

Statistical Hypothesis Testing and Modelling of Peoples' Power: A Causal Study of the #BlackLivesMatter Movement via Hawkes Processes on Social and Mass Media

Alfred Lindström^{1,2} Simon Lindgren³ Raazesh Sainudiin^{1,2}

¹ Department of Mathematics, Uppsala University, Uppsala, Sweden

² Combient Competence Centre for Data Engineering Sciences, Uppsala University, Uppsala, Sweden

³ Department of Sociology, Umeå University, Umeå, Sweden

* Swedish Research Council project 2019-03351 – Project MEP: Meme Evolution Programme in Mathematical Theology & The Wallenberg AI, Autonomous Systems & Software Program funded by Knut and Alice Wallenberg Foundation.

November 29, 2023

We studied the protest in support of the Black Lives Matter (BLM) following the death of George Floyd on 25th of May 2020.

Why these events?

- The decentralized nature of the (BLM), and the way social media has played a key part in its development via the hashtag #BlackLivesMatter.
- Largest protests in U.S. history till then, which also spread internationally.

Question: Is there a causal relation between interactions in social media and reports of street protests in mass media?

- 23.3 million collected tweets from 7.1 million users regarding the Black Lives Matter-movement, along with the smaller counter movements Blue Lives Matter (pro-police movement), and All Lives Matter.
- Tweets posted in the timespan from 24 May 2020 to 30 June 2020.
- 82% of the tweets were some sort of retweeted¹ content.

Note: We only had the tweet data, and thus no data on the relationship of the users in the data set as time-stamped retweet sequences.

Question: Can we still identify which users identified with the “anti-Black Lives Matter-movement”?

¹Retweet - when a user reshares another user's tweet.

Retweet Graph: Map every user to a unique vertex, and let the edges represent that one user has retweeted another.

- From this we could identify the most influential/active users.
- Via the *Label Propagation algorithm* we could identify different communities, one of them being anti-BLM with 26.6 thousand users.
- This anti-BLM community contained quite a few famous persons, non-sympathetic to the BLM movement.

Table: Sample tweets from the pro-BLM and anti-BLM communities.

Pro-BLM community
i can't stand by and continue to live in a world where the color of your skin is an automatic target on my family, friends, and neighbors backs. tri-city we must come together to support our communities. THIS. IS. AMERICA. BE THE CHANGE YOU WANT TO SEE. #blacklivesmatter https://t.co/XIDSNgqx6Q
Thread of people who took it upon themselves to trivialise the current situation going on and #BlackLivesMatter
#BlackLivesMatter Houston is hosting a protest march this FRIDAY at 2PM starting at Discovery Green demanding justice for #GeorgeFloyd White allies, y'all gotta do better and this is a place to start. Everyone who's able should be there. https://t.co/EbWeBrZneP
Aiyana Jones a 7 YEAR OLD CHILD who was shot in the head by an officer, when the officer raided the wrong house. A 7 year old girl didn't deserve to be killed because of disgusting reckless officers. Acab and BLM, never forget this girls name! #BlackLivesMatter https://t.co/HCWzabkFv4
So protest in Huntsville, TX was small, but that was no surprise. We're a small town and most things just caught up to the present on the outside...at the end of the protest on my way home, I saw something I never noticed. This is why we do what we do. #BlackLivesMatter https://t.co/gTuCiIB7mi
Anti-BLM community
Black people are 80 times more likely to kill white people in England/Wales than the reverse! And yet, #BlackLivesMatter more than others? EXPLAIN... Check the stats: https://t.co/DmPDVVGbSo https://t.co/qxXmuNlh2X
#BlackLivesMatter should now be classified as an extreme political hate group.. Simple.. https://t.co/mFh56qCpo9
#DontTakeTheKnee #DontTakeTheKnee please get this trending Sick & tired of the #ScumMedia telling us what we should do! Well I say #DontTakeTheKnee #BLM is a terrorist organisation. Do your homework! #AllLivesMatter #WhiteLivesMatter #ISTANDwithDominic Raab @SkyNews
Then someone gets stabbed and they want the police back after running them out of town. Ha you couldn't make it up #BlackLivesMatter #blm #thugs #brixton https://t.co/1uVXQ63UT2
Just saw a video of #BlackLivesMatter protest in #Reading - looks like 3 white people have been stabbed and in a bad way! Now if this turns out to be a race attack, I'm going to blame the #Media. They've been stoking up tensions between blacks and whites for weeks now!

The GDELT Project

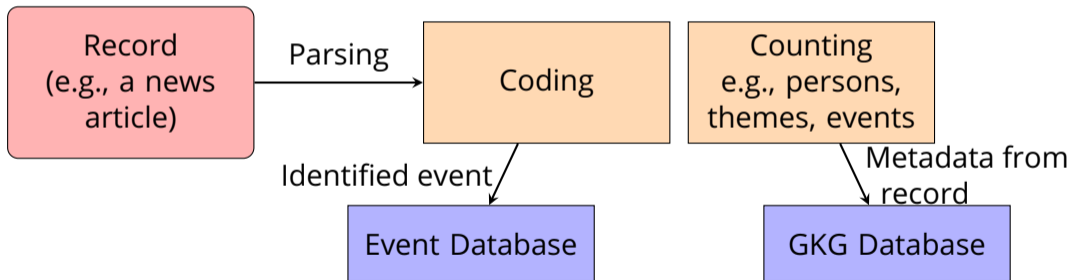
Global Database of Events, Language, and Tone (**GDELT**):

“The GDELT Project is an initiative to construct a catalog of human societal-scale behavior and beliefs across all countries of the world, connecting every person, organization, location, count, theme, news source, and event across the planet into a single massive network that captures what’s happening around the world, what its context is and who’s involved, and how the world is feeling about it, every single day.”

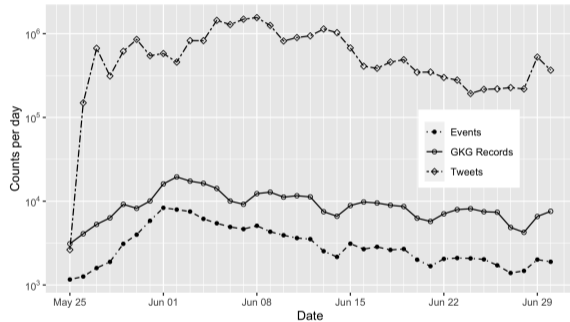
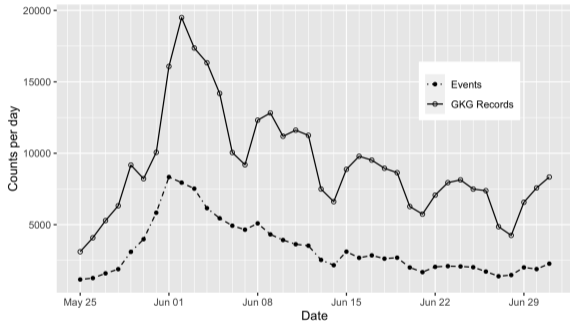
- Monitors the world’s broadcast, print, and web news. Updated every 15 minutes.
- Uses a coding framework that identifies events and actors being reported in these records from mass media.
- Essentially two separate but interlinked databases: The *Global Knowledge Graph* (GKG), consisting of records and the *Event Database*

The GDELT Project

“**President Reagan** has *threatened* further action against **the Soviet Union** in an international television program beamed by satellite to more than 50 countries.”



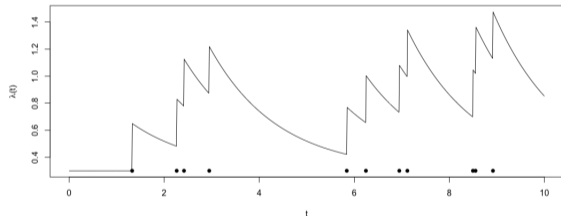
The GDELT Project



Hawkes Processes

- $\lambda(t)$: The *intensity* at time t
- \mathcal{H}_t : *History* up to time t consisting of *timestamps* t_i
- $\phi(t)$: The *kernel*, which regulates how previous events prior to time t affects the rate of new events

$$\lambda(t) = \mu + \sum_{t_i \in \mathcal{H}_t} \phi(t - t_i)$$

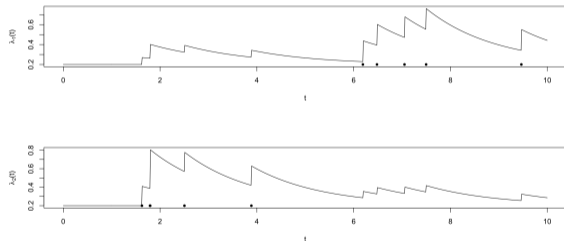


Hawkes Processes

Multivariate case

- $i = 1, \dots, d$: Index for the *dimension*

$$\lambda_i(t) = \mu_i + \sum_{j=1}^d \sum_{t_k \in \mathcal{H}_{t,j}} \phi_{ij}(t - t_k) \quad i = 1, \dots, d$$

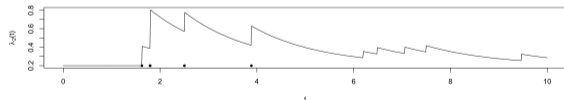
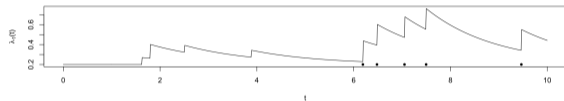


Hawkes Processes

Multivariate case with an Exponential Kernel

$$\lambda_i(t) = \mu_i + \sum_{j=1}^d \sum_{t_k \in \mathcal{H}_{t,j}} \alpha_{ij} \beta_{ij} e^{-\beta_{ij} t} \quad i = 1, \dots, d$$

- μ_i : *Baseline intensity* in dimension i
- α_{ij} : *Excitation rate*, regulates the change of intensity in dimension i , w.r.t. prior events in dimension j
- β_{ij} : *Decay rate*



Hawkes Processes

Granger Causality

Let X_t, Y_t be stochastic processes. X_t Granger-causes variable Y_t , then the past values of X_t contain information that helps predict future values of Y_{t+1} better than doing prediction based only on past values of Y_t .

Theorem

Let $N(t)$ be a multivariate Hawkes process in d dimensions, with kernels $\phi_{ij}(t)$, $i, j \in \{1, \dots, d\}$. Then the j -th component N_j does not Granger-cause the i -th component N_i if and only if $\phi_{ij} = 0, \forall t \in \mathbb{R}$.

\Rightarrow Thus, when $N(t)$ is a multivariate Hawkes process with exponential kernel, by Theorem 1 the j -th component N_j does not Granger-cause the i -th component N_i if and only if $\alpha_{ij} = 0$.

Question: Is there a causal relation between interactions in social media and reports of street protests in mass media?

- Timeline: 24-hours-long period between midnight of 28 May and midnight of 29 May.
- Twitter (dimension 1): Only original (no retweets) tweets from pro-BLM users with at least 1 retweet. → 10,774 tweets.
- GDELT (dimension 2): Records from GKG with containing theme "PROTEST", and mentioning George Floyd → 3,341 records

The data was fitted using python library ticks.

- Requires the the decay parameters β_{ij} as input.
- The baselines μ_i and excitation parameters α_{ij} can then be efficiently computed via convex optimisation.

Joint Media Modeling

Model comparison

$$\mathcal{M}_1: \beta_{ij} = \beta \in (0, \infty), \forall (i, j) \in \{1, 2\}^2$$

Number of parameters = 7

$$\hat{\beta} = \begin{pmatrix} 6.17 & 6.17 \\ 6.17 & 6.17 \end{pmatrix}$$

Maximum log-likelihood = 384.771

AIC = **-755.542**

$$\mathcal{M}_2: \beta_{ij} \in (0, \infty), \forall (i, j) \in \{1, 2\}^2$$

Number of parameters = 10

$$\hat{\beta} = \begin{pmatrix} 16.170 & 3.702 \\ 8.638 & 16.170 \end{pmatrix}$$

Maximum log-likelihood = **384.772**

AIC = -749.544

- The β_{ij} were found via simplectic grid search w.r.t. the maximum log-likelihood.
- Relative likelihood of the models=0.04984
- Likelihood-ratio test statistic = 0.002121
- p -value= 0.9971 implying that we cannot reject \mathcal{M}_1 favouring \Downarrow_2

Estimated parameters using \mathcal{M}_1 :

$$\hat{\boldsymbol{\mu}} = \begin{pmatrix} 1.000 \\ 0.998 \end{pmatrix}$$

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 6.17 & 6.17 \\ 6.17 & 6.17 \end{pmatrix}$$

$$\hat{\boldsymbol{\alpha}} = \begin{pmatrix} 0.986 & 0.0327 \\ 0.0216 & 0.921 \end{pmatrix}$$

The following null hypotheses were proposed:

- $H_{0,12} : \alpha_{12} = 0$, i.e., reports of protests in mass media do not Granger-cause communication events in Twitter related to the BLM-movement.
 - $H_{0,21} : \alpha_{21} = 0$, i.e., communication events in Twitter related to the BLM-movement do not Granger-cause reports of protests in mass media.
 - $H_0 : \alpha_{12} = \alpha_{21} = 0$.
-
- To get the confidence intervals for α_{12}, α_{21} we did a non-parametric bootstrap.
 - 99-th percentile bootstrapped CI for $\alpha_{12} : (0.000, 0.09405) \Rightarrow$ We cannot reject $H_{0,12}$.
 - 99-th percentile bootstrapped CI for $\alpha_{21} : (0.01479, 0.02949) \Rightarrow$ We do reject $H_{0,21}$.

- To estimate Type I errors (rejecting the null hypothesis when it is true), we simulated data from the null H_0 , by restricting $\alpha_{12} = \alpha_{21} = 0$, and then repeating the hypothesis testing via bootstrapping.
- Only one out of 100 such simulations from H_0 was rejected giving 0.01 as the Monte Carlo estimate of the Type I error.

Conclusion

- We reject the null hypothesis that there is no causal relationship, and show that communication events in Twitter, surrounding tweets that supported the BLM movement, Granger-caused the reports of street protests in mass media from the GDELT project.
- However, we cannot show that the reporting of street protests in mass media Granger-caused the corresponding communication events in Twitter.
- We identified such pro-BLM tweets thorough a network analysis of the Twitter data to identify communities of users who have a shared ideology among an ideologically diverse set of communities.
- **<https://github.com/lamastex/mep>**
- **<https://github.com/lamastex/spark-gdelt>**