LDA2Net: Digging under the surface of COVID-19 topics in literature

Topic 112 companion sheet

G. Minello C.R.M.A. Santagiustina M. Warglien

This file contains the following supplementary information for Topic 112 of the manuscript "LDA2Net: Digging under the surface of COVID-19 topics in scientific literature":

- Human label and automatic n-gram label proposals (Table 1)
- Summary measures (Table 2)
- Network of top 25 bigrams (Figure 1)
- Wordclouds of top 25 words by node relevance measure (Figure 2)
- Wordclouds of top 25 bigrams by edge relevance measure (Figure 3)
- Filtered (0.99 percentile) topic network (Figure 4)

Table 1: Human and automatic label proposals. Automatic label candidate for largest word community of the topic. In parenthesis: absolute frequency of the walk out of a sample of size 1000.

Human label	2-gram label	3-gram label	4-gram label
time studies	total->number (9.5%)	significant->increase->number (4.3%)	ED->emergency->visits->total (1.8%)

Here follows the set of topic-specific measures that have been used to classify the topic and to analyse its structural properties (see manuscript for details):

Table 2: Summary measures

JSD	Mean propensity	Variance propensity	Modularity	Barrat Clustering Coeff.
0.698072	0.008700	0.000361	0.319341	0.596008
76	83	71	116	85

Based on the aforementioned measures, Topic 112 has been classified as a CROSS-CUTTING topic.

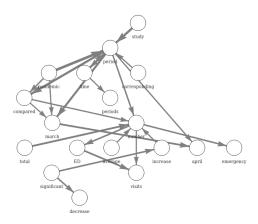
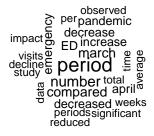
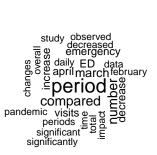


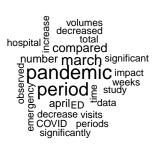
Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

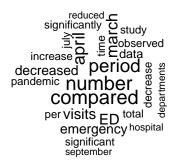












Out-degree Betweenness PageRank

Figure 2: Top 25 unigrams (i.e., nodes) by measure.

patients-admitted morel-coronavirus COVID-19-acception mental statistically-significant health-care statistically-significant health-care covid-19 goronavirus-disease COVID-19-pandemic public-health united-states COVID-19-pandemic statistically-significant health-care severe-active number-patients sudy-period statistically-significant health-care severe-active number-patients patients-COVID-19-goronavirus-production patients-COVID-19-goronavirus-patients-covid-19-goronavirus-pandemic-period

number-compared period-pandemic period-bandemic period-bandemic period-compared period-compared period-compared period-compared period-compared period-compared period-bandemic period-compared period-bandemic period-bandemic period-bandemic period-bandemic period-perio

corresponding-period to 8 compared-number to 10 compared-period march 20 compared-period march 20 compared-period total-number number-yells march-20 compared-period total-number number-yells march-20 compared-period total-number total-number number-yells number number-yells period-number 10 compared-period total-number total-number number-yells number number-yells period-number 10 compared-period total-number 10 compared-period total-number number-yells number number-yells number-yells number-yells number-yells number-yells number-yells number-yells nu

Figure 3: Top 25 bigrams (i.e., edges) by measure.

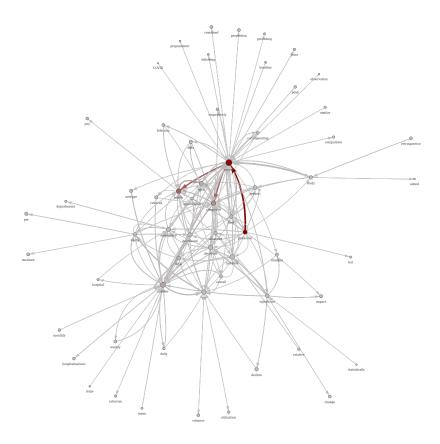


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness centrality. Isolated nodes have been removed after filtration.