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

Topic 42 companion sheet

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

This file contains the following supplementary information for Topic 42 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
symptoms & clinical manifestations	clinical->symptoms (21.9%)	gastrointestinal->infection->symptoms (6.7%)	clinical->symptoms->manifestations->reported (1.9%)

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	0.715687	0.008407	0.000343	0.021911	0.599889
rank	84	69	68	64	94

Based on the aforementioned measures, Topic 42 has been classified as a CROSS-CUTTING topic.

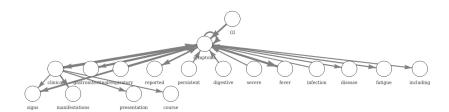


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

persistent to taste history cases history infection fatigue of throat clinical symptoms

Symptoms

loss fever disease manifestations mell gastrointestinal headache diagnosis pain

presented to presented to present digestive to the present digestion to the present digestive to the present digestion to

present right right present right present right right present right righ

manifestations infection digestive digestive separated of the control of the cont

gastrointestinal infection pain Statigue GI 9 gastrointestinal infection gastrointestinal gastro

diarrhea system of the control of th

Out-degree Betweenness PageRank

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

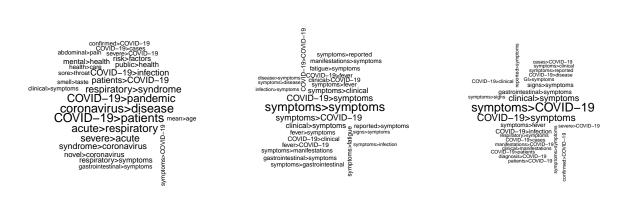


Figure 3: Top 25 bigrams (i.e., edges) by measure.

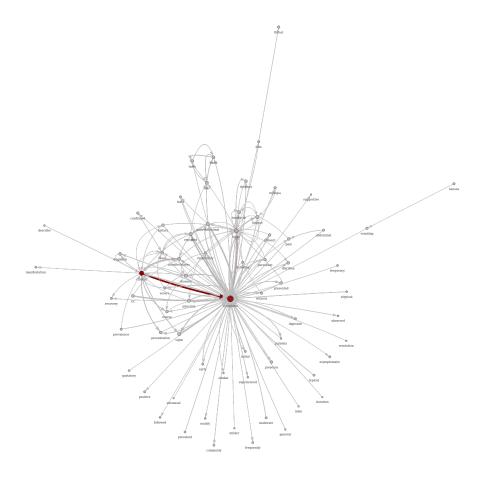


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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