Assessing disinformation through the dynamics of supply and demand in the news ecosystem

Supplementary Information

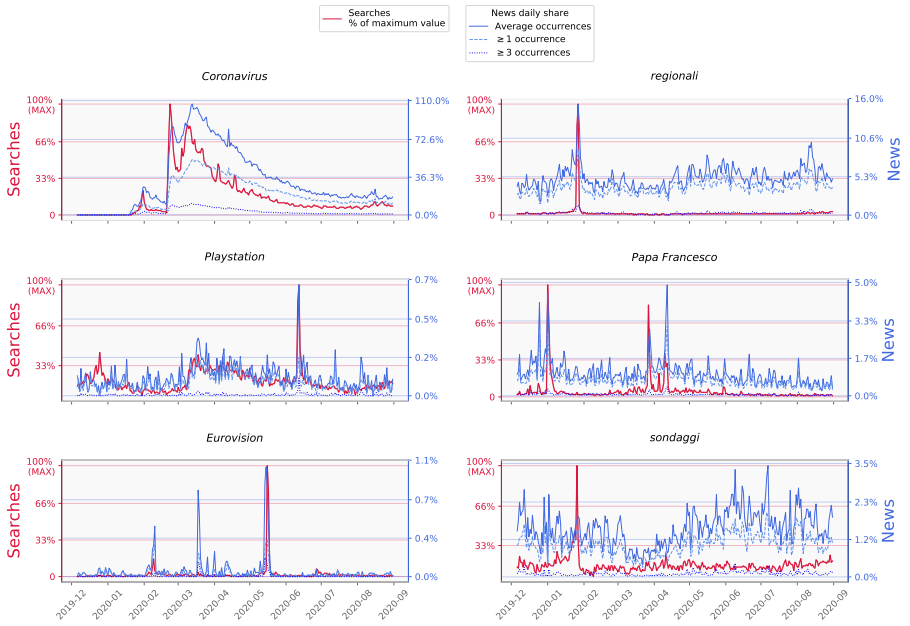

Figure S1: Temporal behaviour of the fractions of Searches (red, left y-axis) and General News (blue, right y-axis), in Italy from early December 2019 to the end of August 2020, about six selected keywords. Searches are reported as a percentage of the maximum observed in the monitored period. General News are represented instead by three curves showing: the daily fraction of articles containing at least one occurrence of the keyword (dotted line), the daily fraction of articles containing at least three occurrences of the keyword (dashed line), and the average number of occurrences per article (continuous line). The six keywords are *coronavirus*, *playstation*, *eurovision*, *regionali* (regional elections), *papa francesco* (Pope Francesco) and *sondaggi* (polls). *coronavirus* was dominant in the information landscape of the observed period, resulting to be the most queried term and and the most frequent keyword in the news. The peak in mid-march, in particular, shows that, on average, every newspaper article mentioned it once.

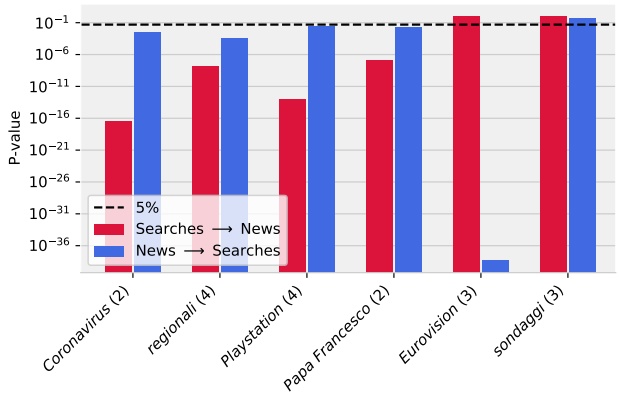

Figure S2: The results of the Granger-causality test for the six keywords investigated. We report the p-values and, next to each keyword, the time windows' size for the best model in brackets. The horizontal black line highlights the 5% significance value. When the bars go above such a threshold, the corresponding test for Granger-causality is not significant. Lower bars suggest a stronger causality. The majority of the most queried keywords shows a Search Granger-caused dynamics.

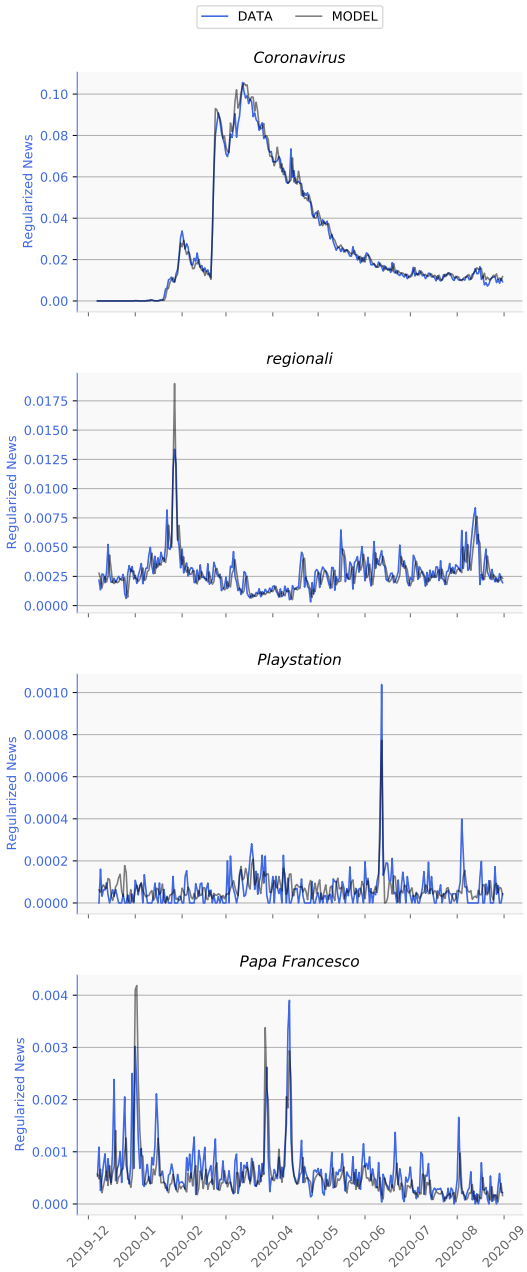

Figure S3: Comparison of the time series for General News and the outcome of the Improved Model from equation 1. Numeric results are reported in Tab. 1

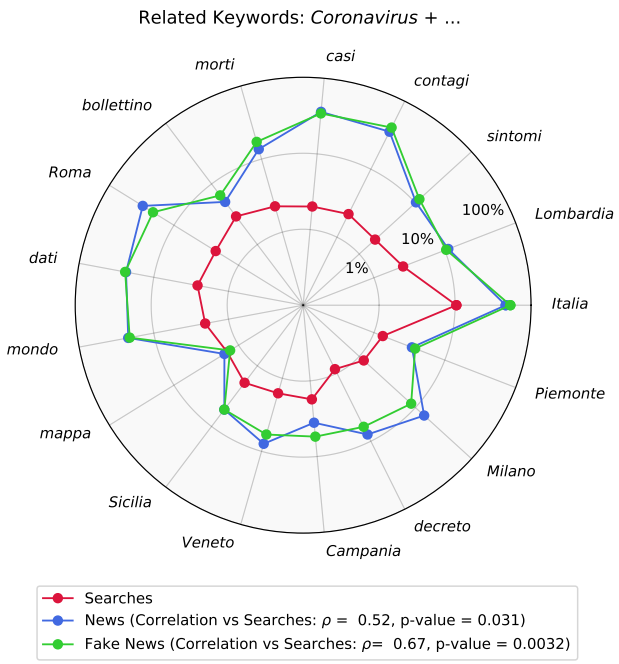

Figure S4: Radar representation of the total semantic vectors for Searches, General News, and Fake News. The radius is the logarithm of the average percentage of the related keyword. The results from Spearman's correlations are described in the legend. Searches are normalized to 1, while General News and Fake News sum to more than 1 given the metrics' non-exclusivity (a piece of news can be counted in more sub-domains if it contains more keywords). For the analysis performed, however, the scales are irrelevant.

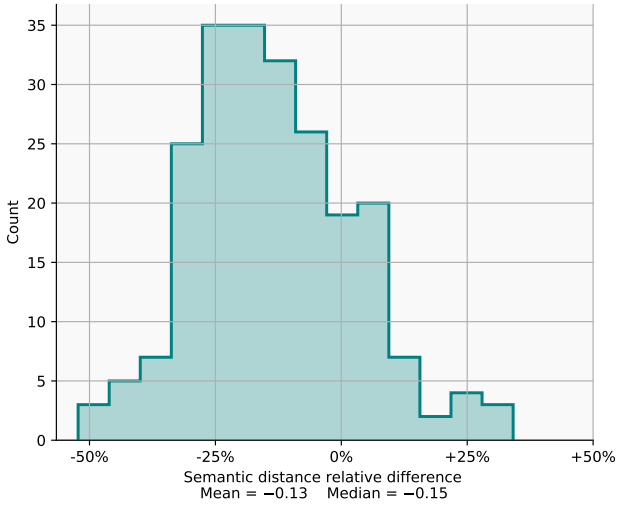

Figure S5: The histogram of the daily relative differences between the cosine distance Searches-Fake News and the cosine distance Searches-General News. Both average and median are below zero, indicating that Fake News are closer to Searches than General News even on daily basis.