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

Topic 78 companion sheet

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

This file contains the following supplementary information for Topic 78 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
geographic (countries)	world->health (7.6%)	around->world->health (3.9%)	around->world->health->response $(1.3\%)$

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	0.587759 $31$	0.007382	0.000194	0.167829	0.554247
rank		14	30	99	31

Based on the aforementioned measures, Topic 78 has been classified as a CROSS-CUTTING topic.

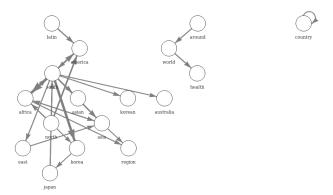


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

regions occupantly
understand the state of t

including australia world of australia world of australia world of westernamerica of successions around africa eigen around africa eigen around africa eigen of australia eigen eigen of australia eigen of australia eigen of australia eigen of

burden health of the sain world country australia malaria international

international including
since korea — western
latin b regions
africa c region
SOUTH southeast
world
central asia east first
america japanasian
around nigeria
australia

prevalence
malaria provide mal

regions low malariaworld including first asia country special and a different policies in the country special and a different policies in the country of the

Out-degree Betweenness PageRank

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

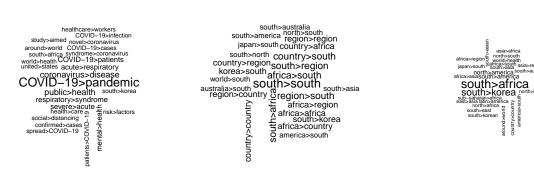


Figure 3: Top 25 bigrams (i.e., edges) by measure.

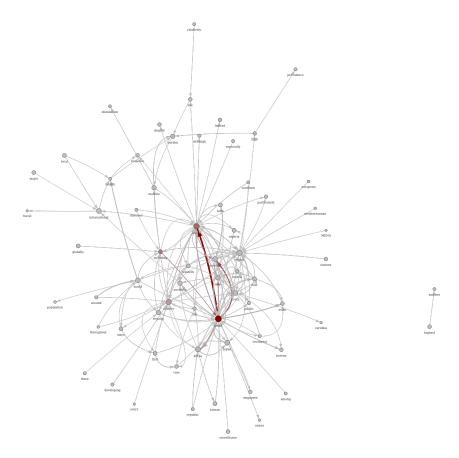


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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