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

Topic 14 companion sheet

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

This file contains the following supplementary information for Topic 14 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
public health response to pandemics	public->system (23.3%)	public->system->crisis (4.8%)	public->system->crisis->health (3.5%)

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.
 0.863546	0.008725	0.000199	0.000000	0.612160
114	86	34	26	100

Based on the aforementioned measures, Topic 14 has been classified as a CROSS-CUTTING topic.

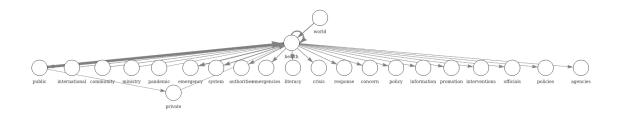
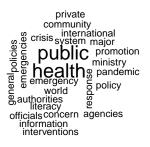


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.









promotion authorities community emergencies improve we pullcy pandemic information public of literacy health by world was a world with the pullcular response of international of interventions of strengthen promotions.

agencies
surveillanceofficials
including pandemic
de international
crisis response
crisis response
under a the crisis response
emergency public
authorities at some control of the crisis response
and the crisis response

Out-degree Betweenness PageRank

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

health-amergendee
COVID-18-cases
united-states
health-crisis novel-coronavirus
health-trisis novel-coronavirus
acute-respiratory
COVID-19-patients
acute-respiratory
ECOVID-19-pademic
public>health
mental-health world-health
health-care health-emergency
respiratory
syndrome-coronavirus
social-distancing
care-workers
patients-COVID-19

health-private of the private of the

health-interventions community-health international-health of the health-literacy (above health health-specified public health-specified public health-specified health-specifie

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

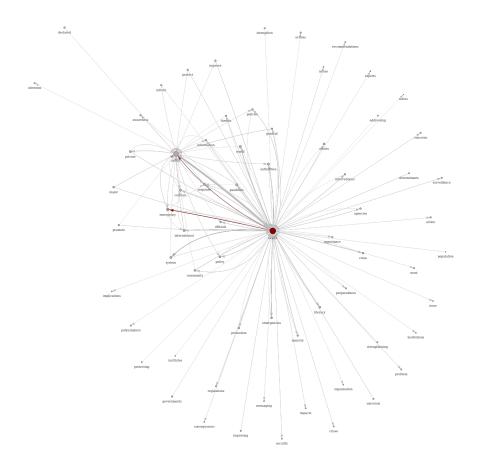


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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