

RESEARCH ARTICLE

Tracking Protests Using Geotagged *Flickr* Photographs

Merve Alanyali*, Tobias Preis, Helen Susannah Moat

Data Science Lab, Behavioural Science, Warwick Business School, University of Warwick, Coventry, CV4 7AL, United Kingdom

* m.alanyali@warwick.ac.uk



OPEN ACCESS

Citation: Alanyali M, Preis T, Moat HS (2016) Tracking Protests Using Geotagged *Flickr* Photographs. *PLoS ONE* 11(3): e0150466. doi:10.1371/journal.pone.0150466

Editor: Lidia Adriana Braunstein, IFIMAR, UNMdP-CONICET, ARGENTINA

Received: February 27, 2015

Accepted: February 15, 2016

Published: March 1, 2016

Copyright: © 2016 Alanyali et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Flickr data are available via the <https://www.flickr.com/services/api/>. The Guardian Data are available via <http://open-platform.theguardian.com/>.

Funding: MA was funded by University of Warwick Chancellor's International Scholarship (<http://www2.warwick.ac.uk/>). TP and HSM were supported by the Research Councils UK Grant EP/K039830/1 (<http://www.rcuk.ac.uk/research/xrcprogrammes/digital/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Recent years have witnessed waves of protests sweeping across countries and continents, in some cases resulting in political and governmental change. Much media attention has been focused on the increasing usage of social media to coordinate and provide instantly available reports on these protests. Here, we investigate whether it is possible to identify protest outbreaks through quantitative analysis of activity on the photo sharing site *Flickr*. We analyse 25 million photos uploaded to *Flickr* in 2013 across 244 countries and regions, and determine for each week in each country and region what proportion of the photographs are tagged with the word “protest” in 34 different languages. We find that higher proportions of “protest”-tagged photographs in a given country and region in a given week correspond to greater numbers of reports of protests in that country and region and week in the newspaper *The Guardian*. Our findings underline the potential value of photographs uploaded to the Internet as a source of global, cheap and rapidly available measurements of human behaviour in the real world.

Introduction

Smartphones and computers are becoming an indispensable part of everyday life in many countries around the globe. Usage of these devices and the online services they connect us to is generating fast and cheap measurements of human behaviour at a global scale. Research in the growing interdisciplinary field of computational social science [1–7] has begun to draw on this rich new data source, exploiting data from search engines such as *Google* [8–13] and *Yahoo* [14, 15], the online encyclopedia *Wikipedia* [16, 17], news sources such as the *Financial Times* [18] as well as social media platforms including *Twitter* [19–22] and the photo sharing website *Flickr* [23–26]. Studies to date have demonstrated that appropriate analyses of these online datasets can offer estimates of key economic and health indicators before official figures are released [10, 27–30] and in some cases, improve forecasts of real world economic decision making [13–15, 17–19].

In recent years, news reports have described a number of prominent outbursts of protests in countries around the world, in some cases leading to political change. Much media attention has been focused on the increasing usage of social media to coordinate and provide instantly

Competing Interests: The authors have declared that no competing interests exist.

available reports on these protests [31–33]. As a result of improved connectivity, posts to social media sites are steadily beginning to shift from solely text based reports to sharing of visual media such as photographs and videos. Here, we explore whether the data created through such widespread usage of online services may offer a valuable new source of measurements of behaviour during protests. Specifically, we investigate whether data on photographs uploaded to the photo sharing website *Flickr* can be used to identify protest outbreaks around the world.

Materials and Methods

We analyse a large corpus of metadata on the 24,944,764 geotagged photographs taken and uploaded to *Flickr* between 1st January 2013 and 31st December 2013. We retrieved data on image uploads to *Flickr* by accessing the *Flickr* API in January 2014, and downloading data in JSON format using R 3.0.1. The metadata we analyse comprise a wide range of information on where and when a photograph was taken, information about the photographer, as well as user chosen title, description and tags for each photograph, and the URL from which the photograph can be downloaded.

For each geotagged photograph, we retrieve data on both the time and the place at which the photograph was taken. For each week, for each of the 242 countries and regions listed in [S1 Table](#), as well as the United Kingdom and the United States, we determine how many photographs were taken and uploaded with the word “protest” in English in either the title, photograph description or photograph tag. We also translate the word “protest” into 33 further languages, by accessing the “Protest” article on the English language *Wikipedia*, and using the title of all articles on versions of *Wikipedia* which are not in English, but which are linked as translations of the article. The complete list of translations is provided in [S2 Table](#). The counts of photographs taken and shared on *Flickr* throughout 2013 in each of the 244 countries and regions are listed in [S3 Table](#).

The overall number of photos taken and uploaded to *Flickr* in different countries and regions may differ. To account for this, we extract the total number of photos taken and uploaded during each week in 2013 for each of the 244 countries and regions analysed. We consider a week as starting on a Monday and ending on a Sunday. Using these counts, we normalise the weekly counts of photographs taken in each country and region in each week, by dividing the number of photographs labelled with a word signifying “protest” by the weekly count of all photos taken in the same country and region.

To determine whether we can find any evidence that changes in the number of protest-tagged photographs taken and uploaded to *Flickr* correspond to changes in the number of protest outbreaks, we require data on when and where protests have occurred. Such ground truth data can be difficult to obtain. Most studies of civil unrest therefore rely on data from newspaper reports as a proxy for ground truth [22, 34–36]. Following this approach, here we determine how many protest related articles for each of the 244 countries and regions were published in the online edition of *The Guardian* in each week in 2013.

We retrieved data on articles in the online edition of *The Guardian* via *The Guardian Developer Toolbox* in January 2016. We deem an article as protest related if it is tagged with the word “protest”, and we deem an article as covering news related to one of the 244 countries and regions analysed if it is tagged with the country and region’s name. To account for differences in coverage of news in different countries and regions by *The Guardian*, we also determine the total number of articles published in each week and tagged with each country’s name. In total, we analyse data on 178,730 articles from *The Guardian*. The counts of *The Guardian* articles published in 2013 and tagged with each of the of the country and region names are listed in [S4 Table](#). We note that *The Guardian* uses a different tagging system for articles

relating to the United Kingdom and the United States. For this reason, we determine the number of articles relating to the United Kingdom by counting articles tagged with the names “England”, “Scotland”, “Wales” and “Northern Ireland”, and we determine the number of articles relating to the United States by counting articles tagged “US news”.

In order to model the relationship between *Flickr* user activity and protest outbreaks, we build a logistic regression panel model. The outcome variable is whether a *The Guardian* article is protest related or not. To control for underlying differences in the number of protests in a given country and region and week, we include country and region and week as fixed effects in our model. We underline that in this analysis, we focus on the relationship between *Flickr* activity and protest reports within the same week. Future analyses may wish to investigate whether photographic data can be used to predict protest activity before it occurs.

Results

We use data on reports of protests in the online edition of *The Guardian* as an approximation of the ground truth of when and where protest outbreaks occurred. For each of the 244 countries and regions analysed, for each month in 2013, we calculate the number of *The Guardian* articles tagged with the country and region’s name. In [Fig 1](#), we depict the percentage of articles for each country and region and each month which were also tagged with the word “protest”. Patterns which can be visually identified in the data reflect known major protest events in 2013: for example, protest outbreaks in both Brazil and Turkey can be observed in June 2013.

We examine to which extent data on the number of photographs tagged with the word “protest” and uploaded to *Flickr* reflect the ground truth data extracted from *The Guardian*. Again, for each of the 244 countries and regions listed in [S1 Table](#), for each month in 2013, we calculate the total number of geotagged photographs taken and uploaded to *Flickr*. In [Fig 2](#), we visualise the percentage of photographs for each country and region and each month which were also labelled with a word signifying “protest” in one of the 34 languages identified above and listed in [S2 Table](#). Visual inspection suggests that while there are clear differences between the spatio-temporal distributions of “protest” labelled *Flickr* photographs and “protest” labelled articles in *The Guardian*, some key similarities can also be identified, such as an increase in “protest” labelled *Flickr* photographs in Brazil and Turkey in June 2013.

To determine whether we can find statistical evidence of a relationship between the number of “protest” labelled photographs taken and uploaded to *Flickr* and reports of protests in *The Guardian*, we consider both datasets at weekly granularity. For each week in 2013, for each country and region, we calculate the number of geotagged photographs taken and uploaded to *Flickr* which are labelled with the character sequence “protest” in 34 different languages, and normalise this count by the total number of geotagged photographs taken and uploaded to *Flickr* in that week and country and region. To analyse the relationship between the data mined from *Flickr* and reports of protests in *The Guardian*, we build a logistic regression panel model. To account for unobserved differences in coverage between countries and regions and weeks, we include country and region and week as fixed effects.

Our results suggest that a greater normalised number of “protest” labelled *Flickr* photographs in a given week and country and region corresponds to a greater proportion of *The Guardian* articles about that country and region being tagged with the word “protest” (*Flickr* predictor: $\beta = 2.95$, $SE = 0.31$, $z = 9.48$, $N = 12932$, $p < 0.001$). The odds ratio corresponding to an increase of 0.1 in the normalised number of “protest” tagged *Flickr* photos is 1.34 (calculated using β and the value of the logistic regression intercept, -5.09). This implies that if we fix the country and region and week effects, increasing the normalised number of “protest” tagged *Flickr* pictures by 0.1 will increase the odds of a protest related *The Guardian* article by 34%.

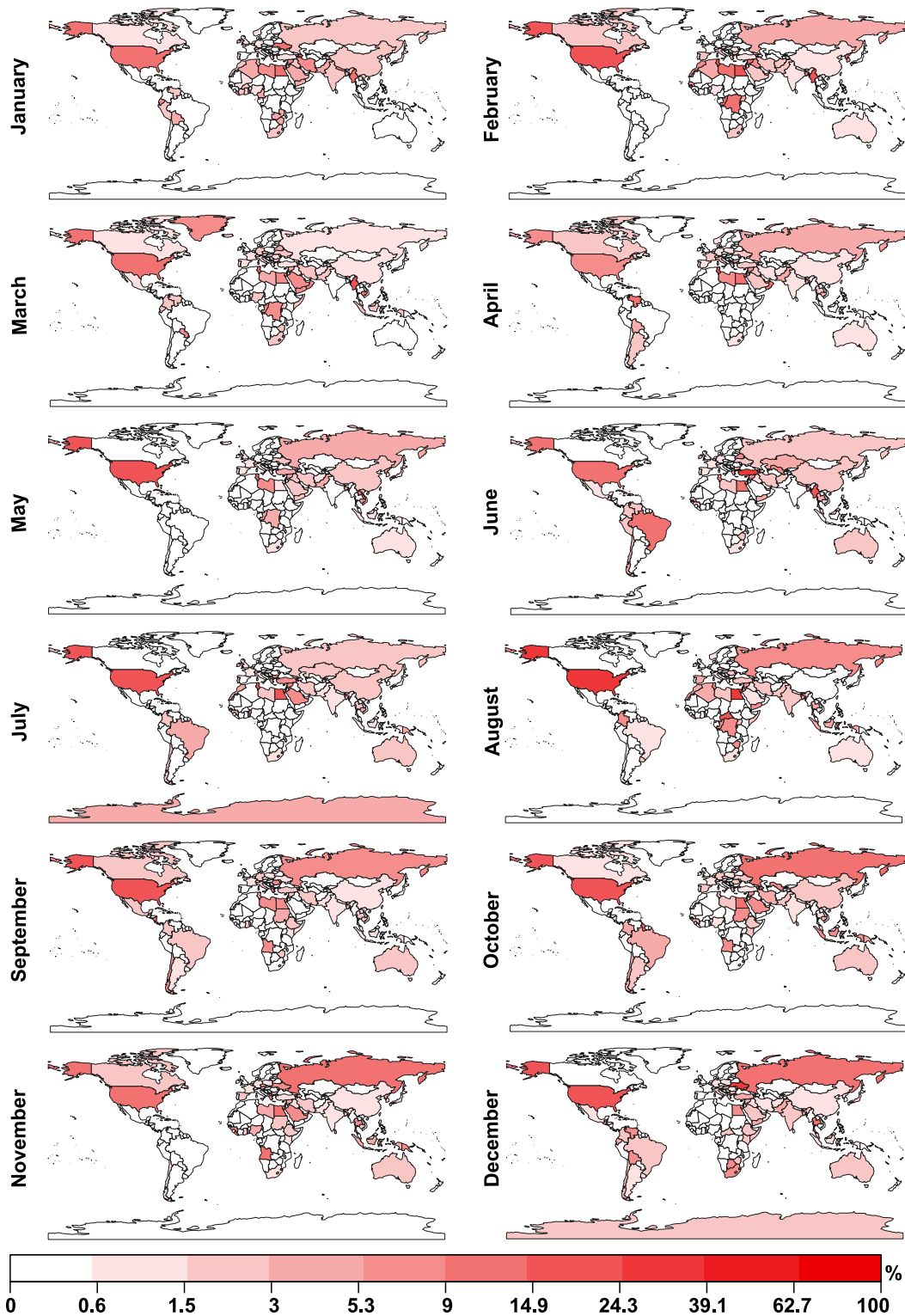


Fig 1. Reports of protests in 2013 in the online edition of *The Guardian*. We use data on reports of protests in the online edition of *The Guardian* as an approximation of the ground truth of when and where notable protest outbreaks occurred. For each of the 244 countries and regions for each month in 2013, we calculate the number of *The Guardian* articles tagged with the country and region's name. Here, we depict the percentage of articles for each country and region and each month which were also tagged with the word "protest". Patterns which can be visually identified in the data reflect known major protest events in 2013: for example, protest outbreaks in both Brazil and Turkey can be observed in June 2013. Equal breaks are calculated for the logarithmically transformed percentages.

doi:10.1371/journal.pone.0150466.g001

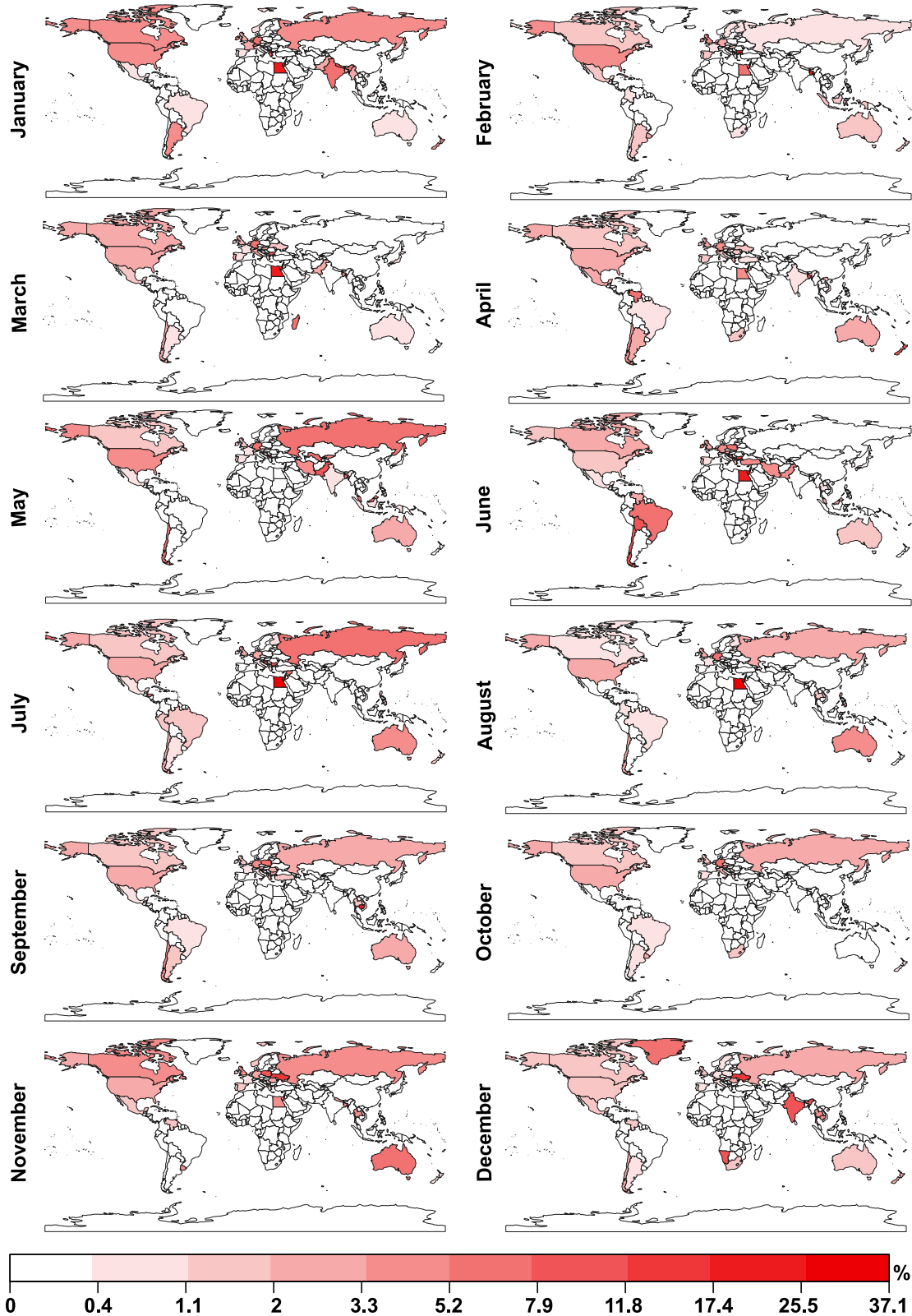


Fig 2. Locations of Flickr photographs labelled with “protest” in 2013. We investigate to what extent data on the number of photographs tagged with the word “protest” and uploaded to Flickr reflect the ground truth data extracted from *The Guardian*. For each of the 244 countries and regions for each month in 2013, we calculate the total number of geotagged photographs taken and uploaded to Flickr. Here, we visualise the percentage of photographs for each

country and region and each month which were also labelled with the character sequence “protest”. Visual inspection suggests that while there are clear differences between the spatio-temporal distributions of “protest” labelled Flickr photographs and “protest” labelled articles in *The Guardian*, some key similarities can also be identified, such as an increase in “protest” labelled Flickr photographs in Brazil and Turkey in June 2013. Equal breaks are calculated for the logarithmically transformed percentages.

doi:10.1371/journal.pone.0150466.g002

For comparison, we construct a simple baseline model which captures differences in protest frequency between countries and regions, and differences in protest frequencies across different weeks, by building a logistic regression panel model with country and region and week as fixed effects, leaving out the Flickr predictor. We find that the model including data on the normalised number of “protest” labelled Flickr photographs allows us to account for more variance in the proportion of *The Guardian* articles tagged with the word “protest” than this simple baseline model of differences between different countries and regions and different weeks (*McFadden R*² for baseline model = 0.34, *McFadden R*² for Flickr model = 0.35, $\chi^2(1) = 84.48$, $p < 0.001$, Likelihood Ratio Test).

In line with other studies of civil unrest, our analysis uses data from newspaper reports of protests as a proxy for ground truth data on protest occurrences [34–36]. As a result, we cannot rule out the possibility that Flickr users are posting photographs labelled with a word signifying “protest” as a result of reading an article about protests in their country and region in *The Guardian*, or another news source. We posit however that the geotagged nature of the Flickr photographs we analyse makes it less likely that such an explanation may hold, in contrast with simple time series analyses of online behaviour on services such as Google or Twitter, where searching behaviour or tweets may reflect reactions to news articles. With this caveat in mind, our results are consistent with the hypothesis that data on photographs posted to Flickr may help us to identify protest outbreaks around the world.

Discussion

We investigate whether data on photographs uploaded to the photo sharing website Flickr may be of use in identifying protest outbreaks. We analyse 25 million photos uploaded to Flickr in 2013 across 244 countries and regions, and determine for each week in each country and region what proportion of the photographs are tagged with the word “protest” in 34 different languages. We find that higher proportions of “protest”-tagged photographs in a given country and region in a given week correspond to greater numbers of reports of protests in that country and region and week in the newspaper *The Guardian*. These results are in line with the striking hypothesis that data on photographs uploaded to Flickr may contain signs of protest outbreaks. Our findings underline the potential value of photographs uploaded to the Internet as a source of global, cheap and rapidly available measurements of human behaviour in the real world.

Supporting Information

S1 Table. Country and region names. List of country and region names used in analysis. (PDF)

S2 Table. Translations of “protest”. List of translations of the word “protest” in different languages. (PDF)

S3 Table. Flickr photograph counts per country and region in 2013. Total number of photographs per country and region taken and uploaded to Flickr during 2013. (PDF)

S4 Table. Counts of articles in The Guardian per country and region in 2013. Total number of *The Guardian* articles released during 2013 covering news relating to countries and regions listed.

(PDF)

Acknowledgments

M.A., T.P. and H.S.M. designed research, performed research, analysed data and wrote the paper.

Author Contributions

Conceived and designed the experiments: MA TP HSM. Performed the experiments: MA TP HSM. Analyzed the data: MA TP HSM. Contributed reagents/materials/analysis tools: MA TP HSM. Wrote the paper: MA TP HSM.

References

1. Lazer D, Pentland AS, Adamic L, Aral S, Barabasi AL, Brewer D, et al. Computational Social Science. *Science*. 2009; 323:721–723. doi: [10.1126/science.1167742](https://doi.org/10.1126/science.1167742) PMID: [19197046](https://pubmed.ncbi.nlm.nih.gov/19197046/)
2. King G. Ensuring the Data-Rich Future of the Social Sciences. *Science*. 2011; 331:719–721. doi: [10.1126/science.1197872](https://doi.org/10.1126/science.1197872) PMID: [21311013](https://pubmed.ncbi.nlm.nih.gov/21311013/)
3. Moat HS, Preis T, Olivola CY, Liu C, Chater N. Using Big Data to Predict Collective Behavior in the Real World. *Behavioral and Brain Sciences*. 2014; 37:92–93. doi: [10.1017/S0140525X13001817](https://doi.org/10.1017/S0140525X13001817) PMID: [24572233](https://pubmed.ncbi.nlm.nih.gov/24572233/)
4. Jiang ZQ, Xie WJ, Li MX, Podobnik B, Zhou WX, Stanley HE. Calling Patterns in Human Communication Dynamics. *Proceedings of the National Academy of Sciences*. 2013; 110:1600–1605. doi: [10.1073/pnas.1220433110](https://doi.org/10.1073/pnas.1220433110)
5. Petersen AM, Tenenbaum JN, Havlin S, Stanley HE, Perc M. Languages Cool as They Expand: Allometric Scaling and the Decreasing Need for New Words. *Scientific Reports*. 2012; 2:943. doi: [10.1038/srep00943](https://doi.org/10.1038/srep00943) PMID: [23230508](https://pubmed.ncbi.nlm.nih.gov/23230508/)
6. Johnson N, Carran S, Botner J, Fontaine K, Laxague N, Nuetzel P, et al. Patterns in Escalations in Insurgent and Terrorist Activity. *Science*. 2011; 333:81–88. doi: [10.1126/science.1205068](https://doi.org/10.1126/science.1205068) PMID: [21719677](https://pubmed.ncbi.nlm.nih.gov/21719677/)
7. Letchford A, Moat HS, Preis T. The advantage of short paper titles. *Royal Society Open Science*. 2015; 2(8):150266. doi: [10.1098/rsos.150266](https://doi.org/10.1098/rsos.150266) PMID: [26361556](https://pubmed.ncbi.nlm.nih.gov/26361556/)
8. Preis T, Moat HS, Stanley HE. Quantifying Trading Behavior in Financial Markets Using Google Trends. *Scientific Reports*. 2013; 3:1684. doi: [10.1038/srep01684](https://doi.org/10.1038/srep01684) PMID: [23619126](https://pubmed.ncbi.nlm.nih.gov/23619126/)
9. Preis T, Reith D, Stanley HE. Complex dynamics of our economic life on different scales: insights from search engine query data. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. 2010; 368(1933):5707–5719. doi: [10.1098/rsta.2010.0284](https://doi.org/10.1098/rsta.2010.0284)
10. Choi H, Varian H. Predicting the Present with Google Trends. *Economic Record*. 2012; 88:2–9. doi: [10.1111/j.1475-4932.2012.00809.x](https://doi.org/10.1111/j.1475-4932.2012.00809.x)
11. Noguchi T, Stewart N, Olivola CY, Moat HS, Preis T. Characterizing the time-perspective of nations with search engine query data. *PLoS ONE*. 2014; 9(4):e95209. doi: [10.1371/journal.pone.0095209](https://doi.org/10.1371/journal.pone.0095209) PMID: [24736725](https://pubmed.ncbi.nlm.nih.gov/24736725/)
12. Kristoufek L. Can Google Trends Search Queries Contribute To Risk Diversification? *Scientific Reports*. 2013; 3:2713. doi: [10.1038/srep02713](https://doi.org/10.1038/srep02713) PMID: [24048448](https://pubmed.ncbi.nlm.nih.gov/24048448/)
13. Curme C, Preis T, Stanley HE, Moat HS. Quantifying the semantics of search behavior before stock market moves. *Proceedings of the National Academy of Sciences*. 2014; 111(32):11600–11605. doi: [10.1073/pnas.1324054111](https://doi.org/10.1073/pnas.1324054111)
14. Bordino I, Battiston S, Caldarelli G, Cristelli M, Ukkonen A, Weber I. Web search queries can predict stock market volumes. *PLoS ONE*. 2012; 7(7):e40014. doi: [10.1371/journal.pone.0040014](https://doi.org/10.1371/journal.pone.0040014) PMID: [22829871](https://pubmed.ncbi.nlm.nih.gov/22829871/)
15. Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ. Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences*. 2010; 107(41):17486–17490. doi: [10.1073/pnas.1005962107](https://doi.org/10.1073/pnas.1005962107)

16. Yasseri T, Sumi R, Rung A, Kornai A, Kertész J. Dynamics of conflicts in Wikipedia. PLOS ONE. 2012; 7(6):e38869. doi: [10.1371/journal.pone.0038869](https://doi.org/10.1371/journal.pone.0038869) PMID: [22745683](https://pubmed.ncbi.nlm.nih.gov/22745683/)
17. Moat HS, Curme C, Avakian A, Kenett DY, Stanley HE, Preis T. Quantifying Wikipedia Usage Patterns Before Stock Market Moves. Scientific Reports. 2013; 3:1801.
18. Alanyali M, Moat HS, Preis T. Quantifying the Relationship Between Financial News and the Stock Market. Scientific Reports. 2013; 3:3578. doi: [10.1038/srep03578](https://doi.org/10.1038/srep03578) PMID: [24356666](https://pubmed.ncbi.nlm.nih.gov/24356666/)
19. Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. Journal of Computational Science. 2011; 2(1):1–8. doi: [10.1016/j.jocs.2010.12.007](https://doi.org/10.1016/j.jocs.2010.12.007)
20. Gonçalves B, Perra N, Vespignani A. Modeling users' activity on twitter networks: Validation of dunbar's number. PLOS ONE. 2011; 6(8):e22656. doi: [10.1371/journal.pone.0022656](https://doi.org/10.1371/journal.pone.0022656) PMID: [21826200](https://pubmed.ncbi.nlm.nih.gov/21826200/)
21. Botta F, Moat HS, Preis T. Quantifying crowd size with mobile phone and Twitter data. Royal Society open science. 2015; 2(5):150162. doi: [10.1098/rsos.150162](https://doi.org/10.1098/rsos.150162) PMID: [26064667](https://pubmed.ncbi.nlm.nih.gov/26064667/)
22. Steinert-Threlkeld ZC, Mocanu D, Vespignani A, Fowler J. Online social networks and offline protest. EPJ Data Science. 2015; 4(1):1–9. doi: [10.1140/epjds/s13688-015-0056-y](https://doi.org/10.1140/epjds/s13688-015-0056-y)
23. Preis T, Moat HS, Bishop SR, Treleaven P, Stanley HE. Quantifying the digital traces of Hurricane Sandy on Flickr. Scientific reports. 2013; 3(3141).
24. Wood SA, Guerry AD, Silver JM, Lacayo M. Using social media to quantify nature-based tourism and recreation. Scientific reports. 2013; 3(2976).
25. Barchiesi D, Moat HS, Alis C, Bishop S, Preis T. Quantifying international travel flows using Flickr. PLOS ONE. 2015; 10(7):e0128470. doi: [10.1371/journal.pone.0128470](https://doi.org/10.1371/journal.pone.0128470) PMID: [26147500](https://pubmed.ncbi.nlm.nih.gov/26147500/)
26. Barchiesi D, Preis T, Bishop S, Moat HS. Modelling human mobility patterns using photographic data shared online. Royal Society open science. 2015; 2(8):150046. doi: [10.1098/rsos.150046](https://doi.org/10.1098/rsos.150046) PMID: [26361545](https://pubmed.ncbi.nlm.nih.gov/26361545/)
27. Askitas N, Zimmermann KF. Google econometrics and unemployment forecasting. German Council for Social and Economic Data (RatSWD) Research Notes. 2009;(41:).
28. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. 2009; 457(7232):1012–1014. doi: [10.1038/nature07634](https://doi.org/10.1038/nature07634) PMID: [19020500](https://pubmed.ncbi.nlm.nih.gov/19020500/)
29. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google Flu: traps in big data analysis. Science. 2014; 343(14 March):1203–1205. doi: [10.1126/science.1248506](https://doi.org/10.1126/science.1248506) PMID: [24626916](https://pubmed.ncbi.nlm.nih.gov/24626916/)
30. Preis T, Moat HS. Adaptive nowcasting of influenza outbreaks using Google searches. Royal Society open science. 2014; 1(2):140095. doi: [10.1098/rsos.140095](https://doi.org/10.1098/rsos.140095) PMID: [26064532](https://pubmed.ncbi.nlm.nih.gov/26064532/)
31. Branigan T. "China Blocks Twitter, Flickr and Hotmail Ahead of Tiananmen Anniversary";. Accessed: 2015-01-29. Available: <http://www.theguardian.com/technology/2009/jun/02/twitter-china>.
32. Christie-Miller A. "Erdogan bans Twitter as corruption claims spread";. 2015-01-29. Available: <http://www.thetimes.co.uk/tto/news/world/europe/article4040322.ece>.
33. Arthur C. "Egypt blocks social media websites in attempted clampdown on unrest";. 2015-01-29. Available: <http://www.theguardian.com/world/2011/jan/26/egypt-blocks-social-media-websites>.
34. Braha D. Global civil unrest: contagion, self-organization, and prediction. PLOS ONE. 2012; 7(10):e48596. doi: [10.1371/journal.pone.0048596](https://doi.org/10.1371/journal.pone.0048596) PMID: [23119067](https://pubmed.ncbi.nlm.nih.gov/23119067/)
35. Compton R, Lee C, Xu J, Artieda-Moncada L, Lu TC, De Silva L, et al. Using publicly visible social media to build detailed forecasts of civil unrest. Security informatics. 2014; 3(1):1–10. doi: [10.1186/s13388-014-0004-6](https://doi.org/10.1186/s13388-014-0004-6)
36. Dos Santos R, Shah S, Chen F, Boedihardjo A, Lu CT, Ramakrishnan N. Forecasting location-based events with spatio-temporal storytelling. In: Proceedings of the 7th ACM SIGSPATIAL International Workshop on Location-Based Social Networks. ACM; 2014. p. 13–22.