Model Info Sheet

**Section 1: Information about paper or report**

1) Author(s): Names of the authors of the paper or report

Jin, Fang et al.

j

2) Title of the paper or report which introduces the model

*Epidemiological modeling of news and rumors on Twitter*

3) DOI or permanent link to the paper or report (for example, link to arxiv.org webpage)

https://doi.org/10.1145/2501025.2501027

4) License: Under which license(s) are the data and/or model shared?

5) Email address of the corresponding author

jfang8@vt.edu

**Section 2: Scientific claim(s) of interest**

6) Does your paper make a generalizable claim based on the ML model? If yes, what is the scientific claim? For example, “Our ML model can be used to diagnose Covid-19 using chest radiographs of adult patients”.

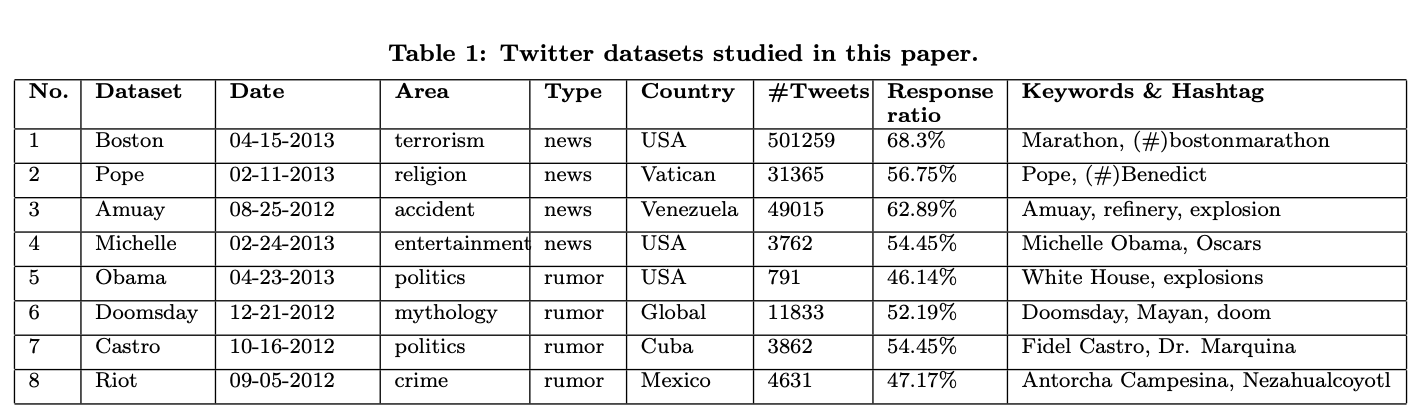
*‘While not an exhaustive list [of topics and geographic regions], [our analysis] demonstrates the wide applicability of the proposed model.’*

*‘We demonstrate the capability of the* ***SEIZ model*** *to quantify compartment transition dynamics. We showcase how such information could facilitate the development of screening criteria for distinguishing rumors from real news happenings on Twitter.’*

*‘SEIZ model fits the Twitter data much more accurately than the SIS model.’*

7) Is the scientific claim made about a distribution or population from which you can sample? If yes: (a) what is the population or distribution about which the scientific claim is being made? (b) What is the sample used for the study? For example, “(a) Population: adult patients with symptoms of Covid-19. (b) Sample: We use a random sample of adult patients who present at a U.S. based hospital between April 2020 and June 2020”.

*Yes – test set is a sample of tweets from across geographical regions and topics, in multiple languages. Tweets were surfaced with a series of keywords pertinent to each topic (see table below):*

**

8) Does the scientific claim only apply to certain subsets of the distribution mentioned in Q6? For example, “Our model works on chest radiographs of U.S.-based adult patients and might not generalize to radiographs taken in other places or using different machines.”

*Sampled tweet stories are specific to geography and language.*

**Section 3: Train-test split is maintained across all steps in creating the model**

9) Train-test split type: How was the dataset split into train and test sets? (For example, cross-validation; separate train and test sets).

*This split wasn’t actually observed for the model identification step of this study, which didn’t require the training on an ML model (only the fitting of a statistical model – one of SEIZ or SIS).*

10) Are there duplicates in the dataset? If yes, explain how duplicates are handled to ensure the train-test split.

*Seems possible that there might be duplicate tweets in the dataset, though it’s unclear if the authors flag 1) multiple tweets from the same account, or 2) retweets of the same tweet from different (or same) accounts.*

11) In case the dataset has dependencies (e.g., multiple rows of data from the same patient), describe how the dependencies were addressed (for example, using block-cross validation).

*Again, a bit unclear. Maybe more productive to describe certain model assumptions in this space, instead?:*

* *transition rates: in both SIS and SEIZ model, unclear how quickly users are transitioning from ‘infected’ to ‘susceptible’ compartments, and between any of the four compartments in the latter model*
* *unclear what the total population size should be – it could be the total number of Twitter accounts in the test sample, but this is probably just a loose upper bound (we’re really interested in the number of individuals* ***who could be exposed to the news or rumor topic****)*
* *it seems like not knowing 1) the total population size N and 2) the initial population sizes for each compartment would be a huge disadvantage to model fit ???*

12) List all the pre-processing steps used in creating your model. For example, imputing missing data, normalizing feature values, selecting a subset of rows from the dataset for building the model.

*The authors fit compartment size during the model fitting process? i.e., they accept S(t\_0), E(t\_0) etc as parameters in the parameter fit routine.*

13) How was the train-test split observed during each pre-processing step? If applicable, use a separate line for each step mentioned in Q12.

*Again, no split observed; seems like the authors performed a non-linear least-squares fit of SEIZ and SIS over the entire dataset. They note that exposure delay seems to yield a better fit of SEIZ than SIS for the eight stories under investigation.*

14) List all the modeling steps used in creating your model. For example, feature selection, parameter tuning, model selection.

*SIS and SEIZ are already established models for modeling birth-death events; most of the methodology of this paper describes adapting analogs of the original setting (literal birth / life of biological organisms) to this setting (life and death of a misinformative narrative) and fitting parameters on real data.*

15) How was the train-test split observed during each modeling step? If applicable, use a separate line for each step mentioned in Q14.

*Not so much test-train as just model-fitting (the model in question isn’t an ML model – it’s a birth-death graphical model) with constraints.*

16) List all the evaluation steps used in evaluating model performance. For example, cross-validation, out-of-sample testing.

*The authors identify a ratio R\_{SI} as follows:*

*R\_{SI} = the sum of the effective transition rates* ***entering*** *the exposed compartment / the sum of the transition rates* ***exiting*** *this compartment to I.*

*R\_{SI} > 1 means that influx > efflux, and R\_{SI} < 1 means the opposite.*

17) How was the train-test split observed during each evaluation step? If applicable, use a separate line for each step mentioned in Q16.

*Maybe not a relevant q for this paper*

**Section 4: Test set is drawn from the distribution of scientific interest.**

18) Why is your test set representative of the population or distribution about which you are making your scientific claims?

*The authors use a set of keywords to surface tweets related to certain misinfo narratives – seems like this might introduce a certain degree of bias, as far as proportion of true / misinformative tweets, and propagation rates (if they’re surfacing viral tweets, for instance, maybe transition rates would be inflated during model fitting?).*

19) Explain the process for selecting the test set and why this does not introduce selection bias in the learning process.

*Surfaced tweets from keyword set.*

20) In case your model is used to predict a future outcome of interest using past data, detail how data in the training set is always from a date earlier than the data in the test set.

*Maybe not a relevant q for this paper*

*The authors do claim that the R\_{SI} scores for ‘true’ news stories do have much higher scores than did the rumor topics. But I wonder if it might be more worthwhile to compare R\_{SI} scores to true / untrue narratives about the same topic, i.e. stories that are merely discussing the Boston bombings vs stories that are actively spreading misinfo about the Boston bombings.*

**Section 5:** **Each feature used in the model is legitimate for the task**

21) List the features used in the model, alongside an argument for their legitimacy. A legitimate feature is one that would be available when the model is used in the real world and is not a proxy of the outcome being predicted. You can also include this list in an appendix and reference the relevant section of your Appendix here.

*N (sample set size); S(t\_0); E(t\_0); I(t\_0); Z(t\_0)*

*probability of belief p;*

*probability that a susceptible person becomes a skeptic b;*

*probability that a susceptible person becomes infected (1 - l )*