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

Topic 77 companion sheet

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

This file contains the following supplementary information for Topic 77 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 & duration	median->time (28.5%)	median->time->period (5.3%)	median->time->period->day (2.5%)

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.
value	0.800624	0.008101	0.000196	0.000000	0.560919
rank	104	53	31	12	39

Based on the aforementioned measures, Topic 77 has been classified as a CROSS-CUTTING topic.

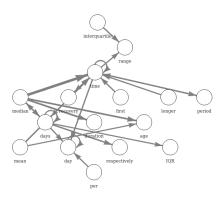


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.



total shorter range incubation interval duration interval median set of the following age days period IQR mean first study recovery average interquartile series



interval Median Bare Lifectovery Bare Lifectovery Bare hours periodprolonged incubation

 average included respectively upon interval control of the point of th

Out-degree Betweenness PageRank

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

social-distancing
cohort-study COVID-19-cases
mental-health median-age
mechanical-ventilation
syndrome-coronavirus
risk-factors severe>accute
care-unit acute-respiratory
coronavirus>disease
COVID-19>patients
patients>COVID-19 patients
patients>COVID-19 patients
patients>COVID-19 novel>coronavirus

COVID-19-infection
severe>COVID-19 novel>coronavirus

patients>admitted

days-sine
longer-time days-day
time-sine first-time
period-time means-age
time-period-time means-age
time-period-time means-age
median-butanon
days-median time-day
days-median time-day
days-median time-day
days-median-day-day-day
interquartile-range days-day
time-day-day-day
interquartile-range days-days

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

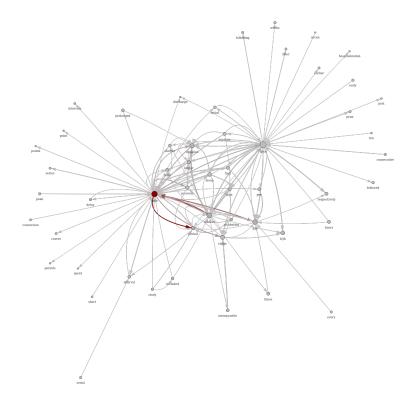


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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