

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

## Topic 33 companion sheet

G. Minello

C.R.M.A. Santagiustina

M. Warglien

This file contains the following supplementary information for Topic 33 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
abstract structure	methods->results (15.4%)	methods->results->study (7%)	materials->methods->results->study (2.4%)

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	0.827364	0.007697	0.000098	0.074756	0.623041
rank	109	28	3	76	109

Based on the aforementioned measures, Topic 33 has been classified as a CROSS-CUTTING topic.

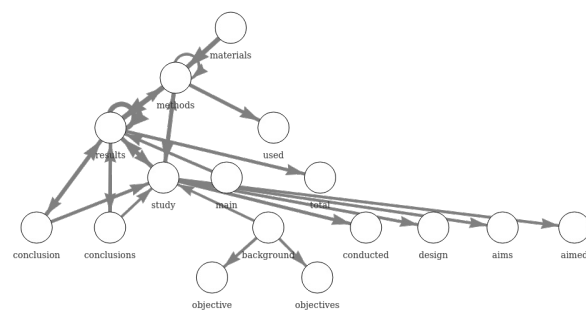


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

A word cloud visualization showing various terms related to research methodology. The most prominent words are "results", "study", "methods", "conclusion", "aimed", "method", "analyzed", "total", "conducted", "aim", "objectives", "materials", "included", "assess", "aims", "used", "design", "main", "background", "conclusions", "relevance", "obtained", "evaluate", "data", and "objective". The words are arranged in a circular pattern, with their size indicating their frequency or importance in the dataset.

A word cloud containing terms such as 'discussion', 'conclusion', 'aims', 'conclusions', 'included', 'used', 'study', 'background', 'conducted', 'abstract', 'main', 'method', 'evaluate', 'objectives', 'results', 'materials', 'aimed', 'obtained', 'analyzed', 'design', 'data', and 'introduction'. The words are arranged in a circular pattern, with 'study' and 'methods' being the most prominent.

[illegible]

3

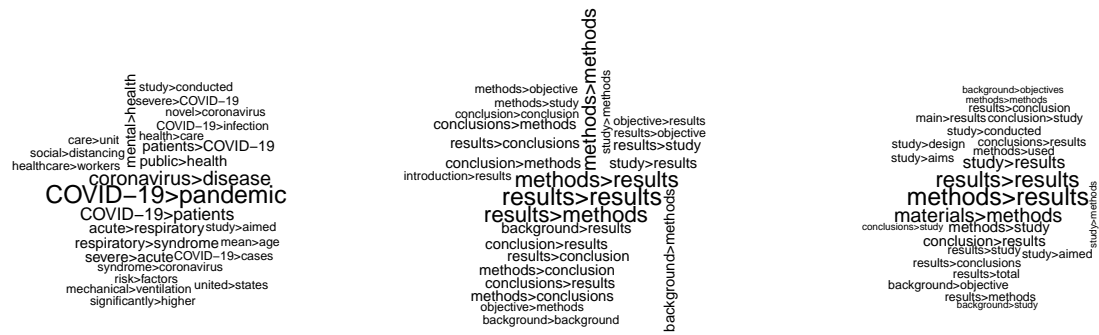


Figure 3: Top 25 bigrams (i.e., edges) by measure.

