

# Mark Huberty

Department of Political Science  
University of California, Berkeley  
210 Barrows Hall #1950, Berkeley, CA 94720-1950  
Tel: 206-550-2157  
e-mail: [markhuberty@berkeley.edu](mailto:markhuberty@berkeley.edu)  
Web: <http://markhuberty.berkeley.edu>  
Code: <https://github.com/markhuberty>

## EDUCATION

**2007-**        **Ph.D. Program, Political Science**  
University of California, Berkeley

Research interests: comparative political economy and economic policy; climate and energy policy; computational social science; machine learning.

Ph.D. expected June 2013.

**2005-2007**    **M.A. with Honors, European Studies and International Economics**  
The Johns Hopkins University, School for Advanced International Studies (SAIS)

**1996-2000**    **B.S. with High Distinction and Honors, Chemistry**  
Harvey Mudd College

## EMPLOYMENT AND AFFILIATIONS

**2010+**        **Visiting Fellow, Bruegel**

**2010**        **Research Assistant, Professor Jasjeet Sekhon**  
Project: GPU-accelerated algorithms for genetic matching.

**2008+**        **Research Associate, Berkeley Roundtable on the International Economy**  
Project: The political economy of climate change and energy systems transformation

**2007**        **Research Associate, A.T. Kearney**  
Global Business Policy Council

**2000-2005**    **Analyst, Consultant, Manager; Accenture Ltd.**  
Projects included post-merger integration of IT systems, and business design and definition of requirements for an actuarial and underwriting system for a large US health insurer. Led teams of 2-25 people during design and testing of that system. Spent 3.5 months on-site in Chennai, India in 2004 working with offshore development teams. Consistently rated in the top 10% of my peer group. Promoted early twice (2002 and 2005).

## TECHNICAL SKILLS

*Programming languages:* R (proficient, 5 years experience), Python (intermediate, 2 years experience), MySQL, some C, some CUDA

*Algorithms:* topic modeling (LDA and related models); ensemble learning; matching for causal inference

*APIs:* Bing Translate, Bitly, Data Science Toolkit, Twitter, Yahoo Geodict

*Version control:* git (3 years experience)

*Other:* Unix systems administration; compute cluster design and setup; OCR (Tesseract);

## RELEVANT APPLIED RESEARCH PROJECTS

### **Entity disambiguation in international patent databases (in process)**

The PATSTAT database provides a wealth of international patent data, but no standardization of names and other data tying inventors to innovations. We adapt the Bayesian approach taken by Lee Fleming and coauthors to the PATSTAT database. We show that automated methods can reproduce results achieved from a combination of intensively tuned data cleaning and manual disambiguation. Infrastructure: R, Python, MySQL. Technical background: regular expressions, geocoding, text mining. See <https://github.com/markhuberty/psClean> and [https://github.com/markhuberty/fung\\_disambiguator](https://github.com/markhuberty/fung_disambiguator) for more detail.

### **Forecasting Congressional election outcomes from the twitter stream (2010-2012)**

In 2010, I built algorithms for predicting Congressional election outcomes with Twitter data. Out-of-sample tests suggested accuracy rates of greater than 80% were possible. For 2012, we published daily predictions of Congressional election outcomes based on these algorithms, and generated dynamic visualizations to make results accessible to a wide audience. Accuracy once again exceeded 85% for districts with incumbent candidates; and 70% for open races. Infrastructure: R; SuperLearner ensemble machine learning package for R; Twitter API. See [http://markhuberty.berkeley.edu/files/twitter\\_paper.pdf.zip](http://markhuberty.berkeley.edu/files/twitter_paper.pdf.zip) for the original paper and <http://californianewsservice.org/category/tweet-vote> for real-time predictions, visualization, and code.

### **The leghist package for R: Automated analysis of legislative history (alpha release)**

leghist automates the discovery of the origins of passed legislation from amendments and other precursors, the creation of synthetic legislation from candidate precursors, the semantic analysis of legislative text, and visualization of resulting flows of legislative origins to final bills. Infrastructure: R. Technical background: document retrieval, topic modeling, regular expressions, text mining and outlier detection in text. See <https://github.com/markhuberty/leghist> for details.

### **The Political Science Compute Cluster (in production since 2010)**

I designed, built, deployed, and continue to maintain an 18-node parallel compute cluster for social science statistics research. As of February 2012, it had been used for over 1800 jobs ranging in duration from a few minutes to several months. Infrastructure: CentOS (operating system); Perceus (cluster management); SLURM (job scheduling); OpenMPI (message passing); R (statistical computation). See <http://pscluster.berkeley.edu> for more detail.

### **GPU-accelerated genetic matching (proof of concept)**

I ported core C / C++ routines in the Matching package for R to CUDA-enabled GPUs. Proof-of-concept prototyping yielded 40-50% reductions in compute times on small-scale (N ~ 1500) datasets common to social science research. Future work should optimize memory usage to permit larger samples.

## GRANTS AND FELLOWSHIPS

**Fulbright Foundation Fellowship**, European Union, 2010-2011

**U.S. Environmental Protection Agency STAR Fellowship**, 2010-2013

**European Union Center of Excellence Grant**, University of California, Berkeley, 2009

**Institute for European Studies Fellowship**, University of California, Berkeley, 2007-2008

**Foreign Language and Area Studies Fellowship**, German, University of California, Berkeley, 2007-2008

## HONORS AND AWARDS

**Peter H. Odegard Award**, for an outstanding 3<sup>rd</sup>-year graduate student, Department of Political Science, University of California, Berkeley, Spring 2010

**Distinguished Teaching Award**, Department of Political Science, University of California, Berkeley, Fall 2009

**C. Grove Haines Prize**, Best paper in European Studies, The Johns Hopkins University, SAIS, 2006

## PUBLICATIONS

Sample publications provided below. A complete list is available at my website.

### *Working Papers*

#### **Political economy of climate and energy**

Mark Huberty and Georg Zachmann, "Green Exports and the Global Product Space: Prospects for EU Industrial Policy", Bruegel Working Paper 2011/07. May 2011.

"The demand for sea-coales in London: the Great Fire as a technology shock". May 2010.

#### **Social media and politics**

"Voting with your tweet: forecasting elections with social media data". May 2012. Real-time forecasts of the 2012 U.S. Congressional election based on this paper can be found at

<http://californianewsservice.org/category/tweet-vote/>.

### *Peer Reviewed Publications*

"Testing the ownership society: ownership and voting in Britain", *Electoral Studies* 30(4) December 2011.

"Shock and Change in the German Venture Capital Market, 1995-2005", *German Politics and Society* 24(3) Fall 2006.

### *Books*

Mark Huberty and John Zysman. *From Religion to Reality: Energy systems transformation and sustainable prosperity*. Stanford University Press. Forthcoming 2013.

### *Book Chapters*

"Energy systems transformation: state choices at the intersection of sustainability and growth", forthcoming in Dan Breznitz and John Zysman, eds, *Can Wealthy Nations Stay Rich?*. Oxford University Press.

"The Dissolution of Sectors: do politics and sectors still go together?", forthcoming in Dan Breznitz and John Zysman, eds, *Can Wealthy Nations Stay Rich?*. Oxford University Press.

## **PRESENTATIONS**

### *Invited*

Keynote Address, Mandag Morgen Green Growth Leaders Forum, Copenhagen, Denmark. April 2011.

“Green Growth and Energy Systems Transformation”. Tianjin Binhai Eco-Cities Conference. Tianjin, China. September 2012.

### *Academic*

“The political economy of Europe’s energy systems transformation”. Council for European Studies Conference, 2012.

“Varieties of low-emissions innovation”. International Studies Association, 2012.