Additional Information

Zhenghong Lieu

Table of contents

1. Resume
2. Abstract of coauthored paper with Professors Jonathan Rodden and Nicholas Eubank
3. Abstract of final year politics thesis
4. Tutors' comments
5. Selected certificates of completion of MOOCs:
    - Algorithms
    - Databases
        - SQL
        - Relational Algebra
        - Relational Design Theory
    - Computer architecture (NAND 2 Tetris I & II)
    - Linear Algebra
    - Multivariate Calculus
    - Machine Learning
    - Big Data Analysis in Scala and Spark
    - Functional Programming in Scala
6. Short snippet of Mathematical Analysis problem set
7. Short snippet of software project I developed

Zhenghong Lieu (lieuzhenghong@gmail.com)

Education

Oct 2017 – present

University of Oxford

BA in Philosophy, Politics and Economics (PPE)

Jun 2017 – present

Massive Open Online Courses (MOOCs) completed

Computer Science

- Algorithms and Data Structures (Stanford)
- Databases (Stanford)
- Machine Learning (Stanford)
- From NAND to Tetris I & II (computer architecture) (IDC Herzilya)
- Competitive Programmer's Core Skills (SPSU)

Data Science

- Functional Programming Principles in Scala (EPFL)
- Parallel Programming in Scala (EPFL)
- Big Data Analysis in Scala and Spark (EPFL)

Mathematics

- Linear Algebra (Imperial College London)
- Multivariate Calculus (Imperial College London)

Experience

Jun 2019 – Sep 2019

Data Scientist (Intern)

Inzura AI

Increased monthly active users by $> 20\%$ by building a Bayesian SMS sender in Python and SQL, meeting a KPI which secured an important client contract

Built deep learning model with Keras that reduces trip processing time from 7s to 0.05s, which made possible a new revenue vertical for the company

Deployed distributed Apache Spark infrastructure and performed Big Data analysis with Scala on ~85 million data points used to assign users more accurate risk ratings

Jun 2018 – Aug 2019

Blockchain Developer (Intern)

Infocomm Media Development Authority of Singapore (IMDA)

Wrote a research paper on blockchain interop protocols and presented it to senior management

Created a fully-automated blockchain demonstration of supply chain management:

- Real-time blockchain visualisation using JavaScript and Canvas
- IoT sensor integration to update asset location data on Hyperledger blockchain
- QR-code based asset tracking with Node.js

Jun 2017 – Aug 2017

Software Engineer

Imcomp International

Greatly increased firm efficiency (75% less time taken to generate reports, 300 engineer hours saved per month) by developing bespoke building inspection software (Electron, Vue.js, JavaScript)

Jan 2017 – May 2017

Venture Capital Analyst (Intern)

iGlobe Partners

Co-wrote 2 papers recommending investment; senior management accepted both recommendations and made Stage B investments

Developed more powerful version of Microsoft Mail Merge in Python that automates away 90% of human errors and saves 50% time when sending mass mailers

May 2016 – Sep 2017

Python Programming Tutor (Volunteer)

Ulu Pandan Stars

Taught a team of underprivileged children Python and led them to win 3rd (amongst ~100 participating teams) in national hackathon CodeXtreme

Feb 2015 – Dec 2016

Soldier (Conscript)

Singapore Armed Forces (SAF), HQ Army Medical Service

Developed the Army's first item tracker and SMS reminder service with React and Node.js:

- Won second prize in the Army Annual Innovation Competition
- Won Best Soldier of the Month out of ~3000 soldiers

Working Papers

2019 "Accounting for Travel Times in Estimating Political Dislocation"

with Nicholas Eubank and Jonathan Rodden

2019 "Using human geography to build a more meaningful compactness measure for automated districting algorithms"

2019 "Can freeform communication increase the rate of Pareto-efficient outcomes in repeated games?"

Honors and Fellowships

Oct 2019 Postmaster

Major college scholarship awarded for sustained academic excellence

Aug 2019 Raff Prize

Awarded for best second-year Economics performance in my college

Oct 2018 Exhbitioner

Minor college scholarship awarded for good performance in the first year

Aug 2018 Sam McNaughton Prize

Awarded for top score in Philosophy in the PPE first-year exams

June 2018 Quantitative Political Essay Prize

Awarded to the best quantitative essay of my Oxford politics cohort (~300 students)

Dec 2017 – present Fowler Prize(s)

Awarded for First-class performance in termly exams, awarded 7/7 terms

Sep 2017 National Infocomm Scholarship

Full-ride undergraduate scholarship awarded to promising Singaporean students by the Infocomm Media and Development Authority (IMDA), a Singaporean statutory board. Awarded on the basis of academic excellence, demonstrated interest in tech, and a will to serve the nation.

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 $prob_k = \frac{numberofdistricts}{numberofvotersinstate} * k$, where k=1,000 for state legislative districts and 5,000 for US Congressional districts. This probability generates k voters per district in expectation. 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.