DSC 180A Data Science Project I Section A02

Wikipedia Edit Wars

Keng-Chi Chang

<kechang@ucsd.edu>

2020-02-05

Plans

- Last week: omitted variable bias, reverse causality
- Assignment 2
- Causal Inference
- Readings

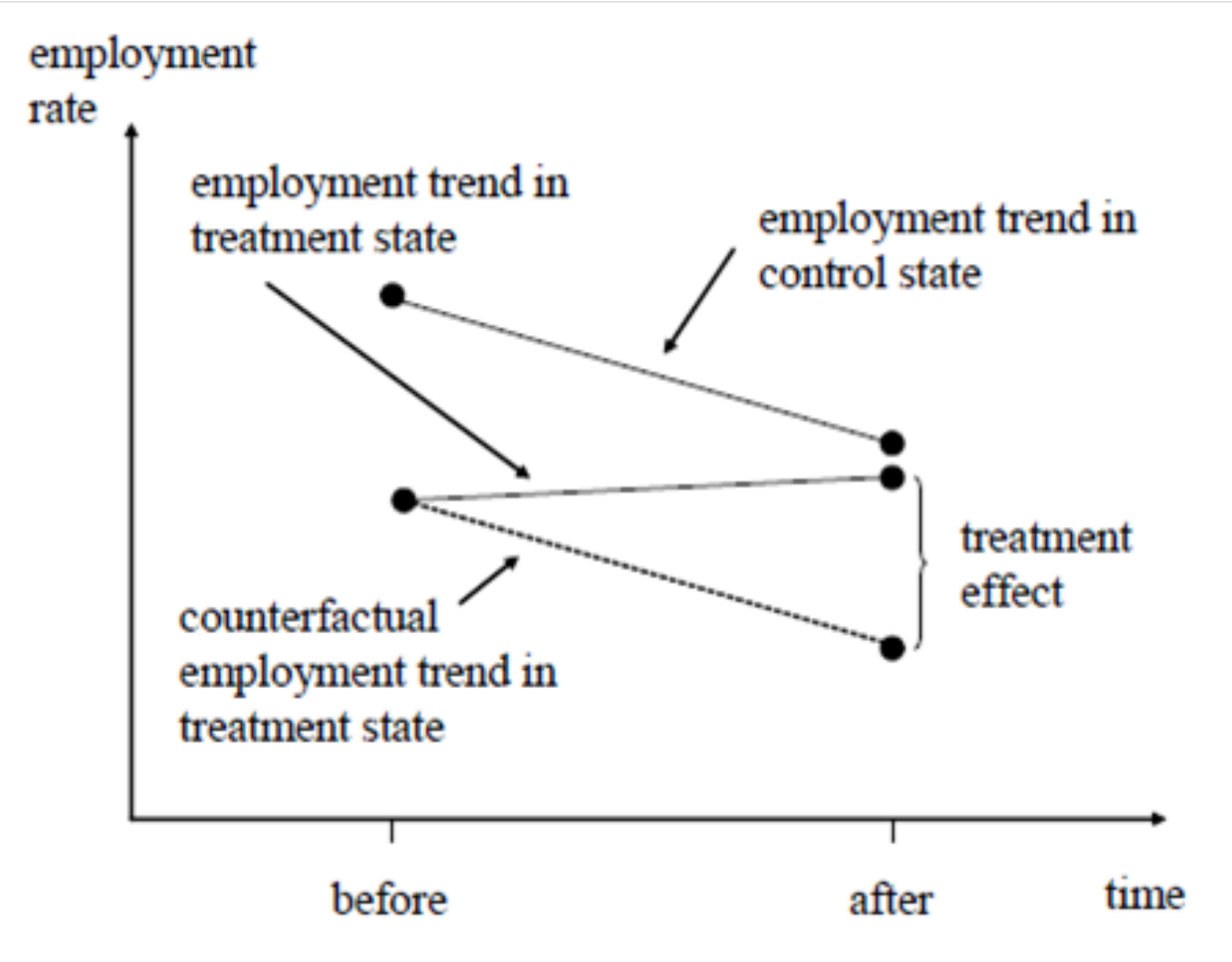
Fundamental Problem of Causal Inference

- We're interested in how change in X causes change in Y
- Observe: Y under some realizations $X = x$
- Problem: We don't know the counterfactual $Y(X = x')$
- Solution: Estimate the missing counterfactual $\widehat{Y}(X = x')$ and recover the causal effect

$$Y(X = x) - \widehat{Y}(X = x')$$

Difference-in-Differences

- Card and Krueger (1994): Minimum wage on employment
- Key assumption: Parallel trends



Wikipedia Censorship

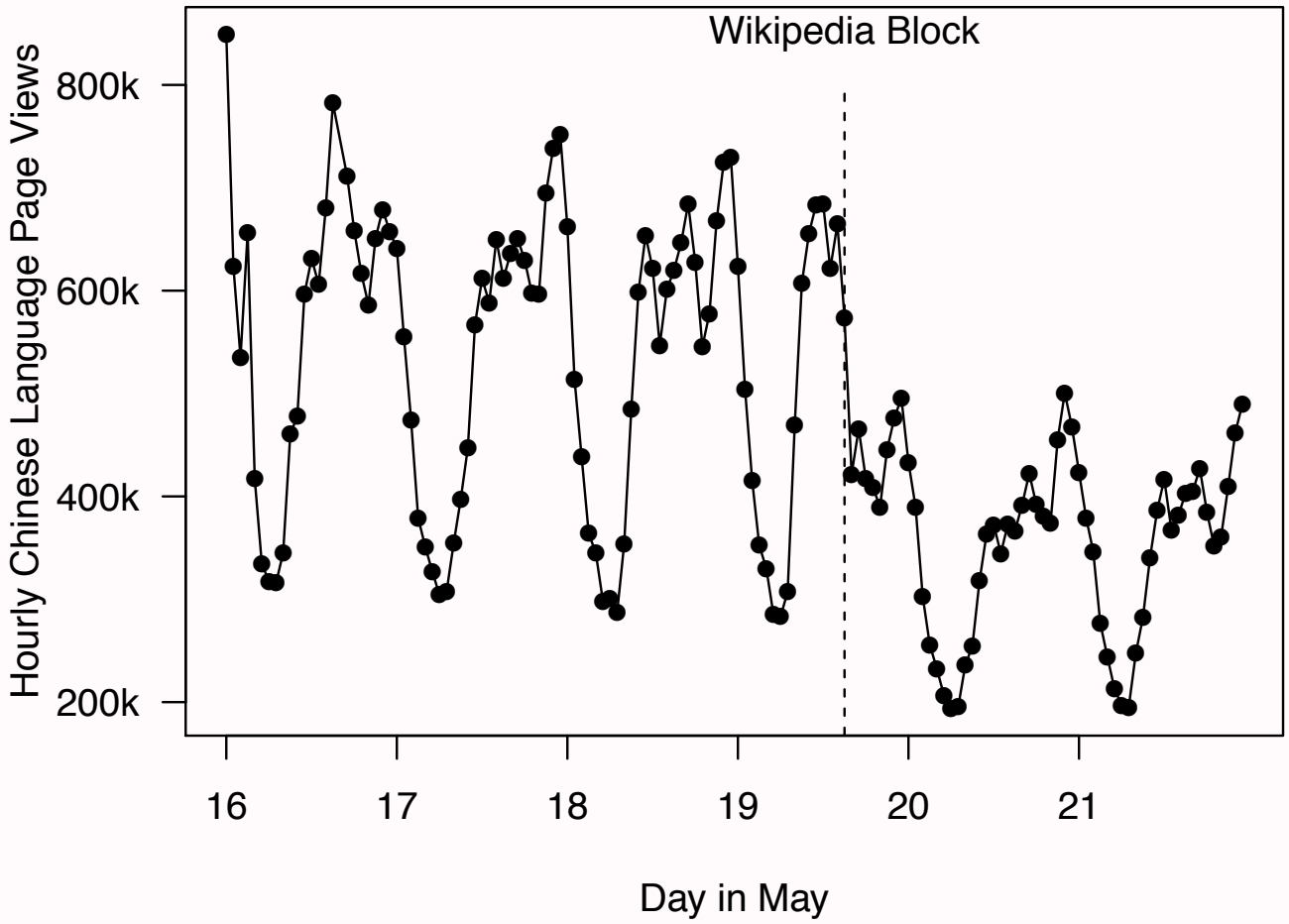
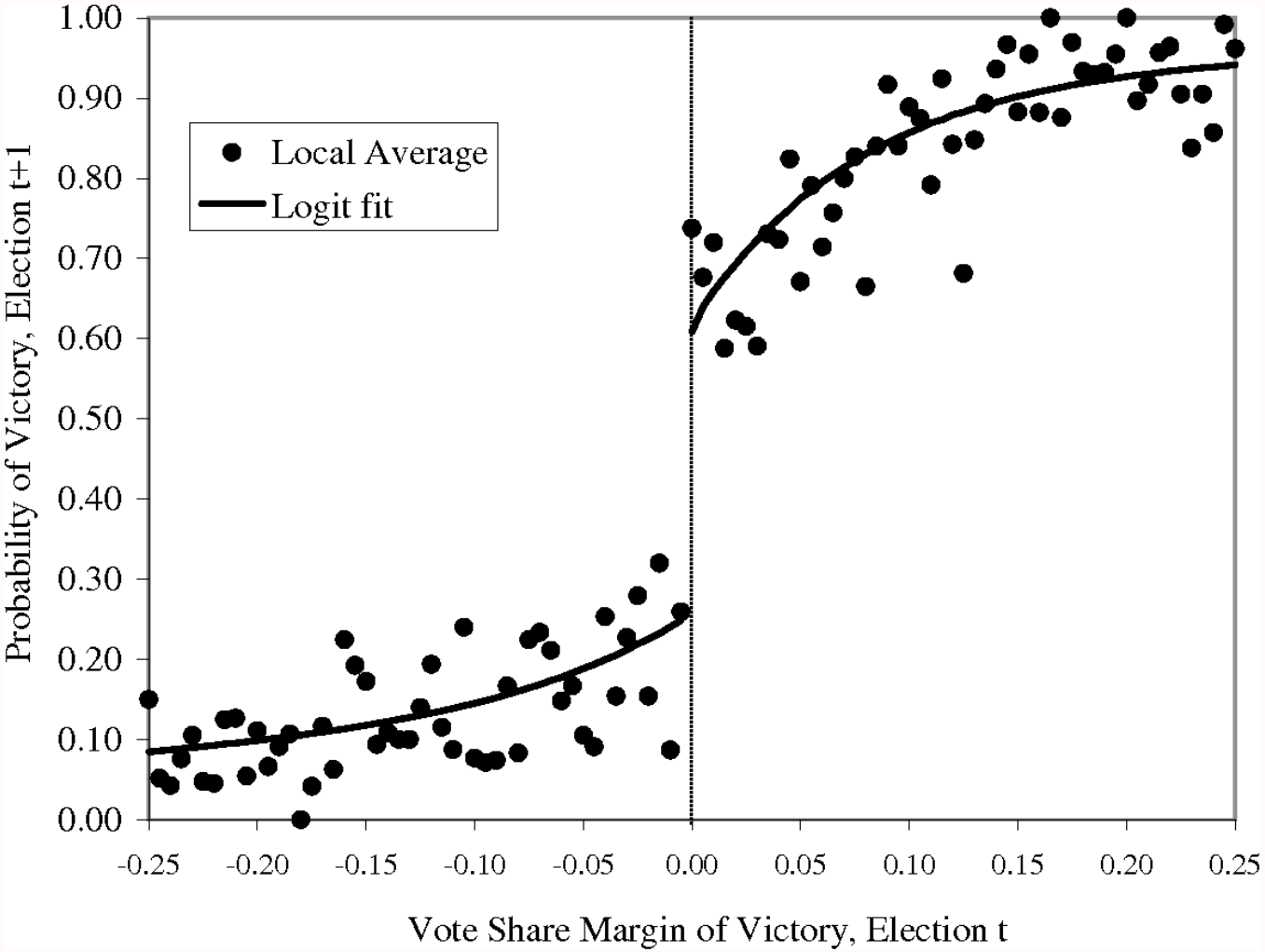


Figure 1. Page views of Chinese language Wikipedia by hour, May 16-21, 2015. The Wikipedia block occurred during the afternoon of May 19, 2015.

Regression Discontinuity Design

- Lee (2007): Incumbent's advantage for elections
- Key assumption: Samples cannot select around cutoff



More on this?

- Matt Blackwell *Causal Inference Lecture Notes*
- Judea Pearl *The Book of Why*
- Angrist and Pischke *Mostly Harmless Econometrics*
- Discuss with me!

Dynamics of Conflicts

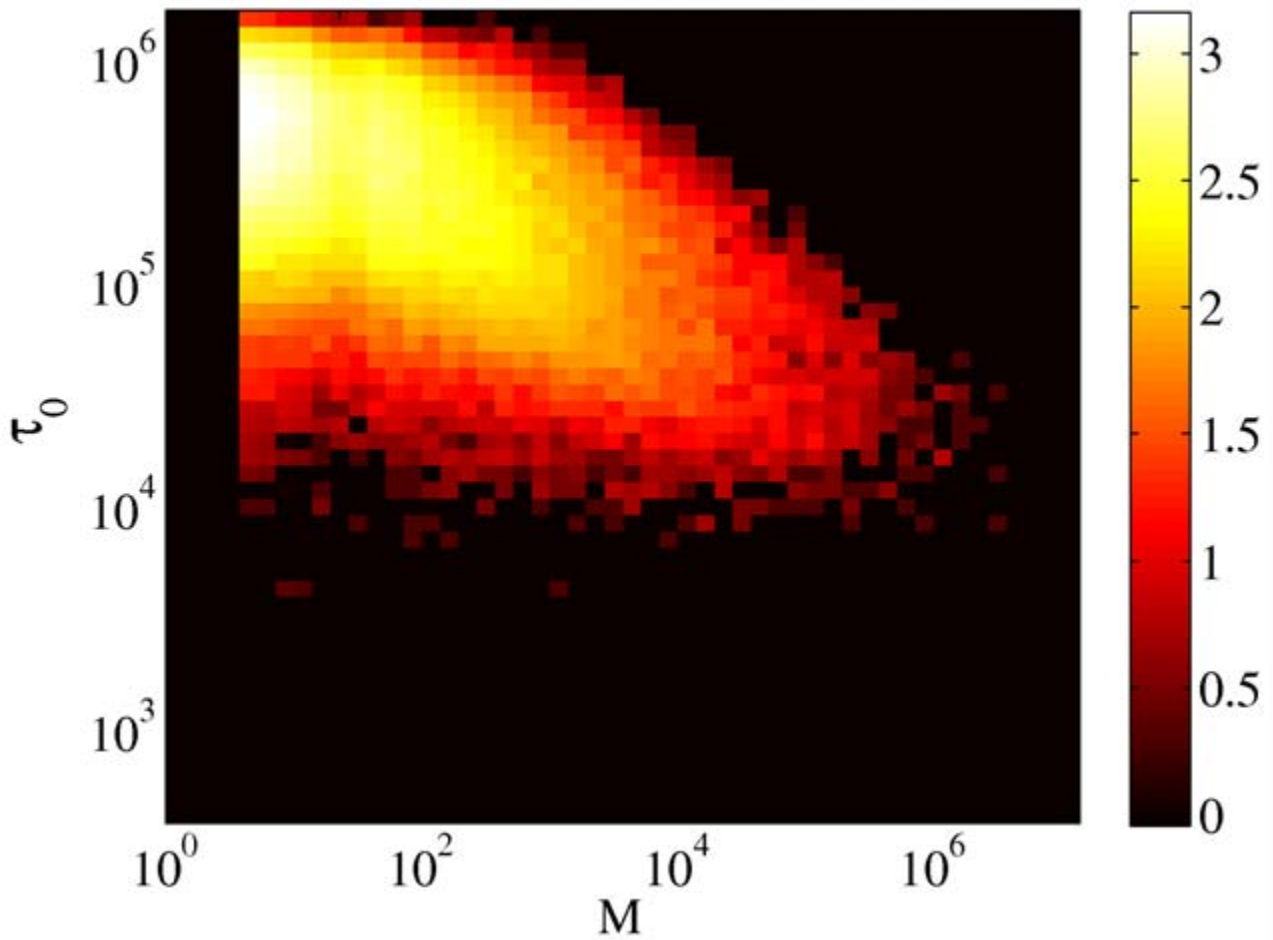


Figure 6. Scatter plot of the average time interval between successive edits and the controversy measure. Color coding is according to logarithm of the density of points. The correlation coefficient $C = -0.03$.

doi:10.1371/journal.pone.0038869.g006

Dynamics of Conflicts

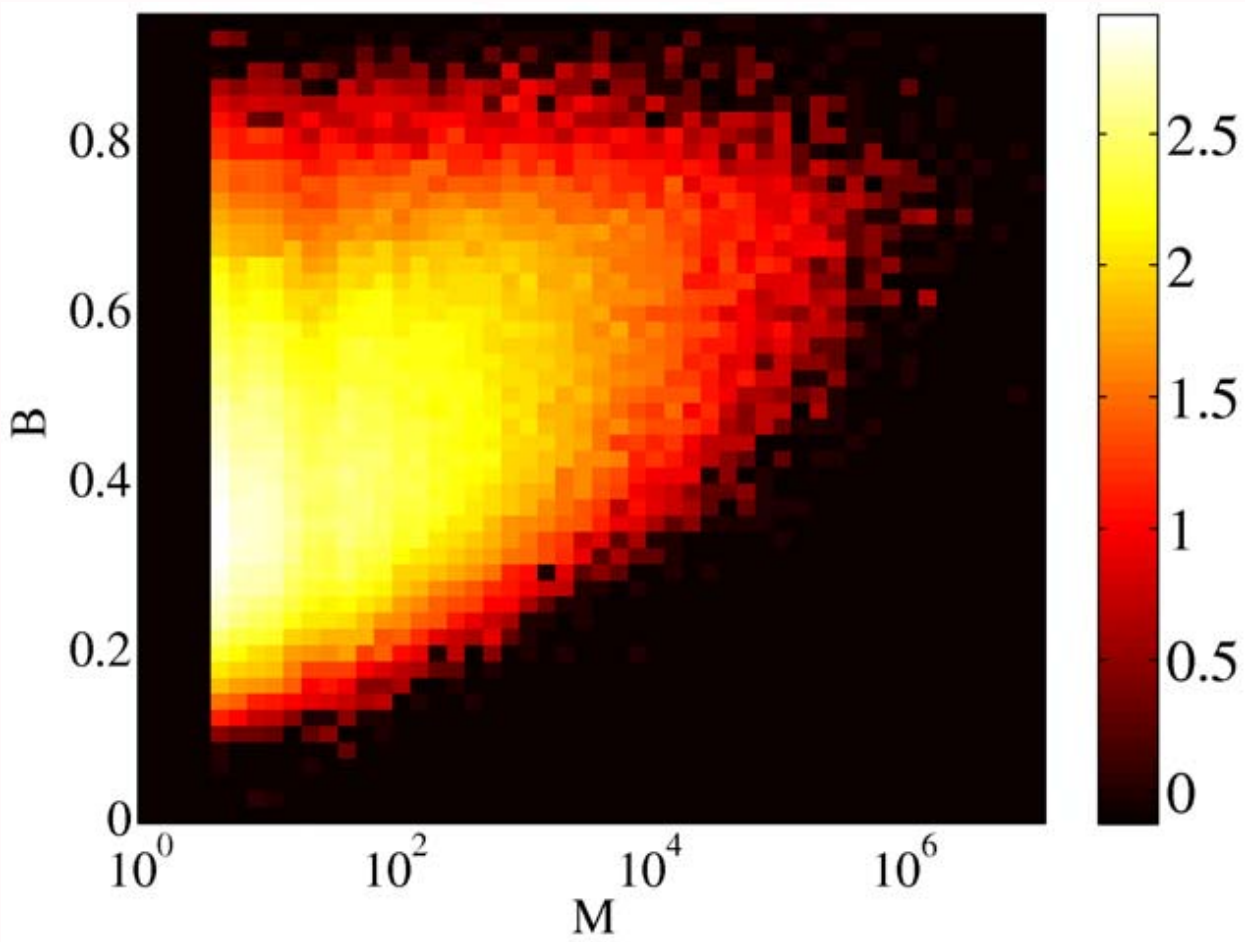


Figure 7. Scatter plot of burstiness and the controversy measure. Color coding according to logarithm of the density of points. The correlation coefficient $C = 0.05$.
doi:10.1371/journal.pone.0038869.g007

Dynamics of Conflicts

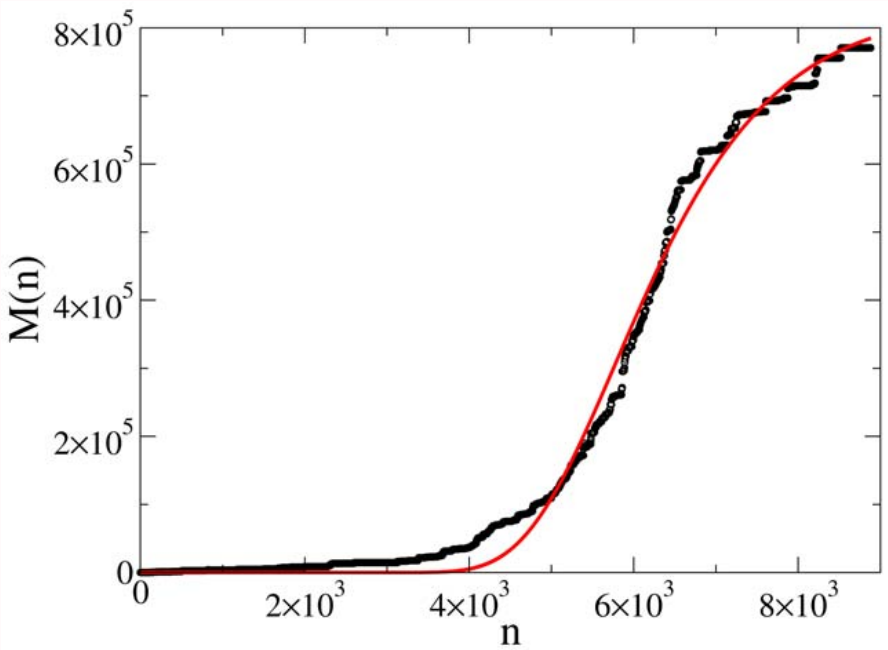


Figure 13. Evolution of controversy measure with number of edits of Jyllands-Posten Muhammad cartoons controversy, with Gompertz fit shown in red. The initial rapid growth in M tends to saturate, corresponding to the reaching to consensus.
doi:10.1371/journal.pone.0038869.g013

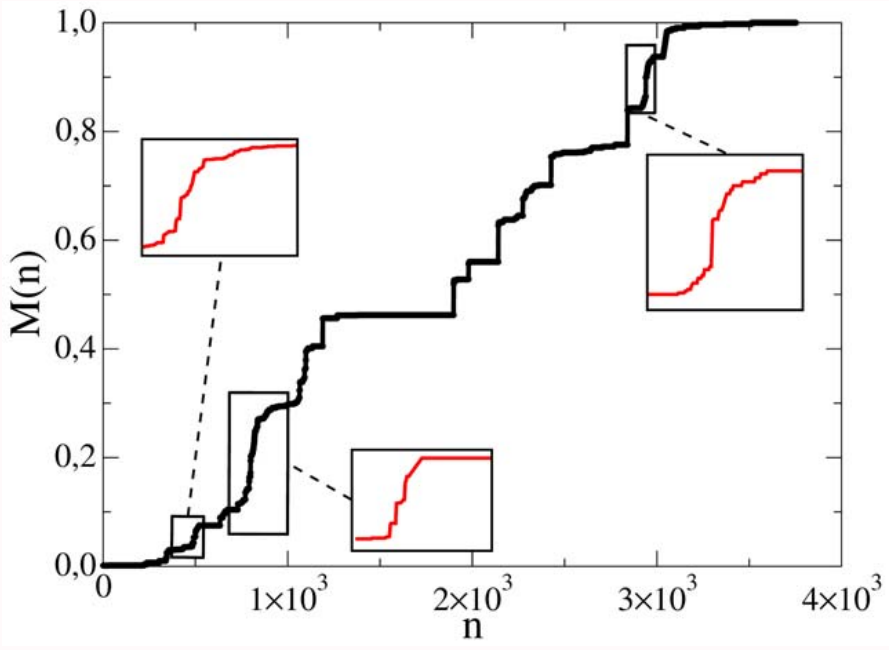


Figure 14. Evolution of controversy measure with number of edits of of Iran – the insets depict focuses of some of the local war periods. $M(n)$ is normalized to the final value M_∞. Cycles of peace and war appear consequently, activated by internal and external causes.
doi:10.1371/journal.pone.0038869.g014

Dynamics of Conflicts

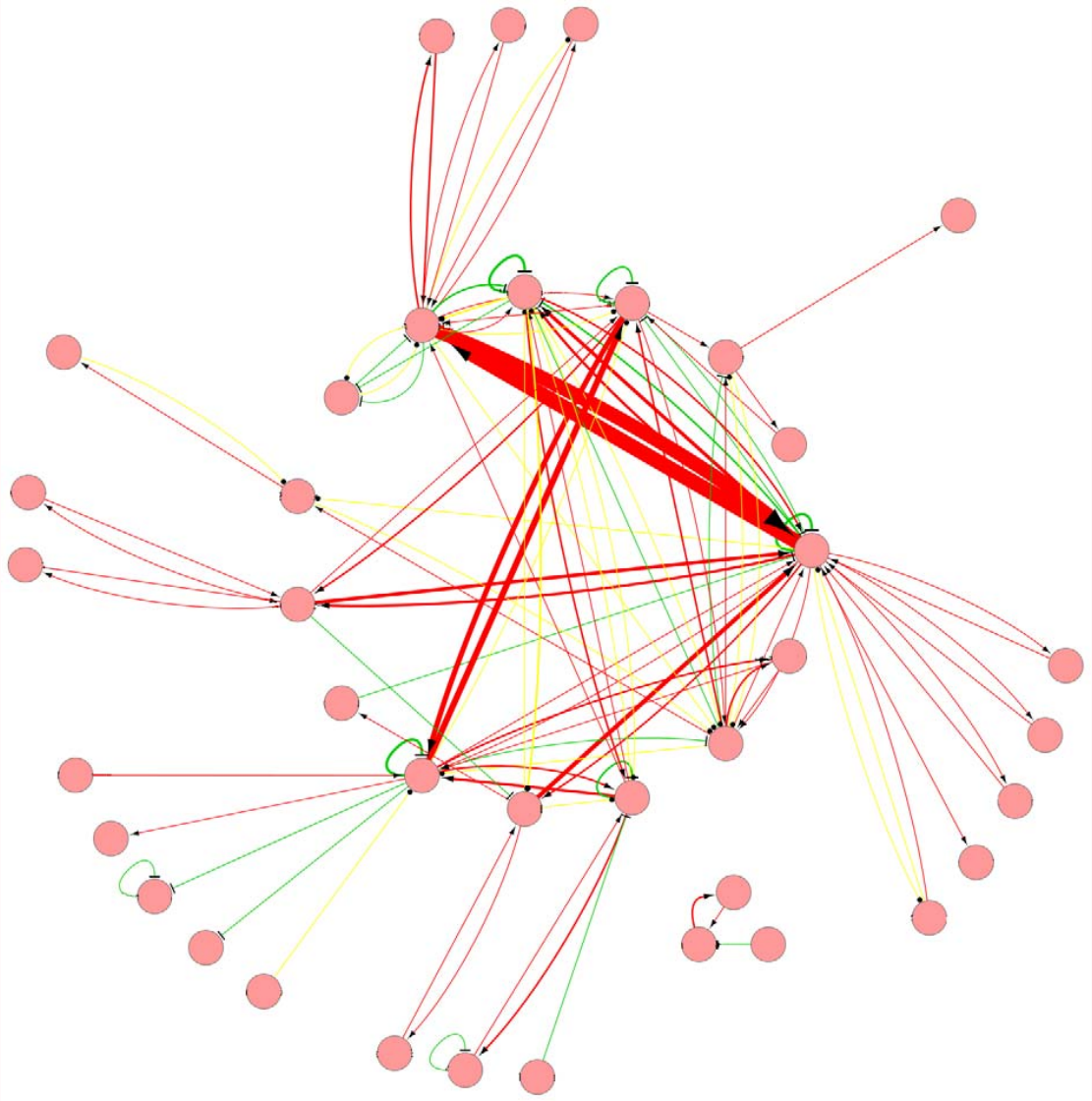


Figure 21. Network representation of editors' interactions in the discussion page of Safavid dynasty.

Multilingual Analysis

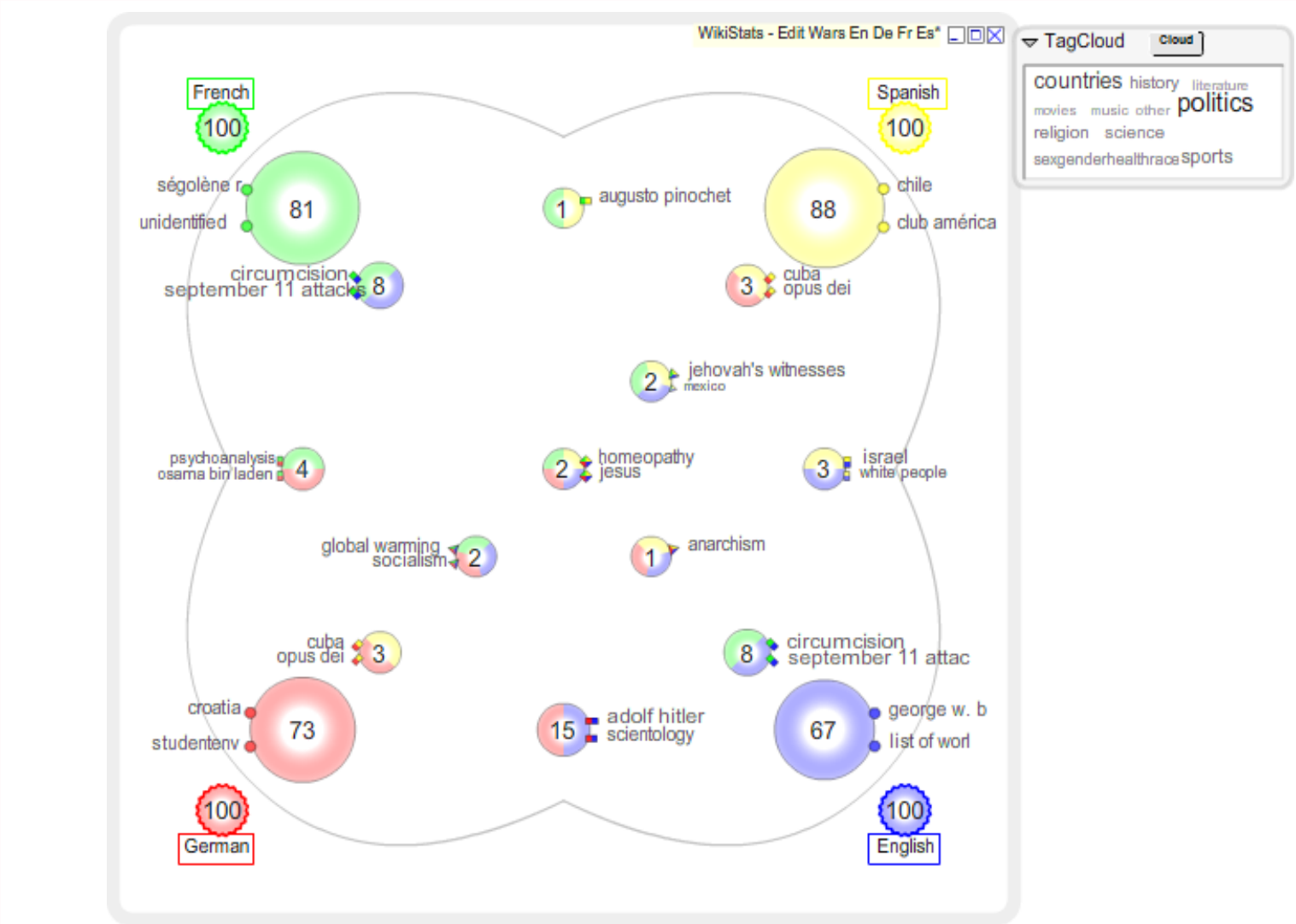


Figure 2 **Category View** of the overlap structure of the most contested Wikipedia pages in *English*, *German*, *French* and *Spanish*.

Geographical Analysis

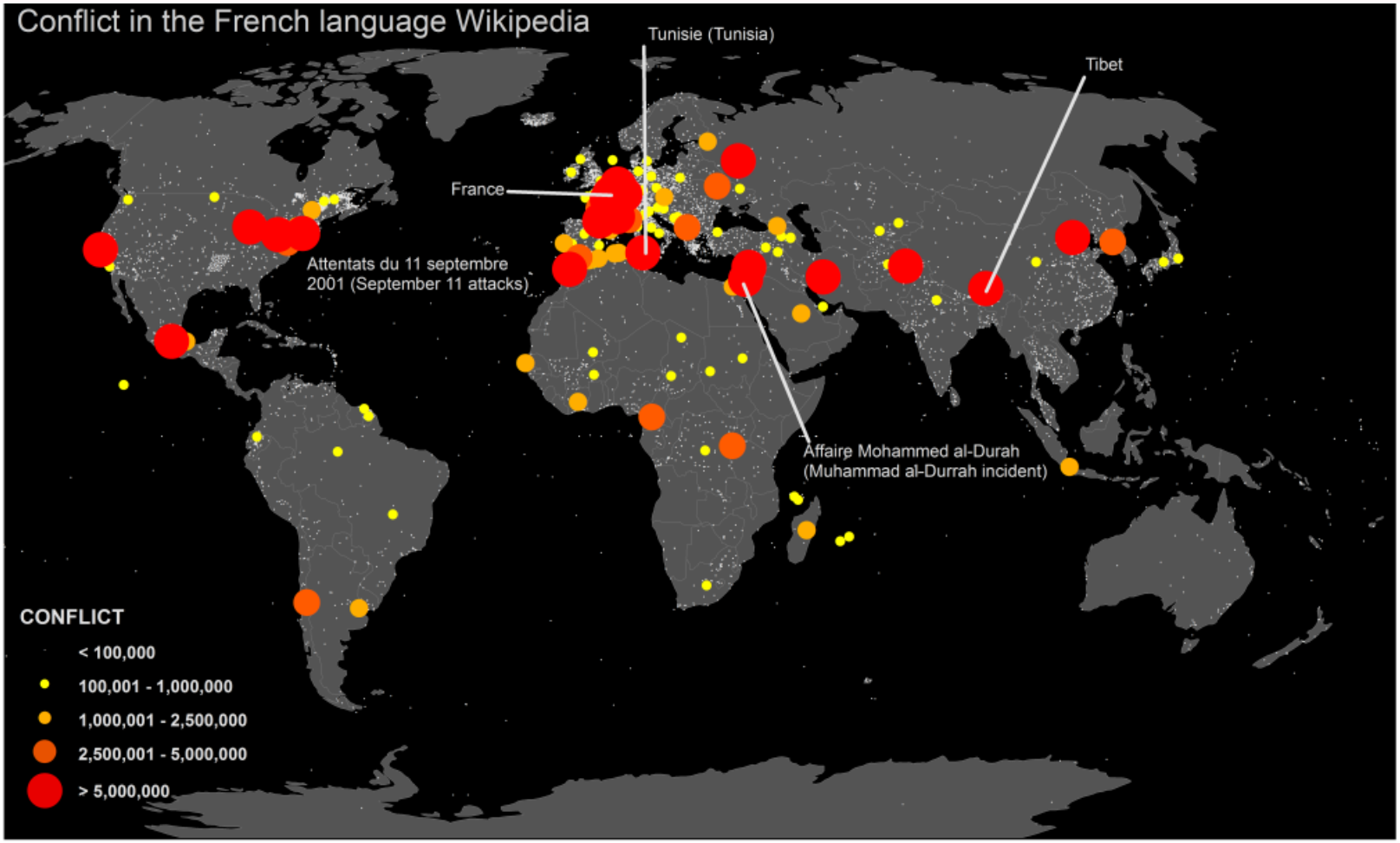


Figure 14 Map of conflict in French edition of Wikipedia. Size of the dots is proportional to the controversy measure M.