Accounting for Travel Times in Estimating Political Dislocation

Nicholas Eubank[*], Zhenghong Lieu[†], Jonathan Rodden[‡]

November 20, 2019

Preliminary Draft
PLEASE DO NOT CITE
Please click here for most recent version.

Abstract

[*]Assistant Research Professor, Social Science Research Institute, Duke University nick@nickeubank.com

[†]Oxford University

[‡]Professor, Department of Political Science and Senior Fellow, Hoover Institution, Stanford University jrodden@stanford.edu

Fundamental to a political of single member districts is the idea that there is value in voters who live in the same area being represented by a single politician. Arguments for this are multifaceted — voters in the same area are likely to share political interests; voters in the same area are better able to communicate and coordinate with one another; politicians can better maintain connections with voters in the same area; voters in the same area are especially likely to belong to the same social communities — but all suggest the importance of voters being located in districts with their geographic peers. [need cites]

The idea that there is value in the constituents of a district being physically proximate to one another is present not only in political theory texts, but also in law. Many states, for example, explicitly state that geometric compactness is one of the desired attributes of electoral districts, and indeed compactness is often a metric used to evaluate the reasonableness of districts in legal cases around districting.

Yet historically, when evaluating whether districts accomplish their goal of creating districts composed of constituents who are "close" to one another, proximity is almost always evaluated on the basis of geographic district. But geographic distance often does not correspond to the human experience of proximity, as anyone who has tried to travel even a few miles across downtown at rush hour can attest. This reliance on purely geometric metrics is understandable given its tractability, but with the rise of ubiquitous data on travel patterns and the amount of time it actually takes for citizens to drive from one location to another, it is now possible to measure the distances between citizens not in feet or miles, but in actual travel times, reflecting for the first time the actual human geography of distance.

In this paper, we build on the work of Lieu (2019) – who develops a measure of district compactness built on a travel-time metric rather than a geographic-distance metric – to create a revised version of the *Political Dislocation* from Eubank and Rodden (2019) that takes into account travel times to more accurately estimate the characteristics of voters' local neighborhoods.

Political Dislocation measures the degree to which a voter's district is aligned with their immediate geographic neighbors. In particular, we examine the degree to which the *partisan composition* of a voter's actual electoral district differs from the partisan composition of their local neighborhood. Where these measures differ dramatically — where, for example, a voter whose k nearest neighbors (where k is the number of people in the voter's actual legislative district) are mostly Democrats, but despite this their district is mostly Republican — we term that voter *politically dislocated*. As shown in Eubank and Rodden (2019), not only is this measure of direct normative importance, it is also a very good measure of the degree to which an individual voter is the victim of packing or cracking, making it a valuable individual-level metric of abusive districting and gerrymandering.

In this paper, we take the *Political Dislocation* measure from Eubank and Rodden

(2019) and update it by identifying each voter's k nearest neighbors not on the basis of geographic proximity, but on the basis of shortest travel times. As we will show, this not only provides an objective basis for identifying and guarding against abusive districting practices (like drawing districts that cross large impassable bodies of water), but it also offers a consistently different picture of the social context of suburban voters. As shown below, we find that our measure generally shows that suburban voters' nearest neighbors tend to be more conservative when one uses travel times as a distance metric, likely because more geographically distant exurban (generally more conservative) voters are often closer on human-scales than voters on the other side of the city (who tend to be more liberal).

1 Data & Methodology

Following Eubank and Rodden (2018), estimation of the partisan composition of each voter's neighborhood is accomplished through a three-step process. First, precinct-level election returns from the 2008 Presidential election are used to estimate the spatial distribution of voters in each state.[1] This is done by creating a number of representative voter points within each precinct, where points are positioned uniformly at random within each precinct's catchment area, and the number of points in each precinct's catchment area is proportional to the number of votes cast for each party.[2] While this down-sampling and placements of points randomly within precincts does introduce some noise, as discussed in Appendix A, the variability contributed to our dislocation measure is empirically very small. This analysis generates an estimate *for each representative-voter point* of the share of neighbors who are co-partisans.

Estimation of the partisan composition of the neighborhood around each of these representative-voter points is then calculated. In our naive nearest neighbor analysis (following Eubank and Rodden (2019)), for each representative-voter point v of a given party $p \in \{D, R\}$, the partisanship of the neighborhood around v is equal to the share of the k nearest points (as measured by geographic distance) who are democrats. The number of nearest neighbors considered – k – is set to ensure the included points

[1] Before calculating these intervals, we apply a uniform swing to account for McCain / Obama vote shares in our 2008 Presidential two-party vote share data. In particular, as McCain's two-party vote share was 46.31%, we apply a 3.69 percentage point uniform swing to all data, so that a Republican voter whose voter neighborhood is 46.31% co-partisan would be said to be in a perfect 50% co-partisan neighborhood. In Congressional races, Democratic victories have been quite rare in districts where McCain's 2008 vote share was higher than 46.31 percent, and Republican victories have been quite rare in districts where Obama's vote share was higher than 53.69 percent.

[2] In particular, the number of points we generate in each precinct for each party is determined by taking a binomial draw from the total number of actual voters. The binomial probability varies by state-chamber, but is equal to $prob_k = \frac{number of districts}{number of voters in state} * k$, where k=1,000 for state legislative districts and 5,000 for US Congressional districts. This probability generates k voters per district in expectation. A larger number of points are used for US Congressional districts to adjust for the fact that the relatively small size of precincts with respect to US Congressional districts reduces the sampling probabilities in each precinct, increasing sampling variance for a given k.

represent the number of voters in the average district in *state* for chamber *chamber*. [3] This estimate is analogous to asking "if a circular electoral district of average district population were centered on this voter, what share of people in that district would be co-partisans?"

In this paper, we extend this method by adopting a new metric for identifying each representative voter's nearest neighbors: rather than identifying nearest neighbors based on geometric distance, we instead measuring distance by calculate average driving times between pairs of representative voters. Each voters neighborhood composition is then calculated as the partisan composition of the k other representative voters who are closest in terms of driving times (details of implementation of this metric can be found in Appendix B).

Finally, in order to ensure comparable estimates of the composition of a voter's geographic neighborhood with the partisanship of their legislative districts, estimation of the partisan composition of each voter's district is accomplished by overlaying post-2010 legislative district boundaries over the precinct returns used to estimate geographic neighborhood partisanship and aggregating these returns to estimate legislative district partisanship.[4]

2 Results

2.1 Identifying Pathological Cases

One feature of this measure is that it provides a mechanism for dealing with districts crossing bodies of water. Figure 1 below plots *Political Dislocation* scores computed both using geometric distance (left) and travel times (right), as well as the difference between the two (below). As the figure shows, dislocation scores are generally similar, the vary dramatically along the southern costs of the Chesapeake Bay where a naive implementation of *Political Dislocation* treats people across the bay as "neighbors."

The ability to properly capture the impact of geographic boundaries represents more than a way of correcting a problem with *Political Dislocation*, however – it shows that that the travel-time based dislocation measure can help police gerrymandering accomplished by running districts across impassable bodies of water, *technically* maintaining compactness while generating districts with residents who are actually quite functionally isolated from one another.

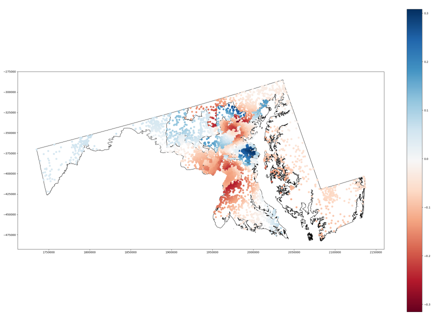
This measure is also able to automatically take into account the passibility of water, however. Consider, for example, the case of Lake Pontchartrain in Louisiana in Figure 2.

[3] To illustrate, consider a state-chamber with 3 districts and 300,000 voters. The average district is home to 100,000 voters, and so the number of points considered in the nearest neighbor analysis should represent 100,000 voters. Note that because of how *prob* is constructed, this will always amount to examining the share of the 1,000 points around each person who are co-partisans.

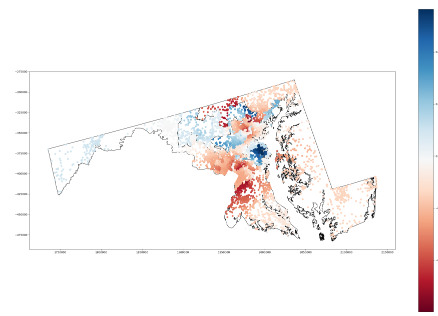
[4] The same uniform swing applied to geographic neighborhoods is also applied here.

Figure 1: Correcting for Biases Caused by Chesapeake Bay

(a) Geometric-Distance Political Dislocation



(b) Driving-Time Political Dislocation



(c) Difference in Political Dislocation Scores

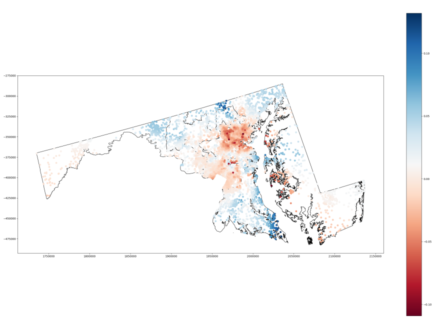
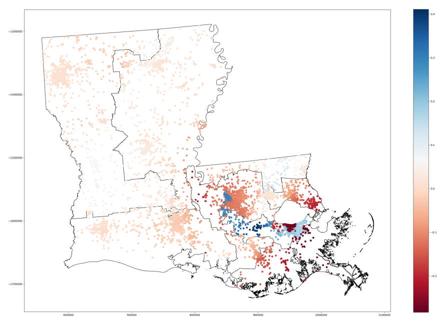


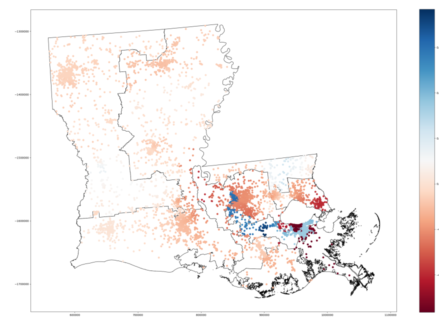
Figure 1a shows the naive, geometric-distance-based Political Dislocation. Figure 1b plots driving-time-based Political Dislocation. Political Dislocation is calculated as the difference in Democratic vote share of each voter's assigned district and the Democratic vote share of her nearest neighbors (calculated either by geographic distance or travel times). Higher values denote individuals whose district is more Democratic than their nearest neighbors. Figure 1c plots the difference between drive-time Dislocation and geometric-distance Dislocation for each representative voter.

Figure 2: Reflecting Road Connections Across Lake Pontchartrain

(a) Geometric-Distance Political Dislocation



(b) Driving-Time Political Dislocation



(c) Difference in Political Dislocation Scores

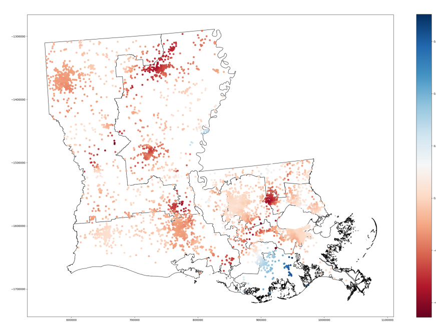


Figure 2a shows the naive, geometric-distance-based Political Dislocation. Figure 2b plots driving-time-based Political Dislocation. Political Dislocation is calculated as the difference in Democratic vote share of each voter's assigned district and the Democratic vote share of her nearest neighbors (calculated either by geographic distance or travel times). Higher values denote individuals whose district is more Democratic than their nearest neighbors. Figure 2c plots the difference between drive-time Dislocation and geometric-distance Dislocation for each representative voter.

As with Maryland, Louisiana has districts that cross this body of water. Unlike in Maryland, however, travel-time based political dislocation does not change radically because the North and South shores of Lake Pontchartrain are connected by a major motorway, reducing their functional isolation.

2.2 General Trends

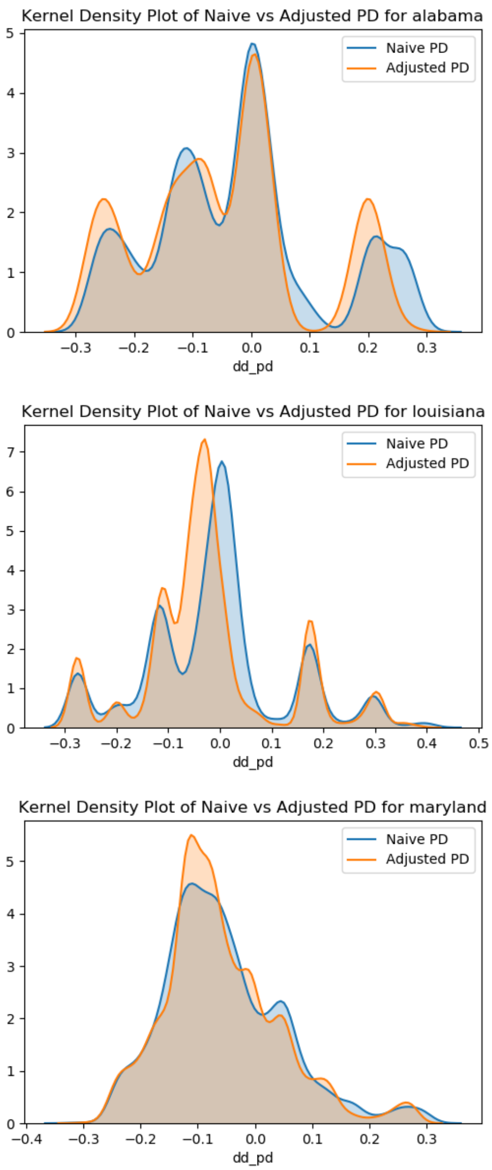
But the ability to objectively accommodate idiosyncratic geographic features is not the only change that comes with using travel-time as a distance metric. Figure 3 below plots the distributions of dislocation scores under naive Dislocation and travel-time-adjusted Dislocation.

As the Figure shows, adjusting for travel time tends to result in a left-shifting of Dislocation distributions. This suggests that voters tend to be in less disproportionately Democratic districts / more disproportionately Republican districts.

What is driving this effect? The difference plots in Figures 1 and 2. Many of the largest changes observed are around smaller cities. What appears to be happening is that when we move from measuring geographic distance to travel times, more of the nearest neighbors for those in these cities smaller cities become people living outside the cities, since travel times are more compressed in areas accessible by low-traffic high-speed roads. As these more rural residents tend to be more conservative, more of the nearest neighbors of people on the periphery of cities become rural residents, resulting in a lower score for the Democratic vote share of their nearest neighbors.

Note that this only happens in cities that are small enough that voters' k (where k is the number of people in a district – here 700,000 people) includes people who don't live in their city. In much larger cities (Baltimore, New Orleans), districts often fall entirely within city limits, precluding this dynamic from coming into play.

Figure 3: Change in Dislocation Distributions from Using Travel Time



Political Dislocation is calculated as the difference in Democratic vote share of each voter's assigned district and the Democratic vote share of her nearest neighbors (calculated either by geographic distance or travel times). Higher values denote individuals whose district is more Democratic than their neighbors.

References

Eubank, Nicholas and Jonathan Rodden. 2018. “Who is my Neighbor? The Spatial Efficiency of Partisanship.” *Working Paper* .

Eubank, Nicholas and Jonathan Rodden. 2019. “Political Dislocation: A Voter-Level Measure of Partisan Representation and Gerrymandering.” *Working Paper* .

Lieu, Zhenghong. 2019. “Using Human Geography to Build a More Meaningful Compactness Measure for Districting Algorithms.” *Working Paper* .

A Sampling Variability

As noted in Section **??**, our estimates of voter dislocation are subject to two forming of sampling variability: downsampling the number of voters, and then placement of these voters within each precinct.

The first source of variance comes from our need to downsample the universe of all US voters for computational tractability. In particular, we create a set of "representative voters" in each precinct for each party by taking a binomial draw from the total number of actual voters for each party in each precinct. The binomial probability varies by state-chamber, but is equal to $prob_k = \frac{numberofdistricts}{numberofvotersinstate} * k$, where k=1,000 for state legislative districts and 5,000 for US Congressional districts. This probability generates k voters per district in expectation. This downsampling makes it computational feasible to calculate the partisan composition each representative voter's k nearest neighbors. A larger k is used for US Congressional districts as they are much larger with respect to individual precincts, resulting in lower binomial draw probabilities for each precinct, thus increasing sampling variance.

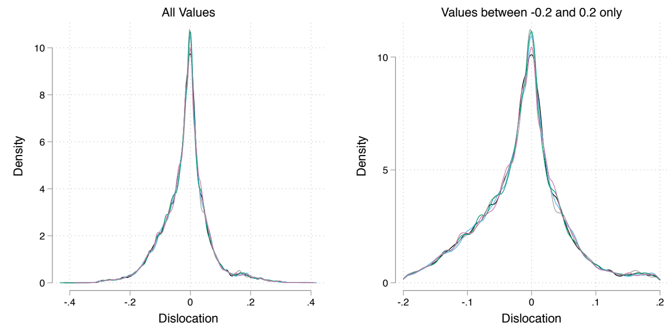
The second source of variance comes from distributing points uniformly within each precinct. Thankfully, US precincts are generally quite geographically compact, limiting the amount of variation introduced by this process.

To evaluate the impact of these sources of variability, Figure 4 below plots the distribution of (representative) voter-level dislocation scores across five rounds of representative-voter point generation. As the Figures show, variation across each round is extremely small, especially within respect to cross-voter simulation: between-round standard deviations constitute only 0.101 %, 0.103 %, and 0.104 % of total variation for these five rounds for state lower, state upper, and US House chambers respectively.

Figure 5 presents analogous diagnostic distribution at the level of legislative districts (plotting the distribution district-level average absolute dislocation scores). Again, between-round standard deviations constitute only 1.11 %, 0.99 %, and 1.67 % of total variation for these five rounds for state lower, state upper, and US representative chambers respectively.

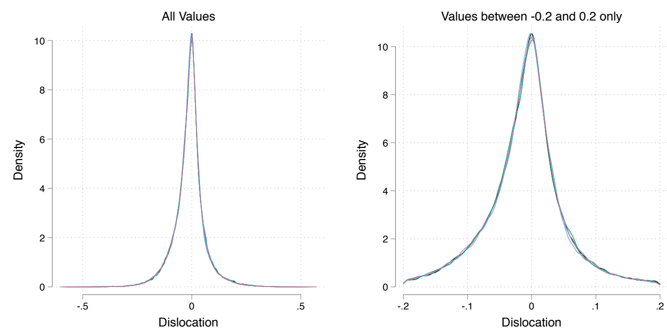
Figure 4

Voter-Level Dislocation Distributions, US Congress
Across 5 Generations of Representative Points



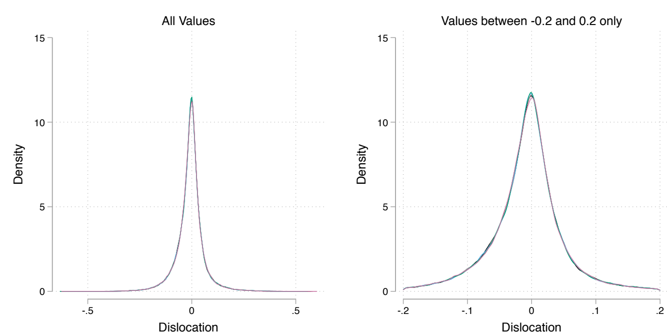
Within Simulation Std. Dev.: 0.0776, Between Simulation Std. Dev.: 0.0001
Between As Pct of Total Std. Dev.: 0.104%
Kernel densities plotted from 10% sample; variance decomposition from full sample.

Voter-Level Dislocation Distributions, State Upper
Across 5 Generations of Representative Points



Within Simulation Std. Dev.: 0.0709, Between Simulation Std. Dev.: 0.0001
Between As Pct of Total Std. Dev.: 0.103%
Kernel densities plotted from 10% sample; variance decomposition from full sample.

Voter-Level Dislocation Distributions, State Lower
Across 5 Generations of Representative Points

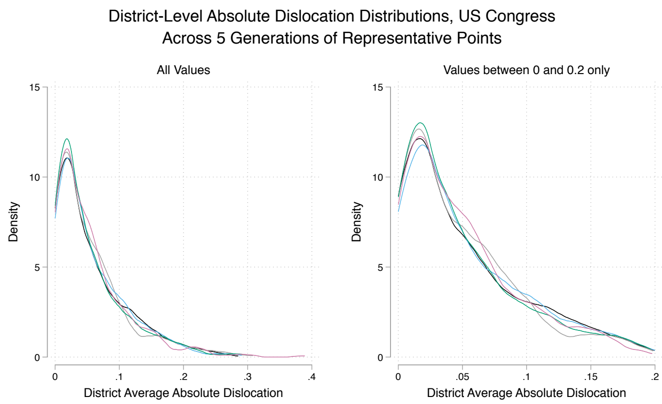


Within Simulation Std. Dev.: 0.0659, Between Simulation Std. Dev.: 0.0001
Between As Pct of Total Std. Dev.: 0.101%
Kernel densities plotted from 10% sample; variance decomposition from full sample.

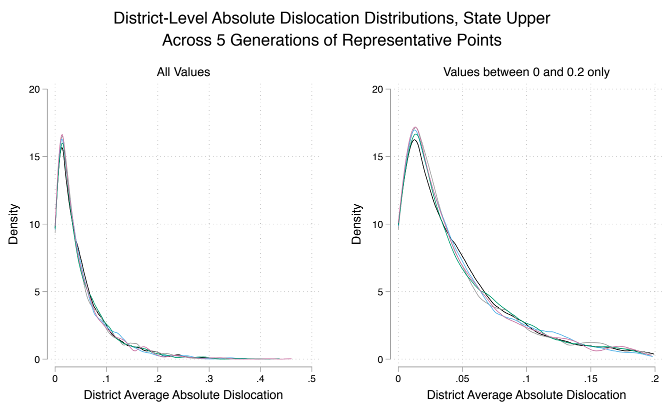
B Driving Times Calculations

[details here]

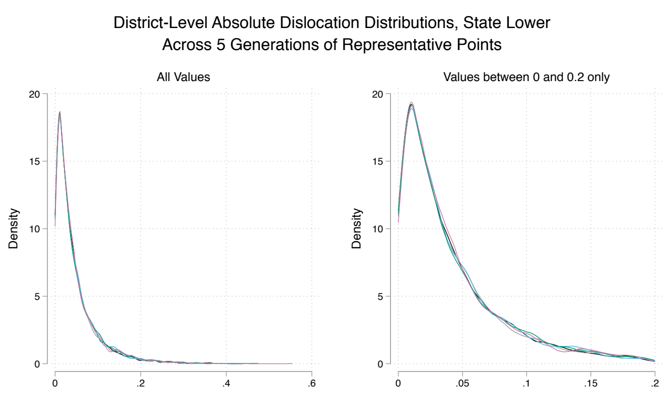
Figure 5



Within Simulation Std. Dev.: 0.0558, Between Simulation Std. Dev.: 0.0009
Between As Pct of Total Std. Dev.: 1.67%



Within Simulation Std. Dev.: 0.0520, Between Simulation Std. Dev.: 0.0005
Between As Pct of Total Std. Dev.: 0.99%



Within Simulation Std. Dev.: 0.0512, Between Simulation Std. Dev.: 0.0006
Between As Pct of Total Std. Dev.: 1.11%