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

Topic 56 companion sheet

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

This file contains the following supplementary information for Topic 56 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
nursing programs residents	programs->program (32.3%)	programs->program->value (8.2%)	programs->program->value->virtual (4.9%)

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	0.577114 27	0.007484	0.000308	0.138525	0.534736
rank		18	59	91	18

Based on the aforementioned measures, Topic 56 has been classified as a SPECIALIZED topic.

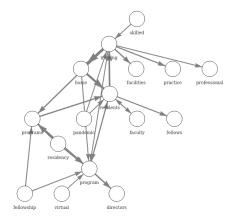


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.



clinical members resident directors facilities program

pandemic nursing program

pandemic nursing program second programs fellows residency including pacademic community members residency programs fellows future academic programs fellows future including pacademic programs fellows future including programs fellows residency programs fellows future including programs fellows fu

security of the complete of th



residency process
application members
application members
applicants
by applicants
clinical system to provided to home by applicants
clinical system to provided to home by applications
continue residents fellows
programs faculty virtual facilities faculty virtual directors completed

completed pandemic process trainees fellows of programs virtual control of the practice of the process trainees of trainees of

Out-degree Betweenness PageRank

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

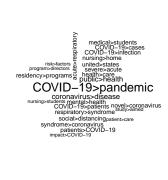






Figure 3: Top 25 bigrams (i.e., edges) by measure.

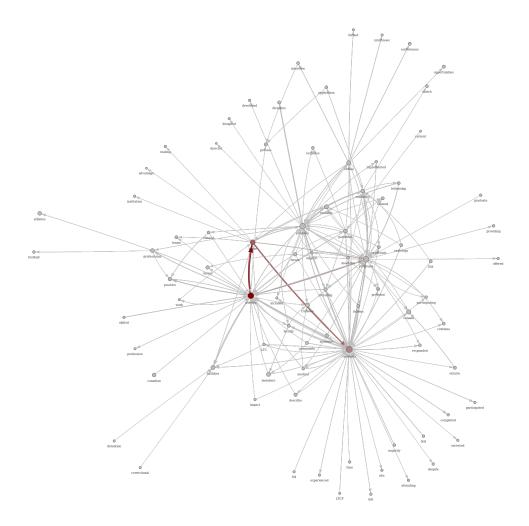


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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