

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

Topic 89 companion sheet

G. Minello

C.R.M.A. Santagiustina

M. Warglien

This file contains the following supplementary information for Topic 89 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
risk factors	risk->factors (21.3%)	risk->factor->exposure (8.2%)	risk->factor->exposure->factors (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	0.859537	0.007829	0.000125	0.000000	0.624794
rank	113	37	13	5	110

Based on the aforementioned measures, Topic 89 has been classified as a CROSS-CUTTING topic.

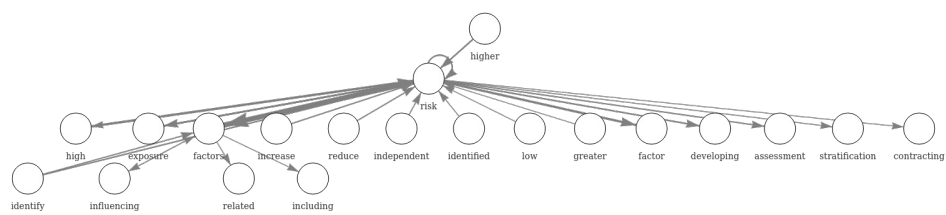


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

A word cloud of terms related to risk factors. The most prominent words are 'risk factors' and 'exposure'. Other visible words include 'independent', 'population', 'developing', 'contracting', 'higher', 'related', 'general', 'including', 'among', 'individual', 'assessment', 'identified', 'history', 'reduce', 'multiple', 'increase', 'identify', 'considered', 'influencing', and 'exposures'.

A word cloud visualization showing various terms associated with risk factors. The words are arranged in a circular pattern around a central point. The most prominent word is "risk", which is significantly larger than the others. Other visible words include "factors", "population", "determine", "greater", "related", "exposure", "increase", "identified", "low", "high", "reduce", "higher", "important", "individual", "independent", "influencing", "assess", "multiple", "considered", "explore", "identify", "highest", and "increases". The font size of each word corresponds to its frequency or importance in the dataset.

Out-degree

Betweenness

PageRank

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

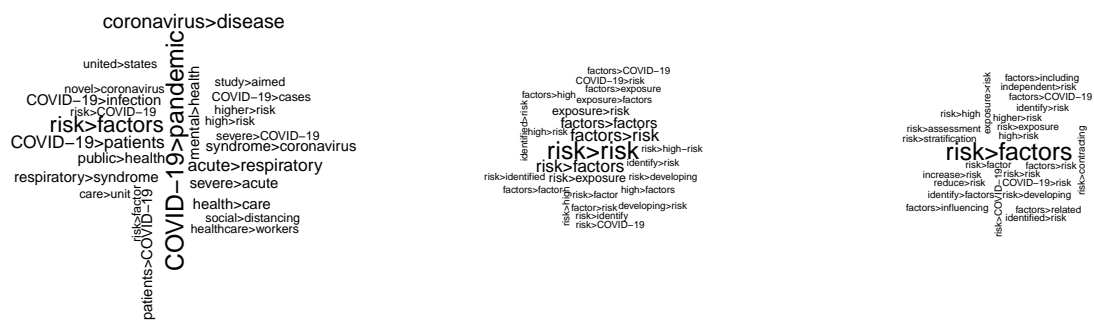


Figure 3: Top 25 bigrams (i.e., edges) by measure.

