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

Topic 25 companion sheet

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

This file contains the following supplementary information for Topic 25 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
comparatiove analysis of viremic levels	significantly->higher (16.6%)	levels->significantly->higher (6.7%)	significantly->higher->lower->level (3.7%)

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	0.803931	0.007717	0.000147	0.175756	0.606167
rank	106	31	19	100	97

Based on the aforementioned measures, Topic 25 has been classified as a CROSS-CUTTING topic.

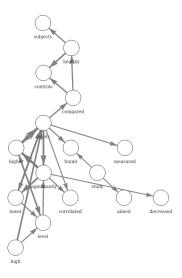


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.







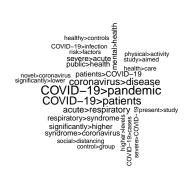


correlation
decreased whether
high significantly
investigated controlshigher
level levels normal
level levels normal
level levels purpose healthy lower
subjects Freduced
values O
measured I
observed
frequency



Out-degree Betweenness PageRank

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



levels-significantly levels-healthy levels-healthy or compared levels glevels-significantly levels-significantly levels-significantly levels-significantly levels-significantly levels-significantly levels-levels levels-level levels-levels levels-levels levels-levels levels-subjects-levels healthy-levels subjects-levels healthy-levels significantly-level levels-levels levels-leve

levels-correlated levels-correlated levels-injeff levels-measured levels-injeff levels-measured levels-injeff levels-injeff levels-injeff levels-injeff levels-significational levels-shigher compared-controls levels-significativity a higher levels-significativity levels-significativity levels-significativity levels-injeff levels-significativity-levels levels-injeff levels-significativity-levels levels-injeff levels-inje

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

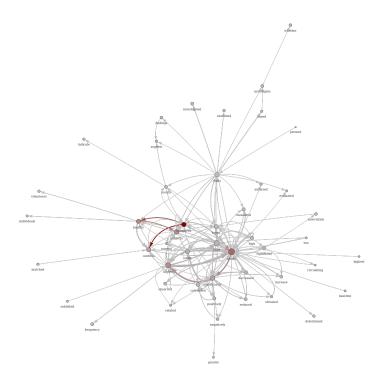


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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