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

Topic 94 companion sheet

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

This file contains the following supplementary information for Topic 94 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
sampling viral rna	viral->RNA (21.2%)	viral->RNA->load (7%)	viral->RNA->loads->samples (4%)

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):

 ${\bf Table\ 2:\ Summary\ measures}$

	JSD	Mean propensity	Variance propensity	Modularity	Barrat Clustering Coeff.
value	0.617761	0.008290	0.000555	0.000000	0.570664
rank	47	64	97	38	52

Based on the aforementioned measures, Topic 94 has been classified as a SPECIALIZED topic.

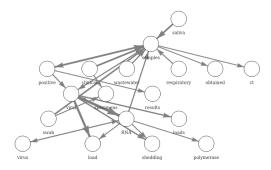
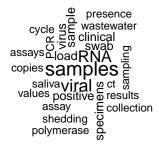


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.





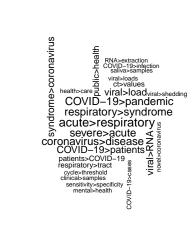
extraction obtained by copies its values in copies its value in copies its values in cop



values shedding extraction positive test virus RNA extraction positive test loads specimens VIRA compared to the sample sample sample sample sample sample sample sample sample culture transport PCR of colline transport proculture to the sample sam

Out-degree Betweenness PageRank

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



samples>specimens
samples>clinical samples>positive
swab>samples\rial samples>positive
swab>samples\rial>viral load>viral
samples>samples RNA>samples
clinical>samples>rial>load\rial>samples
samples>samples
samples>viral>samples
samples>RNA\rial>samples
samples>RNA\rial>samples
samples>RNA\rial>samples
samples>RNA\rial>samples>virus
samples>RNA\rial>samples>virus
samples>RNA\rial>samples>virus
samples>RNA\rial>samples>samples
positive>samples\rial\rial>samplessamples
load>samples
specimens>samples



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

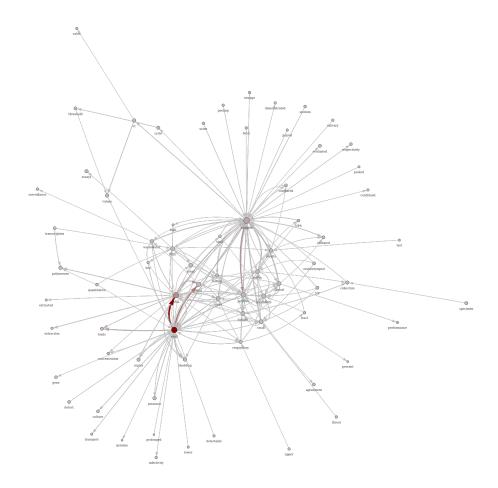


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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