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

Topic 98 companion sheet

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

This file contains the following supplementary information for Topic 98 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
icu and patinet care	ICU->patients (6.7%)	care->unit->patients $(4.7\%)$	patients->admitted->hospitalized->hospital (1.8%)

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	0.775632	0.009010	0.000402	0.000085	0.620277
rank	99	95	76	43	106

Based on the aforementioned measures, Topic 98 has been classified as a CROSS-CUTTING topic.

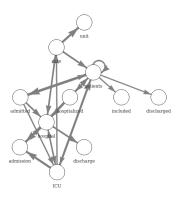


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

inpatient patient
beds hospitals requiring
march admission
units patients
required hospital required hospital admitted Deare
hospitalized Oward hospitalized Oward hospitalized of unit substitution discharge of patients.

outcomes
march required
discharged Unit tertiary
discharged Unit tertiary
discharged Unit tertiary
died hospital g,
hospitals patients to days
admission of the discharged died hospitalized
beds patient of received
hospitalization of the discharged died hospitalization

hospitalization discharged unitscare hospital received hospital received admission patients required hospitalized hospitals outcomes

inpatient died requiring admission of discharge of the critical ICU of the critical of the critical ICU of the critical of the

ambulatory outcomes p outcomes p

hospitalization discharged medical nospitalized unit confirmed uni

Out-degree Betweenness PageRank

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



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

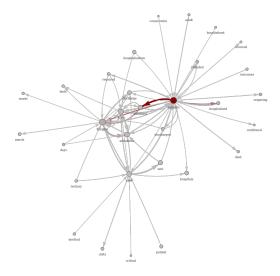


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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