

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

Topic 7 companion sheet

G. Minello

C.R.M.A. Santagiustina

M. Warglien

This file contains the following supplementary information for Topic 7 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.

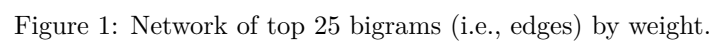
Human label	2-gram label	3-gram label	4-gram label
literature review	various->aspects (10.9%)	several->specific->aspects (1.6%)	aspects->human->papillomavirus->related (0.5%)

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	0.729031	0.008548	0.000164	0.018322	0.564478
rank	91	78	21	63	46

Based on the aforementioned measures, Topic 7 has been classified as a CROSS-CUTTING topic.



A word cloud of terms related to the review. The most prominent words are 'current' and 'review', which are the largest and most central. Other significant words include 'comprehensive', 'understanding', 'management', 'focus', 'present', 'including', 'provides', 'future', 'currently', 'knowledge', 'discuss', 'various', 'article', 'literature', 'regarding', 'aspects', 'overview', 'implications', 'summarize', 'highlight', 'specific', 'recent', 'several', and 'focus'. The words are arranged in a circular pattern around the central 'current review' text.

reviewed
summarizes
including
discuss
article
scientific
existing
finally
review
current
present
various
comprehensive
brief
knowledge
several
regarding
management
overview
aims
narrative
provides
summarize
recent

review
literature
current
knowledge
discuss
including
management
summarize
studies
particular
comprehensive
studies
overview
understanding
focus
given
present
article
evidence
specific
aspects
aims
regarding
provides
recent
several

Out-degree

Betweenness

PageRank

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



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

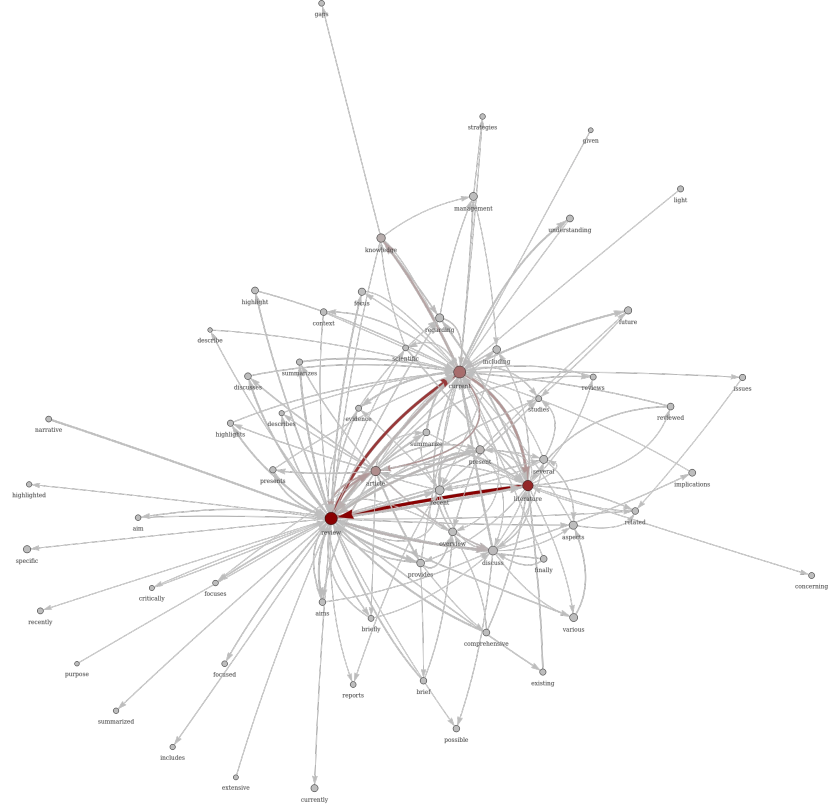


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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