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

Topic 84 companion sheet

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

This file contains the following supplementary information for Topic 84 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
questionnaire studies	study->participants (18.7%)	study->participants->conducted (5.8%)	study->participants->conducted->aimed (2.2%)

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	0.661441	0.009120	0.000425	0.380720	0.580382
rank	63	98	81	119	66

Based on the aforementioned measures, Topic 84 has been classified as a SPECIALIZED topic.

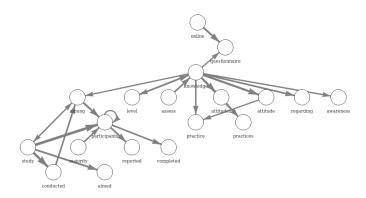


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.



questions completed
by total attitudes
completed
attitudes
completed
attitudes
completed
attitudes
completed
attitudes
attitudes
completed
attitudes
attitudes
aimed jevel
level
perception positive
and attitudes
aimed jevel
perception positive

perception
awareness practice
aimed study asked
practices questionnaire
completed knowledge
participants
online pop attitudes we perception
pelevelo partitude questions
perception
awareness practice
guestionnaire
completed knowledge
participants
online pop attitude questions
pelevelo partitude questions
perception
awareness practice
attitude se

perception positive practices attitudes practicequestionnaire assess online completed regarding level among majority Study which conducted conducted questions total awareness aimed overall

awareness p completed

output

positive a participants of attitudes attitudes of attitudes practices awareness of used information

Out-degree Betweenness PageRank

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

regarding>participants
conducted>participants
conducted>participants
conducted>participants
conducted>participants
knowledge-questionnaire
participants>online among>participants
knowledge-suctionnaire
participants>online among>participants
knowledge-sudy
online>participants knowledge-study
questionnaire>participants
knowledge-sudy
online>participants knowledge-participants
knowledge-sparticipants
participants>participants
participants>participants
participants>participants
participants>participants
participants>participants
participants>participants
participants>coronavirus
participants>coronavirus
participants>coronavirus
participants>coronavirus
participants>coronavirus
participants=questionnaire
study>knowledge-participants
participants=questionnaire
study>knowledge-participants
participants=questionnaire
study>knowledge-participants
participants=questionnaire
study>knowledge-participants
participants=questionnaire
study>knowledge-participants
participants=questionnaire
study>knowledge-participants
participants=coronavirus
participants=questionnaire
participants=coronavirus
participants=questionnaire
participants=coronavirus
participants=questionnaire
participa

attitude>practice mige-recipient confine to conducted>among conducted>among conducted>among conducted>among conducted>among conducted>among conducted>among conducted>among conducted cond

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

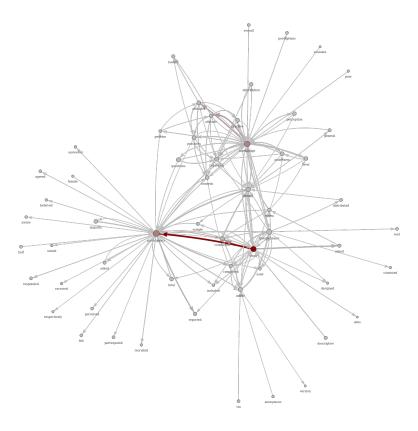


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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