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

Topic 34 companion sheet

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

This file contains the following supplementary information for Topic 34 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
prevention	prevention->measures (26.8%)	control->prevention->measures (9.4%)	prevention->measures->spread->strategies (3.6%)

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	0.742385	0.007710	0.000173	0.024285	0.573755
rank	92	29	24	65	59

Based on the aforementioned measures, Topic 34 has been classified as a CROSS-CUTTING topic.

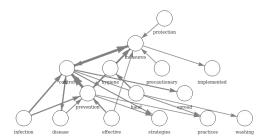


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.



implemented compliance efforts precautionary preventiative prevention reduce precaution infection important provided current spread practices approaches a precaution provided current including washing proventing proventing proventing important provided current provided current preventing proventing proventing important provided current preventing proventing proventing proventing important provided current preventing proventing provided current preventing proventing proventing proventing proventing provided proventing provention proventing prove

Out-degree Betweenness PageRank

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

covid-health syndrome-scoronavirus
covid-health syndrome-scoronavirus
covid-health syndrome-scoronavirus
covid-health syndrome-scoronavirus
covid-health syndrome-scoronavirus
disease-control Severe-acute
covid-19-cases acute-respiratory spread>covid-19
coronavirus-disease,
covid-health spread-scorid-health-scare
prevention-scontrol production of the scarce of the

control>compliance
preventions:COVID-19
hands-prevention
measures.bright |
hands-measures |
hygiene-control Control |
Control>control |
Control>control |
measures.bright |
me

priterion-measure effective-control control-strateges provention-strateges provention-strateges provention-strateges prevention-control-spread prevention-control control-prevention-control control-prevention-control-control-prevention-control-prevention-control-prevention-control-control-prevention-control-control-prevention-control-control-prevention-control-cont

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

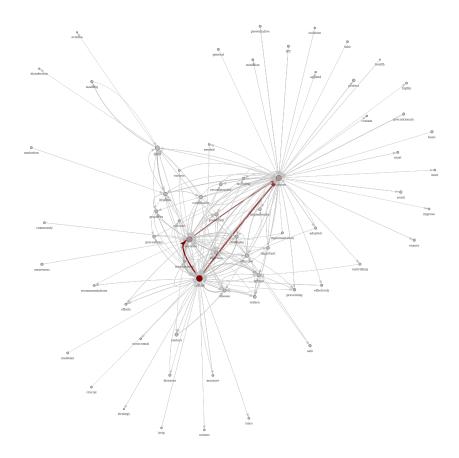


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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