**Supplementary Information for:**

Understanding Public Perception of COVID-19 Social Distancing on Twitter

**Authors:**

Sameh N. Saleh, MD1,2 [ORCID: 0000-0001-5959-1659]

Christoph U. Lehmann, MD2 [ORCID: 0000-0001-9559-4646]

Samuel A. McDonald, MD2,3

Mujeeb A. Basit, MD1,2 [ORCID: 0000-0002-4948-6158]

Richard J. Medford, MD1,2 [ORCID: 0000-0001-9814-8043]

**Corresponding Author:**

Sameh N. Saleh

M: 571-338-3680

F: 214-648-9478

[sameh.n.saleh@gmail.com](mailto:sameh.n.saleh@gmail.com)

**Section A.**

Methods: data processing, transformation, and exploration.

We analyzed tweet metadata to provide descriptive characteristics. For sentiment, emotion, and topic modeling analyses, we transformed tweets into plain text and removed hyperlinks and user mentions. For topic modeling, we further converted the tweet plain text to lowercase, changed words to their root forms (e.g., ‘viruses’ to ‘virus’ or ‘went’ to ‘go’) using the WordNetLemmatizer module of the *NLTK* library in Python(1), and removed stop words (frequently used words with little semantic meaning, such as ‘of’, ‘it’, and ‘is’). For additional dimensionality reduction, we created a list of n-grams (one-word and two-word terms or unigrams and bigrams, respectively) from tweets and removed all extremely low and high frequency terms. These steps decreased the dictionary of terms from 469,039 to 9,984. Using a word cloud, we visualized the top 200 words for each hashtag (after excluding each hashtag and its subdivisions) with larger font size representing greater frequency.

**Table S1.** Metadata associated with tweets

|  |  |  |
| --- | --- | --- |
| user\_id | lang | retweet\_description |
| status\_id | quoted\_status\_id | place\_url |
| created\_at | quoted\_text | place\_name |
| screen\_name  text | quoted\_created\_at | place\_full\_name |
| source | quoted\_source | place\_type |
| display\_text\_width | quoted\_favorite\_count | country |
| reply\_to\_status\_id | quoted\_retweet\_count | country\_code |
| reply\_to\_user\_id | quoted\_user\_id | geo\_coords |
| reply\_to\_screen\_name | quoted\_screen\_name | coords\_coords |
| is\_quote | quoted\_name | bbox\_coords |
| is\_retweet | quoted\_followers\_count | status\_url |
| favorite\_count | quoted\_friends\_count | name |
| retweet\_count | quoted\_statuses\_count | location |
| quote\_count | quoted\_location | description |
| reply\_count | quoted\_description | url |
| hashtags | quoted\_verified | protected |
| symbols | retweet\_status\_id | followers\_count |
| urls\_url | retweet\_text | friends\_count |
| urls\_t.co | retweet\_created\_at | listed\_count |
| urls\_expanded\_url | retweet\_source | statuses\_count |
| media\_url | retweet\_favorite\_count | favourites\_count |
| media\_t.co | retweet\_retweet\_count | account\_created\_at |
| media\_expanded\_url | retweet\_user\_id | verified |
| media\_type | retweet\_screen\_name | profile\_url |
| ext\_media\_url | retweet\_name | profile\_expanded\_url |
| ext\_media\_t.co | retweet\_followers\_count | account\_lang |
| ext\_media\_expanded\_url | retweet\_friends\_count | profile\_banner\_url |
| ext\_media\_type | retweet\_statuses\_count | profile\_image\_url |
| mentions\_user\_id | retweet\_location | profile\_background\_url |
| mentions\_screen\_name | retweet\_verified |  |

**Table S2.** Keywords about social distancing rules

|  |  |  |  |
| --- | --- | --- | --- |
| **Distance** | feet | foot | meter |
| meters | metre | metres |
| **Crowds** | gathering | crowd | masses |
| groups | large groups |  |

**Figure S1.** Probability of user being a bot based on the Botometer tool (2). A random subset of 3,900 users were evaluated.

**References**

1. Loper E, Bird S. NLTK: the Natural Language Toolkit. In: Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics - [Internet]. Philadelphia, Pennsylvania: Association for Computational Linguistics; 2002 [cited 2020 Jul 6]. p. 63–70. Available from: http://portal.acm.org/citation.cfm?doid=1118108.1118117

2. Davis CA, Varol O, Ferrara E, Flammini A, Menczer F. BotOrNot: A System to Evaluate Social Bots. In: Proceedings of the 25th International Conference Companion on World Wide Web - WWW ’16 Companion [Internet]. Montr&#233;al, Qu&#233;bec, Canada: ACM Press; 2016 [cited 2020 Jul 18]. p. 273–4. Available from: http://dl.acm.org/citation.cfm?doid=2872518.2889302