Extremist ideology as a complex contagion: the spread of far-right radicalization in the United States between 2005-2017

Mason Youngblood[a,b,1]

[a]Department of Psychology, The Graduate Center, City University of New York, New York, NY, USA
[b]Department of Biology, Queens College, City University of New York, Flushing, NY, USA
[1]myoungblood@gradcenter.cuny.edu

Supplementary information

0.1 Imputation Check

To ensure that the coding procedure did not bias the estimation of the epidemic predictors, the full twinstim model was re-run after multiple imputation with chained equations using the R package *mice* [1] and random forest machine learning using the R package *missForest* [2]. All four epidemic predictors (plot success, anticipated fatalities, and group membership) were used in fitting and training. The maximum iterations was set to 10 and number of trees was set to 100. The results of 100 rounds of both imputation methods can be seen in Table S1.

	Chained equations		Random forest	
	RR	p-value	RR	p-value
Group membership	5.014	0.0040	6.61	0.00040
Social media	2.29	0.065	4.092	0.00013
Anticipated fatalities	1.15	0.43	0.90	0.53
Plot success	0.93	0.78	1.11	0.55

Table S1: The average rate ratios and p-values for all epidemic predictors after 100 rounds of imputation and estimation using the full model.

Since the observed estimate of social media, the only epidemic predictor with missing data in the best fitting model, is between those from the two imputation methods, and as random forest in *missForest* outcompetes chained equations in *mice* in most [2–7] (but not all [8, 9]) direct comparisons, I assume that the coding method did not significantly influence the results.

0.2 Spatial interaction

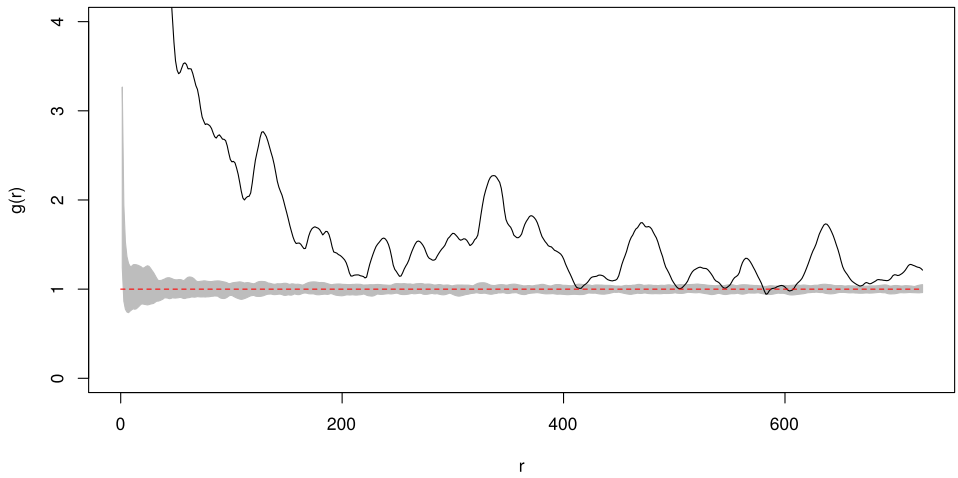

Figure S1: The pair correlation function at different pairwise distances in km (x-axis). The black line is the observed function for the data, the red line is the theoretical function assuming spatial randomness, and the grey envelope shows the upper and lower bounds of the functions from 100 simulated point patterns demonstrating spatial randomness.

0.3 Diagnostics

The residuals, or the fitted cumulative intensities over time, were calculated and transformed to fit a uniform distribution according to Ogata [10]. The cumulative density function diverges from expectations for $U_i < 0.58$, which appears to be the result of tie-breaking with small temporal distances (0.5 days) [11]. Increasing the tie-breaking distance to > 20 days to improve the cumulative density function and reduce serial correlation did not significantly change the predictor estimates, so I chose to use the original model.

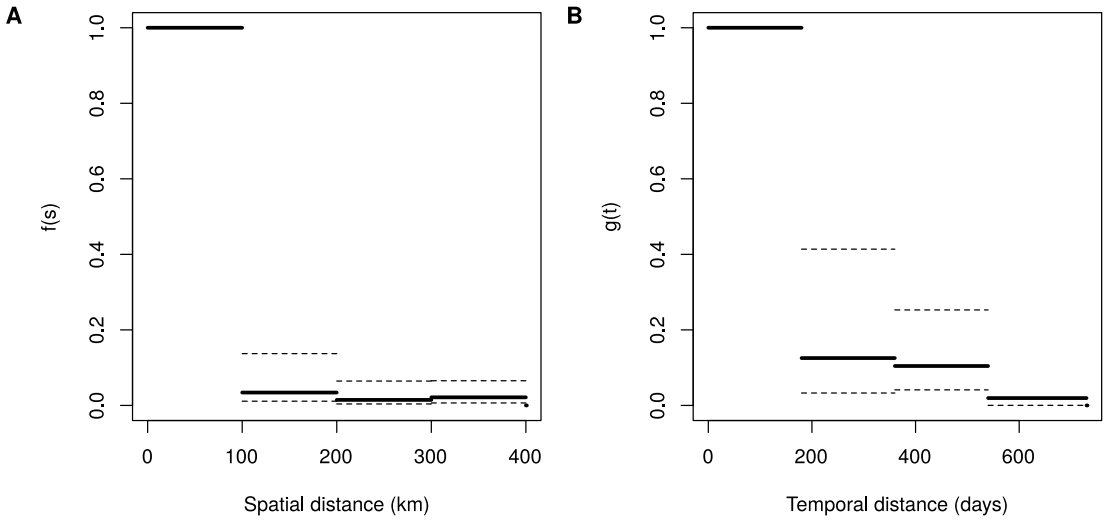

Figure S2: Estimates of the scaled spatial (left panel) and temporal (right) step functions. The 95% Monte Carlo confidence intervals were each calculated from 100 samples.

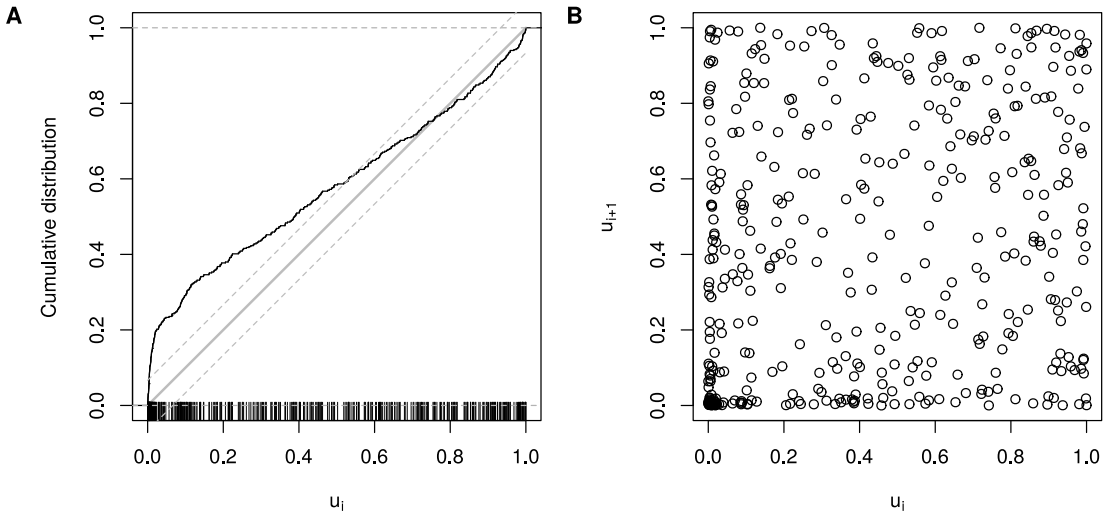

Figure S3: (A) The empirical cumulative density function of U_i, or the standardized residuals according to Ogata [10], with 95% Kolmogorov-Smirnov confidence bands. (B) A scatterplot of U_i and U_{i+1} to look for serial correlation.

References

1. van Buuren, S. & Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* **45,** 1–67 (2011).
2. Stekhoven, D. J. & Bühlmann, P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28,** 112–118 (2012).
3. Waljee, A. K. *et al.* Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open* **3,** 1–7 (2013).
4. Liao, S. G. *et al.* Missing value imputation in high-dimensional phenomic data: Imputable or not, and how? *BMC Bioinformatics* **15,** 1–12 (2014).
5. Muharemi, F., Logofatu, D. & Leon, F. *Review on General Techniques and Packages for Data Imputation in R on a Real World Dataset* in *Computational Collective Intelligence* (eds Nguyen, N. T., Pimenidis, E., Khan, Z. & Trawinski, B.) (Springer, Bristol, UK, 2018), 386–395.
6. Misztal, M. A. Comparison of Selected Multiple Imputation Methods for Continuous Variables – Preliminary Simulation Study Results. *Folia Oeconomica* **6,** 73–98 (2019).
7. Cui, Y. & Wang, J. *Impact of Dimension and Sample Size on the Performance of Imputation Methods* in *Communications in Computer and Information Science* (eds He, J. *et al.*) (Springer, Ningbo, China, 2019), 538–549.
8. Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O. & Hemingway, H. Comparison of random forest and parametric imputation models for imputing missing data using MICE: A CALIBER study. *American Journal of Epidemiology* **179,** 764–774 (2014).
9. Penone, C. *et al.* Imputation of missing data in life-history trait datasets: Which approach performs the best? *Methods in Ecology and Evolution* **5,** 1–10 (2014).
10. Ogata, Y. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association* **83,** 9–27 (1988).
11. Meyer, S., Elias, J. & Höhle, M. A Space-Time Conditional Intensity Model for Invasive Meningococcal Disease Occurrence. *Biometrics* **68,** 607–616 (2012).