

Spatial DNA: Measuring Similarity of Geolocation Data Sets with Applications to Forensics

Christopher Galbraith[†]

Padhraic Smyth[‡]

[†]Department of Statistics

[‡]Department of Computer Science



UCIRVINE
UNIVERSITY *of* CALIFORNIA • IRVINE

July 31, 2019

Acknowledgements



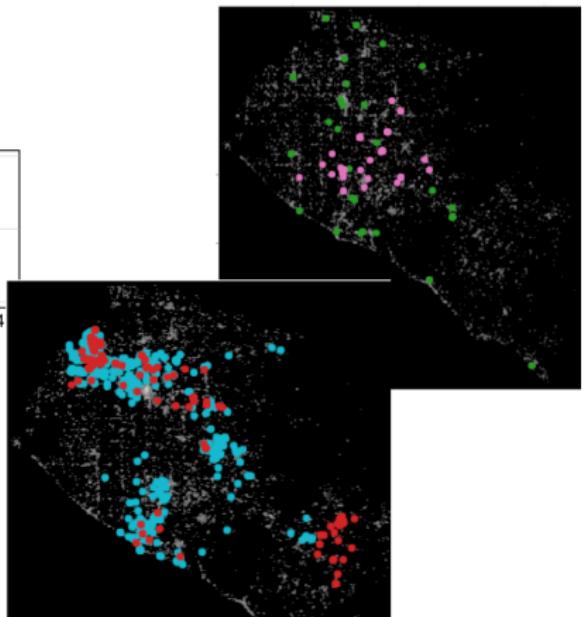
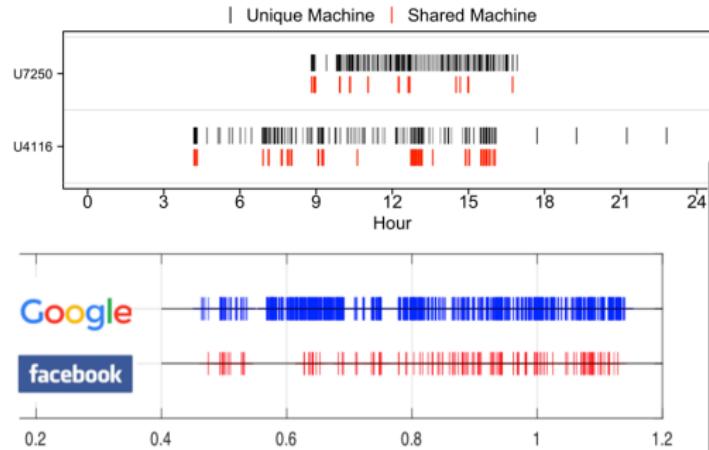
The material presented here is based upon work supported by the National Institute of Science and Technology under Award No. 70NANB15H176. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Institute of Science and Technology, nor of the Center for Statistics and Applications in Forensic Evidence.

Logs of User-Generated Event Data



Problem Statement

Given a pair of user-generated event point patterns (A,B)
quantify the likelihood that the pair was generated by the
same source.

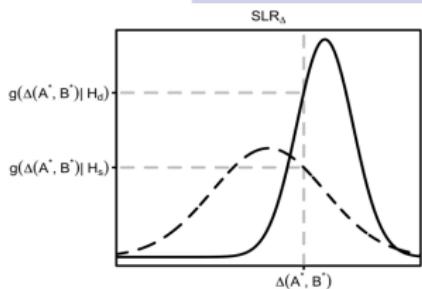


Methodology

(A^*, B^*)
Score Function Δ

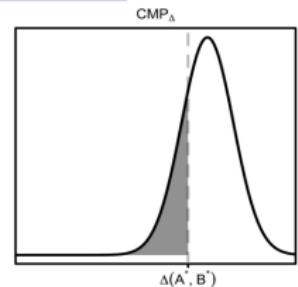
Population-based Approach

- Sample from relevant population:
 $M_i = (A_i, B_i)$ for $i = 1, \dots, N$
- Estimate score-based likelihood ratio (SLR)



Resampling Approach

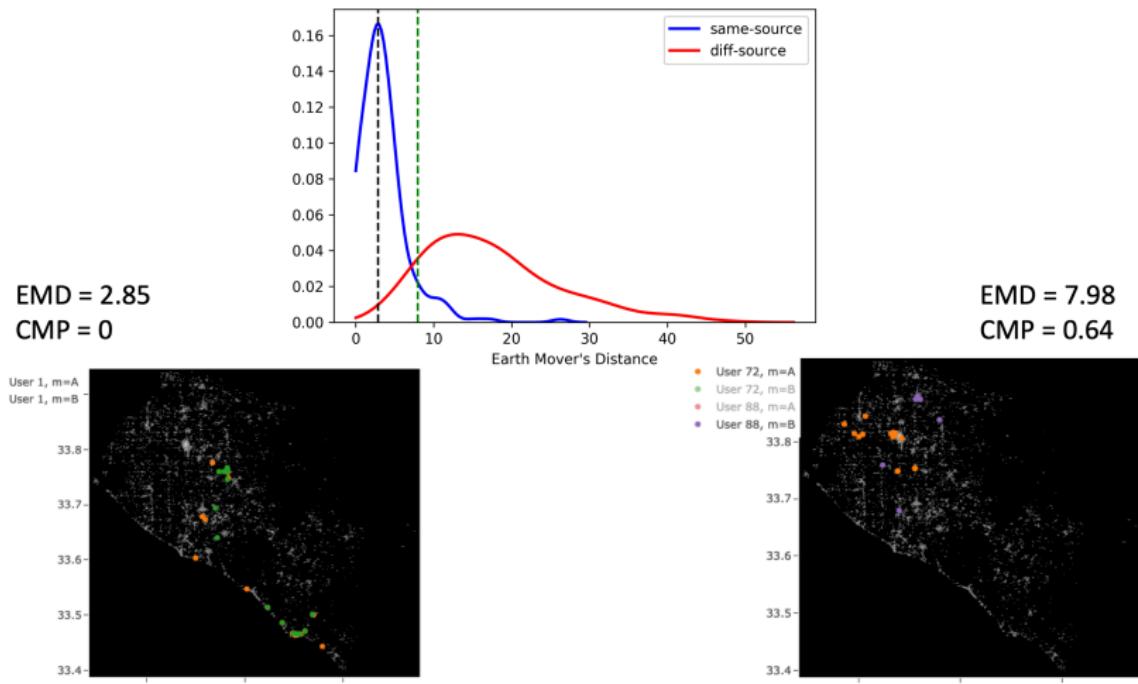
- Single pair: (A^*, B^*)
- Estimate coincidental match probability (CMP)



Degree of Association

Results

- Temporal point patterns, see (Galbraith & Smyth, 2017; Galbraith, Smyth, & Stern, 2019) and R package `assocr`
- Spatial point patterns:



References

- Galbraith, C., & Smyth, P. (2017). Analyzing user-event data using score-based likelihood ratios with marked point processes. *Digital Investigation*, 22, S106 - S114. doi: <https://doi.org/10.1016/j.dii.2017.06.009>
- Galbraith, C., Smyth, P., & Stern, H. S. (2019). Quantifying the association between discrete event time series with applications to digital forensics. *Submitted to J. R. Stat. Soc. A*.