

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

Topic 45 companion sheet

G. Minello

C.R.M.A. Santagiustina

M. Warglien

This file contains the following supplementary information for Topic 45 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
impact of air pollution	concentration->meteorological (8.6%)	concentration->meteorological->conditions (7.9%)	concentration->meteorological->conditions->observed (5.2%)

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	0.528307	0.008035	0.000632	0.075814	0.537123
rank	12	51	101	77	22

Based on the aforementioned measures, Topic 45 has been classified as a SPECIALIZED topic.

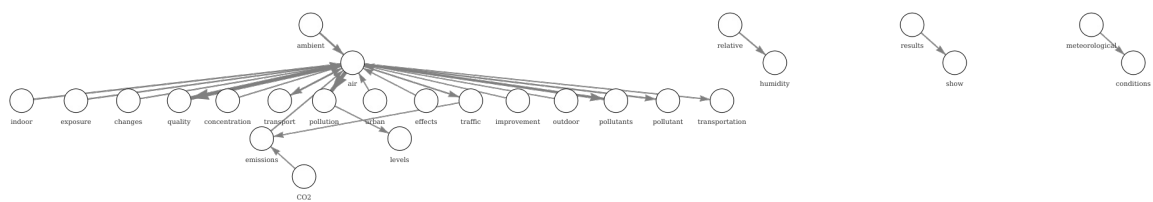


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

exposure  
reduced average  
cities conditions  
PM humidity  
ozone CO<sub>2</sub> transport  
relative quality  
results  
period  
pollution  
emissions  
concentration  
pollutants weather  
transportation

significant  
reductions  
exposure  
concentration  
quality  
urban  
transport  
meteorological  
effects  
levels  
pollution  
average  
indoor  
atmospheric  
reduced  
results  
traffic  
relative  
changes  
pollutants  
significant

A word cloud visualization of the abstract 'air'. The word 'air' is the largest and most central. Other prominent words include 'pollution', 'emissions', 'quality', 'levels', 'reduced', 'transportation', 'humidity', 'sources', 'temperatures', 'meteorological', 'respectively', 'index', 'human', 'pollutant', 'traffic', 'concentration', 'change', 'exposure', 'transport', 'changes', 'reductions', and 'compared'. The words are arranged in a circular pattern around the central 'air' word.

A word cloud of terms related to air quality. The words are arranged in a circular pattern. The most prominent words are 'air', 'pollution', 'quality', 'PM', 'emissions', 'pollutants', 'significant', 'increase', 'transport', 'sources', 'observed', 'results', 'changes', 'index', 'levels', 'exposure', 'concentration', 'decreased', 'reductions', 'show', 'conditions', 'average', 'human', 'traffic', and 'reduced'. The words are in various sizes and orientations, with 'air' and 'pollution' being the largest.

3

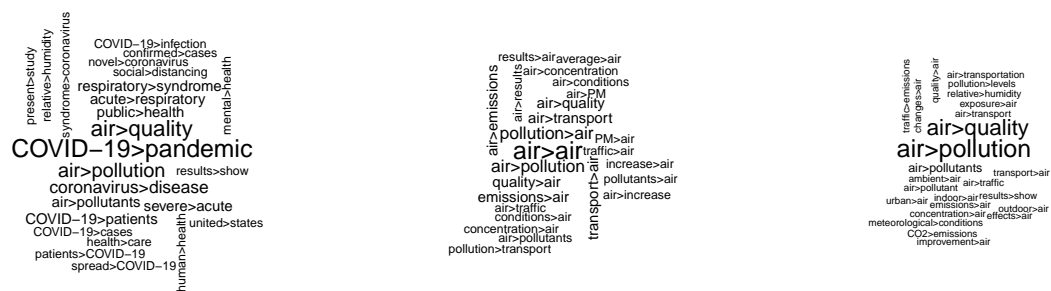


Figure 3: Top 25 bigrams (i.e., edges) by measure.

