

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

Topic 119 companion sheet

G. Minello

C.R.M.A. Santagiustina

M. Warglien

This file contains the following supplementary information for Topic 119 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
related works	related->pandemic (10%)	related->pandemic->including (3.4%)	status->related->pandemic->including (0.6%)

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	0.969251	0.008297	0.000101	-0.229904	0.573267
rank	120	66	5	1	57

Based on the aforementioned measures, Topic 119 has been classified as a CROSS-CUTTING topic.

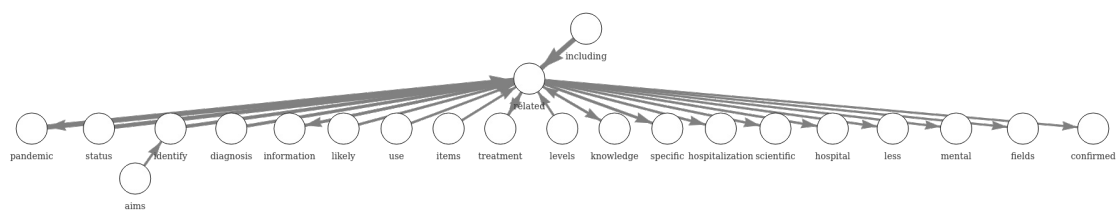


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

A word cloud visualization showing various terms associated with COVID-19. The most prominent words are "related", "pandemic", and "status". Other visible words include "mayo", "addition", "clinic", "hospital", "amongst", "virus", "levels", "aims", "specific", "including", "diagnosis", "use", "significant", "confirmed", "treatment", "definite", "following", "identity", "el", and "ss". The words are arranged in a circular pattern, with some overlapping.

A word cloud of COVID-19 related terms. The word 'related' is the largest and most central. Other prominent words include 'hospitalization', 'knowledge', 'status', 'confirmed', 'specific', 'pandemic', 'diagnosis', 'fields', 'information', 'new', 'treatment', 'including', 'hospital', 'scientific', 'use', 'among', 'conditions', 'mental', 'high', 'less', 'items', 'levels', 'identify', and 'de'. The words are arranged in a circular pattern around the central word 'related'.

3

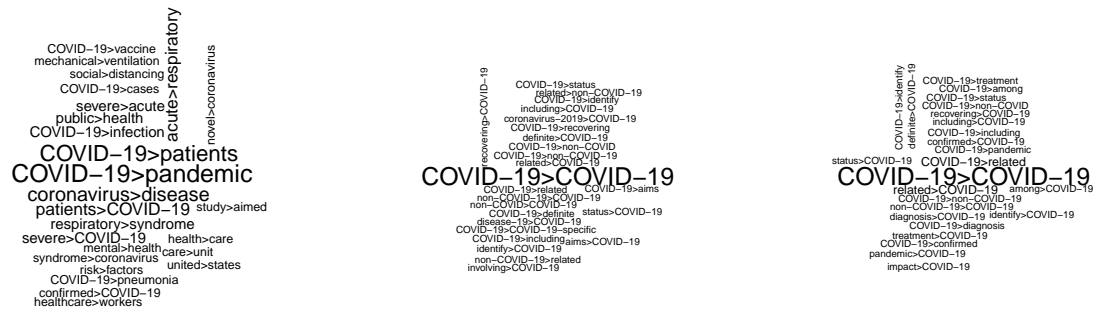


Figure 3: Top 25 bigrams (i.e., edges) by measure.

