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

Topic 74 companion sheet

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

This file contains the following supplementary information for Topic 74 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
clinical cases reports	patient->presented (14.3%)	patient->presented->developed (7.1%)	patient->presented->developed->acute (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.
value	0.620061	0.009522	0.000526	0.317559 115	0.554442
rank	49	108	93		32

Based on the aforementioned measures, Topic 74 has been classified as a SPECIALIZED topic.

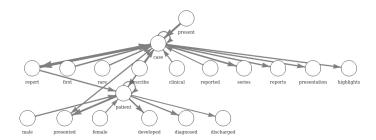


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

woman
Gmanifestations
Developed
daysing developed
daysing diagnosis
clinical of the diagnosed
present of the describe
positive patient days
presentation
Series cases
cutaneous

described reported reported reported rediagnosed positive FepOrt acute female patient of developed diagnosis first series on highlights in the discharged reported re

symptoms diagnosis dischargedpresentation clinical presented series patient established by the series patient established establishe

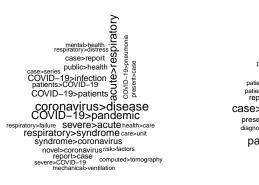
positive and processing the processing positive and processing pro

diagnosis reported remained right fever spresentation patient trace history of the presented page of the prese

reported
supplied in the property of the prope

Out-degree Betweenness PageRank

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



presents-case presentation-case presentation-case presentation-case patient-history case-history patient>case patient-presented case-patient case-presented case-presented case-patient case-presented ca

patient-discharged clinical-case cases-highlights cases-cases-highlights cases-cases-highlights presented-case-cases-highlights presented cases-

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

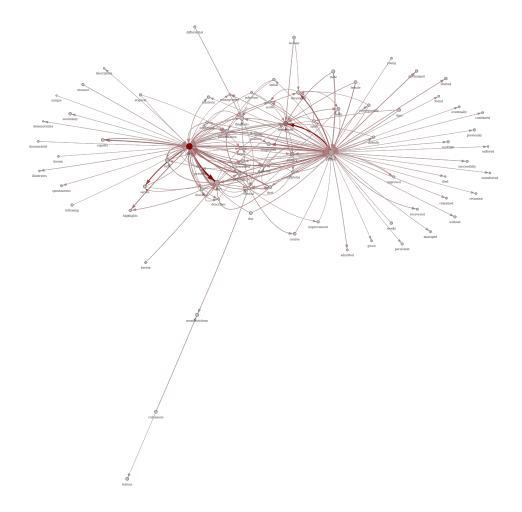


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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