relationship that approaches zero as the size of the neighborhood grows. *Turnout Clustering* is primarily about the closest neighbors. In total, voters turnout at higher rates in neighborhoods with more politically affiliated neighbors, a relationship that both appears in every election with at least a 1 turnout rate between 2010 and 2020 and attenuates toward zero as the size of the residential neighborhood grows in k. This relationship is descriptive and more work is required understand the mechanisms.

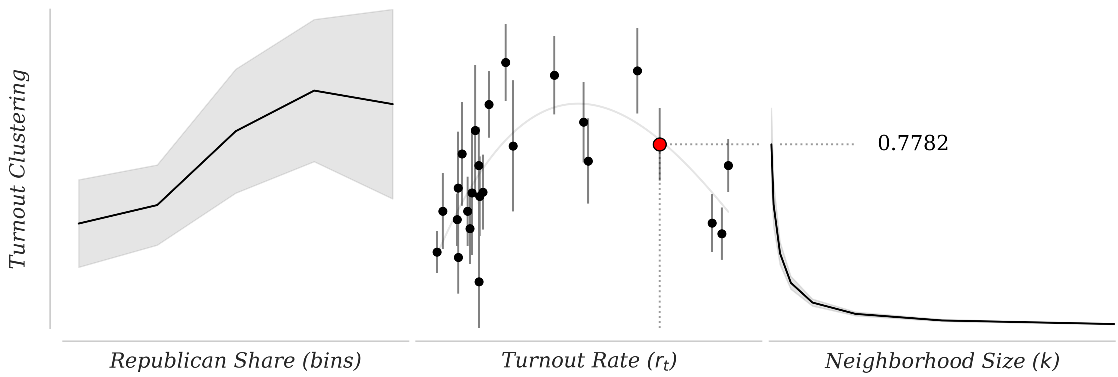

Figure 4. *Highlighting the election on 11/06/2018: the leftmost panel shows the estimate for turnout clustering across Republican share; the center panel shows the estimate by election (t) across turnout rates; and the rightmost panel shows the estimate by neighborhood size (k).*

This analysis highlights the type of investigation possible with point-neighborhoods in cross sections. *Turnout Clustering* would be nearly invisible using neighborhood composed of 1000 voters, even using a radial neighborhood. This relationship would be further obscured using neighborhoods produced with tessellations to tile the plane. This dataset also observes variation through time in the individual-level decisions responsible for the evolution of *Party Clustering* and *Turnout Clustering*. The next section uses the voter-level through-time variation to observe how changes in the voter's three decisions and neighborhood contribute to *Party Clustering* and *Turnout Clustering* in ways that are impossible to observe directly in cross-sections.

IV. Panel Dynamics

The panel structure of the data makes it newly possible to observe the *changes* in a voter's closest neighbors and their three choices (*turnout, residence, and party*). Three examples highlight how the new observable relationships between the individual voter's decisions and their closest neighbors can be used. The first example follows voters as they make a residential move, observing the change in their chosen neighborhoods and the subsequent impact on *Party Clustering* and *Turnout Clustering*. But while the neighborhood must be redefined after a voter moves, the second example uses changes in the point-neighborhood around the voter without changing the neighborhood. Instead, disaggregating neighborhood changes makes it possible to isolate the role of party switching from the other explanations for the evolution of *Party Clustering* and *Turnout Clustering* without redefining the neighborhood.

While the first two examples show how individual decisions impact aggregate segregation, the panel can also be used to study how changes in the residential environment are related to voter behavior. In this frame, the last example uses variation through time to capture changes in the voter's turnout decision and the affiliation of their neighbors. This variation is used in other work to disentangle a number of potential explanations for the descriptive relationships shown here *[Weidman (2023)]*. Similar approaches may prove useful in understanding the role of the neighborhood in other areas of research, like segregation, networks, and spillovers.

III.a. Residential Moves

First, voters' residential decisions are the most widely cited explanation for clustering *[Martin and Webster (2020)]*. When voters move, they may directly care about the political affiliation of their new neighborhood or may prefer local conditions that correlate with political affiliation, like walkability *[Gimpel and Hui (2015)]*. Further, these neighborhood choices might also be weaker among less active voters, leading to sorting on political activity. This panel observes this variation directly. Residential moves would mechanically increase *Party Clustering* if voters choose more politically aligned neighborhoods when they move. Moves would also increase *Turnout Clustering* if active voters move into more aligned neighborhoods than their inactive counterparts. Figure 5 plots the frequency of residential moves by the voter's party affiliation. Summary tables are included in the appendix.

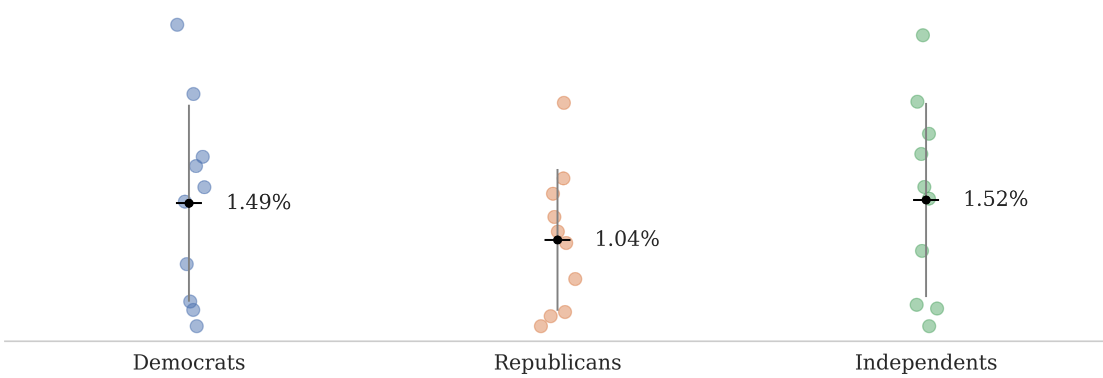

Figure 5. *Frequency of residential moves are shown on the vertical axis, with the empirical standard error and the mean residential move rate, broken out by party.*

Each year on average roughly 1.33% of voters report moving their residential address. This variation is observable using through-time variation in the panel. To measure the impact of residential moves on *Party Clustering* and *Turnout Clustering* separately from party switching simply requires isolating voters who moved but did not switch party. Since these voters are selected conditional on engagement with the electoral system to avoid any 'activity bias' from less active voters being less interested in updating their address when their address changes. This approach isolates these voters to estimate equations (1) and (2) with the voter's in-party neighbors before (*pre*) and after (*post*) their move, producing an elasticity.

$$\varepsilon_{PC} = 2\frac{P_{post} - P_{pre}}{P_{post} + P_{pre}} \tag{3}$$

$$\varepsilon_{TC} = 2\frac{\beta_{post} - \beta_{pre}}{\beta_{post} + \beta_{pre}} \tag{4}$$

These elasticities capture the change in neighborhood segregation among those who moved. Equation (3) uses equation (1) to account for the statewide change in the electorate. Equation (4) uses equation (2) to compare voters in the same precinct, similarly accounting for statewide changes. Figure 6 presents these elasticities for every election with at least 1% turnout, broken out by party. The horizontal axis plots the change in *Party Clustering* due to a residential move, and the vertical axis plots the change in *Turnout Clustering* due to the same residential move, with larger circles representing more moves in the two years prior to the election. A weighted mean on both axes is reported as dashed lines with standard errors on both dimensions. Summary tables are included in the appendix.

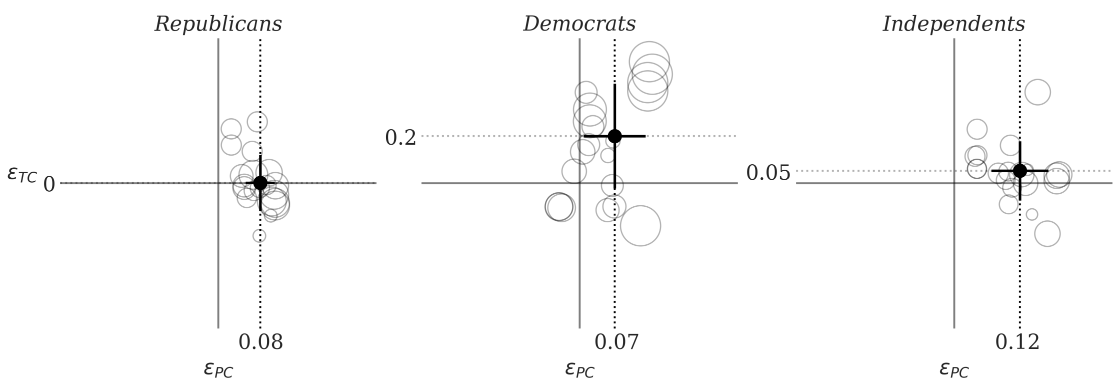

Figure 6. *The directional of the evolution of Party Clustering is shown on the horizontal and the directional of the evolution of Turnout Clustering is shown on the vertical due to residential moves. The mean and standard errors of both dimensions are shown. The size of the points represents the residential move rate.*

This uncovers two new patterns. First, residential moves deepen *Party Clustering*, shown by the average to the right of the vertical axis for all three groups. For example, Republicans move into neighborhoods which are composed of roughly 8% more Republicans than their last neighborhood. Voters of all three groups move into neighborhoods with more in-group neighbors. The larger changes in the economic and political landscape have continued to separate the residential decisions of the electorate along party lines in ways that are consistent with a continued Big Sort. The many underlying mechanisms are the object of much future research: changing incentives of the urban wage gradient; changing preferences for the neighbors with a rise in political polarization; changes in the relationship between political affiliation and preferences for local public goods *[Gimpel and Hui (2015); Autor (2019); McCartney et al. (2021)]*.

Second, residential moves asymmetrically deepen *Turnout Clustering*, shown by the averages on the vertical axis. While the residential moves of Republican and Independent voters show little to no contribution to *Turnout Clustering*, active Democrats appear to disproportionately choose neighborhoods with more Democrats than their inactive counterparts, although within the standard error. This suggests that the forces driving *Party Clustering* are stronger for active Democrats.

III.b. Party Choice

While the first example uses the panel structure to follow a voter through space, the second panel example explores the variation of the neighbors' party decisions through time as another potential explanation for *Party Clustering* and *Turnout Clustering*. Party switches would mechanically increase *Party Clustering* if voters switch to the dominant party in their neighborhood. Party choice would further increase *Turnout Clustering* if disproportionately active voters switch parties towards the dominant party in their neighborhood. The panel observes this variation directly. Figure 7 shows the rates of party switching by the voter's new and former group. Summary tables are included in the appendix.

Among those leaving the major parties, more switch from the Democratic party than from the Republican party. For example, on average $0.016\% + 0.060\% = 0.077\%$ of Democrats have switched each year, 0.016% from affiliating as Republican, and 0.060% from affiliating as Independent. Unsurprisingly, voters switch parties primarily between a major party and a non-major party.

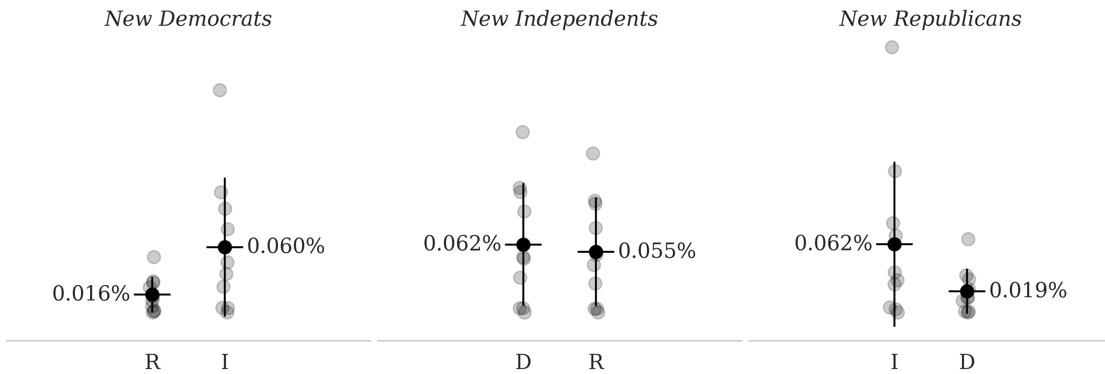

Figure 7. *Share of group members due to switches from the other two groups are shown on the vertical axis, with the empirical standard error and the mean party switch rate, broken out by party.*

This variation in party choice offers a direct time-consistent measure of the evolution of *Party Clustering* and *Turnout Clustering* by isolating voters who themselves did not move or switch parties but experienced a change in their neighborhood composition due to a change in party affiliation. Estimating equations (1) and (2) with both the voter's neighborhood with (*post*) and without (*pre*) the neighbor's party switch produces an elasticity. This isolates the impact on segregation due to party switching independent of other changes.[6] These elasticities are measured according to the following equations for every election with at least 1% turnout.

$$e_{PC} = 2\frac{P_{post} - P_{pre}}{P_{post} + P_{pre}} \tag{5}$$

$$e_{TC} = 2\frac{\beta_{post} - \beta_{pre}}{\beta_{post} + \beta_{pre}} \tag{6}$$

Figure 8, like Figure 6, presents these elasticities for every election with at least 1% turnout, broken out by pary. Points to the left of the horizontal axis represent party switching into the voter's party, and points above the vertical axis represent disproportionate party switching toward the party of active voters. The weighted mean is shown on both axes. Summary tables are included in the appendix.

This uncovers two new descriptive relationships. First, party switching deepens *Party Clustering* for major-party voters. Neighbors tends to switch into the voter's party if the voter belongs to a major party, shown by the horizontal mean to the left of the origin for these two groups. For example, a Republican with a neighbor switching parties will see their

[6] Using all voters who did not switch or move would attenuate the estimates without changing sign.

neighborhood become roughly 10% more Republican due simply to party switches of their existing neighbors. But party switches also weaken *Party Clustering* among Independents, as the neighbors of Independents tend to join the major parties.

Second, the neighbors' party switching tends to weaken *Turnout Clustering* for Democrats, as shown by the vertical mean below the origin. Neighbors tend to join the Democratic party in the neighborhood of less active Democrats. For example, turnout clustering is roughly 15% lower for Democrats with a neighbor switching parties. Party switching appears to have no effect on *Turnout Clustering* for Republicans, and a small effect for Independents.

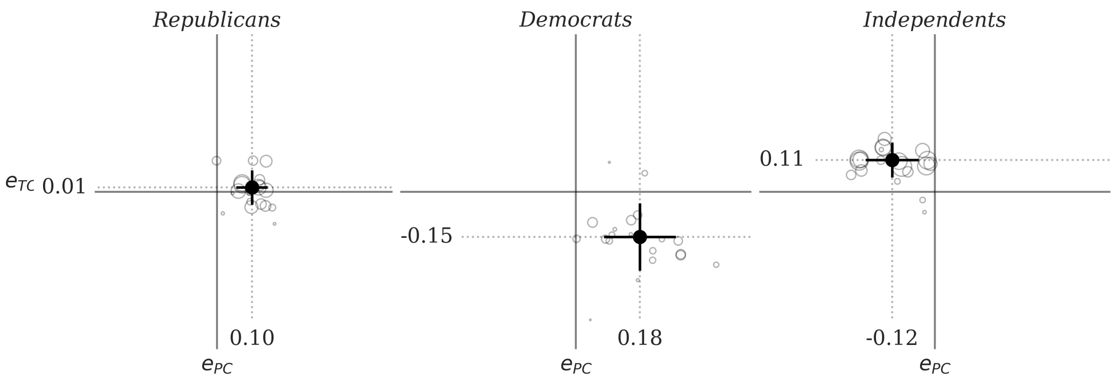

Figure 8. *The directional of the evolution of Party Clustering is shown on the horizontal and the directional of the evolution of Turnout Clustering is shown on the vertical due to party switches. The mean and standard errors of both dimensions are shown. The size of the points represents the party switch rate.*

III.c. Changing Turnout

While the first two examples measured the impact of voter's choices on aggregate measures of segregation, the last example highlights the use of neighborhood changes to study the role of the residential network in voter behavior is newly possible through the observation of the voter and their point-neighborhood together through time. As introduced in *Weidman (2023)*, there are two primary reasons to expect that turnout itself may increase in neighborhoods with new in-party neighbors. First, broad regional realignments may activate existing party members in neighborhoods that see new party members. A precinct with voters who are particularly receptive to a new party message would expect to see both more political engagement among existing members and an increase in the number of voters affiliated to that party. Second, strong cross-domain evidence supports the role of spillovers in behavior, attitudes, and norms *[Gerber et al. (2008); Carrell et al. (2011); Bertrand et al. (2000); Sunstein (1999); Currarini et al. (2009)]*. One conforms to the behavior, norms, and beliefs of those one is exposed to through many interrelated mechanisms. If the turnout decision is influenced by one's neighbors, having more in-party neighbors may lead to higher turnout in political clusters.

Isolating the turnout decision requires holding constant a voters' residential address and party affiliation between elections, leaving only the variation in turnout and neighborhood. Further, a voter is mechanically more likely to vote in an election with high turnout. Measuring the change in turnout across elections with very different turnout rates requires normalizing the turnout decision according to their baseline propensity.[7] This normalization, Δy_{its}, measures the change in turnout between any two elections t and s. The following regression equation estimates the relative increase in a voter's turnout rate for every new in-party neighbor, ΔP_{its}.

$$\Delta y_{its} = \alpha + \beta \Delta P_{its} + \epsilon_{its} \tag{7}$$

Figure 9 plots β for all election pairs within two years, against average turnout, $(R_t + R_s)/2$, further visualizing the turnout rate with the size of the point. Points above (*below*) the horizontal axis represent an increase in turnout among voters who experience an increase (*decrease*) in the number of in-party neighbors within their hyper-local neighborhood. A line of best fit runs across the horizontal axis.

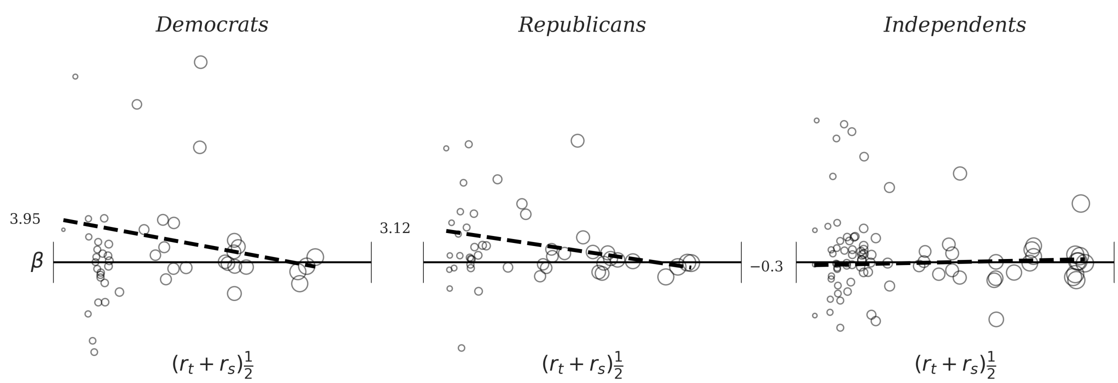

Figure 9. *The relationship between changes in neighborhood composition and changes in turnout, stratified by the average size of the election*

This highlights two relationships. First, major party voters appear to increase their turnout

[7] This normalization is done using the following score:

$$\Delta y_{its} = \frac{y_{it}}{R_t} - \frac{y_{is}}{R_s}$$

.

in neighborhoods that experience new in-party neighbors, as shown by the regression lines above zero. Campaigns in larger elections saturate the electorate with television, radio, and social media advertising, possibly drowning out the role played by neighborhoods in the turnout decision. If this relationship is due to neighborhood spillovers, the effect should appear more clearly in off-year and lower-information elections relative to presidential elections. But if the relationship is due to regional realignments, it should appear across all election types and sizes. So second, the downward slope in regression lines suggests there may be some role played by neighborhood spillovers and not simply regional realignments.

While suggestive, this analysis isn't sufficient to disentangle these two possibilities. *Weidman (2023)* treats this more thoroughly, showing how neighborhood spillovers are responsible an increase in a voter's turnout rate with a new in-party voter in their point-neighborhood. Together with the estimates above, this suggests regional realignments may also play a role. Using different sources of neighborhood change, similar approaches may prove useful in investigating the many potential spatial dependencies in voter behavior.

Conclusion

To date, two important econometric tools have been incompatible: nearest neighbors and panels. This tradeoff has made it impossible to observe how point neighborhoods evolve through time, meaningfully limiting the study of geographic relationships between individuals. Point-neighborhoods available with nearest neighbors approaches offer a point-specific view unique for individuals in such a way that it is not available even with the smallest tesselations. Panels make it possible to observe a unit of observation through time, with significant statistical advantages. This project solves this tradeoff between using point-neighborhoods and through-time variation, making it newly possible to observe how the voter's nearest neighbors and their decisions evolve together, making new types of research questions feasible.

Three examples with North Carolina voters highlight the primary improvement available when combining the geographic resolution of point-neighborhoods with the through-time variation of panels. The first example follows a voter as they move residence, measuring their impact on the two clustering relationships *Party Clustering* and *Turnout Clustering* that were introduced in cross sections. This shows that residential moves intensify *Party Clustering* for all three voter groups, and might intensify *Turnout Clustering* for Democrats and, more weakly, Independents. The second example uses changes in the neighborhood around voters who themselves did not change. This source of variation shows that the party switching of one's neighbors 1) intensifies (weakens) *Party Clustering* for major party voters (Independents), and 2) weakens (intensifies) *Turnout Clustering* for Democrats (Independents). The third example shows how observing these disaggregated neighborhood changes can be used to study the role of the residential environment in voter-level decisions. Variation in the point-neighborhood around voters with unchanging address and party observes that turnout rates increase in neighborhoods that become more politically aligned. This relationship is used in other work to show that neighborhood spillovers may play a role in the turnout decision.

This project makes the algorithm for feasibly computing time-consistent point-neighborhoods publically available. The core data can be further supplemented with other data sources like Census statistics with the method used by *Kaplan and Yuan (2020)*. This builds on previous approaches, providing clear new advantages with many potential applications to study the role of neighborhoods, networks, spillovers in urban political economy.

References

Autor, D. H. (2019). Work of the past, work of the future. *AEA Papers and Proceedings*, 109.

Autor, D. H., Dorn, D., Hanson, G. H., and Majlesi, K. (2016). Importing political polarization?: The electoral consequences of rising trade exposure. *Annual Review of Economics*, 8.

Bertrand, M., Luttmer, E. F. P., and Mullainathan, S. (2000). Network effects and welfare cultures. *The Quarterly Journal of Economics*, 115(3).

Bishop, B. (2009). *The Big Sort*. Houghton Mifin Harcourt.

Brown, J. R. and Enos, R. D. (2021). The measurement of partisan sorting for 180 million voters. *Nature Human Behavior*, 5.

Carrell, S. E., Hoekstra, M., and West, J. E. (2011). Is poor fitness contagious?: Evidence from randomly assigned friends. *Journal of Public Economics*, 95(7-8).

Choi, J., Kuziemko, I., Washington, E. L., and Wright, G. (2021). Local economic and political effects of trade deals: evidence from nafta. *NBER Working Paper Series*, (29525).

Currarini, S., Jackson, M. O., and Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4).

Einstein (2011).

Gerber, A. S. and Green, D. P. (2017). Field experiments on voter mobilization: An overview of a burgeoning literature. *Handbook of economic field experiments*, 1.

Gerber, A. S., Green, D. P., and Larimer, C. W. (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review*, 102(1).

Gimpel, J. G. and Hui, I. S. (2015). Seeking politically compatible neighbors? the role of neighborhood partisan composition in residential sorting. *Political Geography*, 48:130–142.

Igielnik, R., Keeter, S., Kennedy, C., and Spahn, B. (2018). Commercial voter files and the study of u.s. politics. *Pew Research Center*.

Kaplan, E., Spenkuch, J. L., and Sullivan, R. (2022). Partisan spatial sorting in the united states: A theoretical and empirical overview. *Journal of Public Economics*, 211(104668).

Kaplan, E. and Yuan, H. (2020). Early voting laws, voter turnout, and partisan vote composition: Evidence from ohio. *American Economic Journal: Applied Economics*, 12(1).

Martin, G. J. and Webster, S. W. (2020). Does residential sorting explain geographic polarization? *Political Science Research and Methods*, 8.

McCartney, W. B., Orellana, J., and Zhang, C. (2021). Sort selling: Political polarization and residential choice. *FRB of Philadelphia Working Paper*, 21(14).

Shertzer, A., Walsh, R. P., and Logan, J. R. (2016). Segregation and neighborhood change in northern cities: New historical gis data from 1900–1930. *Historical Methods*, 49(4).

Sunstein, C. R. (1999). The law of group polarization. *University of Chicago Law School, John M. Olin Law and Economics Working Paper*, 91.

Sussel, J. (2013). New support for the big sort hypothesis: An assessment of partisan geographic sorting in california, 1992–2010. *PS: Political Science and Politics*, 46(4).

Weidman, T. J. (2023). Homophily turnout: Network effects and voter turnout.