

Topic 6 companion sheet

Digging under the surface of COVID-19 topics in scientific literature

G. Minello

C.R.M.A. Santagiustina

M. Warglien

10 November, 2021

This file contains the following supplementary information for Topic 6 of the manuscript *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 25bigrams (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 (99 percentile) topic network (Figure 4)
- Automatic n-gram label proposals of subtopics, if multiple subtopics of large size exist (Table 4)
- Network of top 25bigrams (Figure 5)

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
retrospective clinical study	hospitalized->confirmed (5.4%)	hospitalized->confirmed->without (2.5%)	hospitalized->confirmed->without->diagnosed (0.8%)

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	0.894972	0.009792	0.000263	0.000000	0.658041
rank	117	110	47	6	120

Based on the aforementioned measures, Topic 6 has been classified as a CROSS-CUTTING topic.

A word cloud of COVID-19 related terms. The word 'patients' is the largest and most central. Other prominent words include 'diagnosed', 'hospitalized', 'cohort', 'undergoing', 'recovered', 'admitted', 'included', 'received', 'without', 'confirmed', 'discharged', 'positive', 'hospital', 'analyzed', 'age', 'suspected', 'presented', 'clinical', 'study', 'enrolled', 'died', 'required', 'underwent', 'receiving', and 'analyzed'. The words are arranged in a circular pattern around the central 'patients' word.

Out-degree

Betweenness

PageRank

2

counts

probs

weights (probs*counts)



Figure 2: Top 25 bigrams (i.e., edges) by measure.

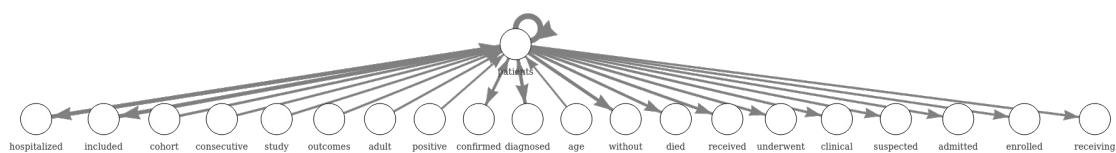


Figure 3: Network of top 25 bigrams (i.e., edges) by weight.