

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

## Topic 75 companion sheet

G. Minello

C.R.M.A. Santagiustina

M. Warglien

This file contains the following supplementary information for Topic 75 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
numbers	one->three (22.8%)	one->three->study (9.2%)	one->three->study->five (1.5%)

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	0.794277	0.007406	0.000089	0.059564	0.626822
rank	102	15	2	73	112

Based on the aforementioned measures, Topic 75 has been classified as a CROSS-CUTTING topic.

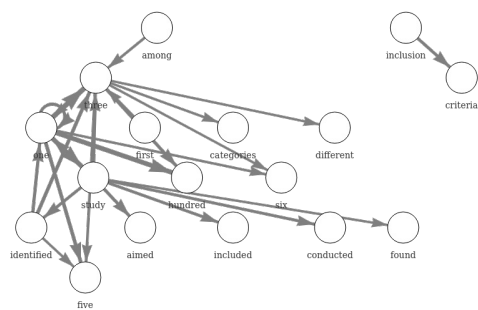


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

A word cloud containing various terms related to the study, including: developed, included, found, including, among, eight, six, met, inclusion, five, study, seven, categories, nine, one, total, hundred, three, identified, criteria, ten, followed, required, and none. The words are of varying sizes and orientations, with 'three' and 'one' being the largest.

3



