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

Topic 81 companion sheet

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

This file contains the following supplementary information for Topic 81 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
medical organization and management	emergency->medicine (26.7%)	medical->personnel->emergency (7.4%)	emergency->medicine->doctors->management (5.3%)

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	0.728658	0.007522	0.000172	0.000016	0.588177
rank	89	22	23	42	77

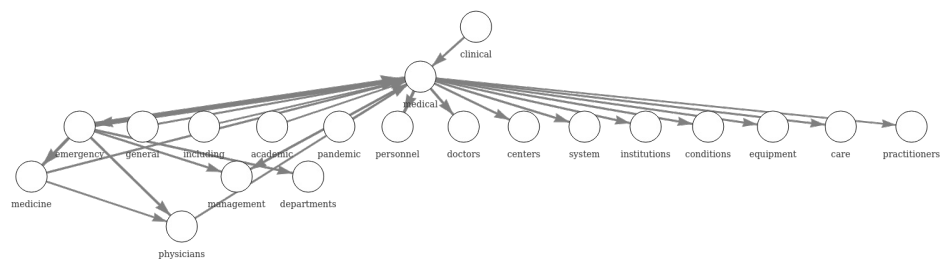Based on the aforementioned measures, Topic 81 has been classified as a CROSS-CUTTING topic.

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

LDA probability

equipment
system patient
general physician
hospitals
EMS doctors clinical support
hospital emergency
centers medical
military medicine
physicians
departments provided including management pandemic
personnel
institution
institutions

Degree

practitioners
academic equipment
institutions
departments
care clinical
management pandemic
personnel including
hospitals emergency medical patient
general conditions medicine doctors support
physicians
centers
system
hospital
practice

In−degree

resources
specialties
physician
doctors teams conditions
departments
clinical management
support emergency
hospital medical
institutions medicine centers
care physicians
practice system
equipment
institution
practitioners assistance

Out−degree

shortage
support departments
pandemic
doctors physicians
hospital clinical seek
electronic medicine
emergency hospitals
medical patient
system general including
management.
military care
personnel internal academic help

Betweenness

practice
clinical medicine emergency
departments provided resources
conditions care centers patient
medical physicians help
practitioners hospital
equipment system personnel management
doctors plan
support
institutions
physician

PageRank

institution
practitioners
conditions
resources hospital centers general
departments care clinical
management doctors hospitals
personnel
medical practice
emergency patient
system support
physicians medicine
institutions including
equipment

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.