

Topic 2 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 2 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
diagnostics	method->developed (8.6%)	method->developed->used (3.4%)	method->developed->rapid->sensitive (1.7%)

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	0.572876	0.008702	0.000496	0.087023	0.554008
rank	23	84	90	81	30

Based on the aforementioned measures, Topic 2 has been classified as a SPECIALIZED topic.

LDA probability



Degree



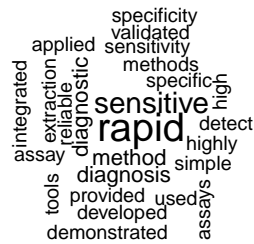
In-degree



Out-degree



Betweenness



PageRank



Figure 1: Top 25 unigrams (i.e., nodes) by measure.

counts

probs

weights (probs*counts)

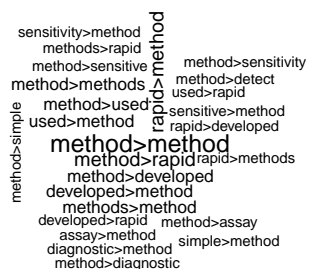
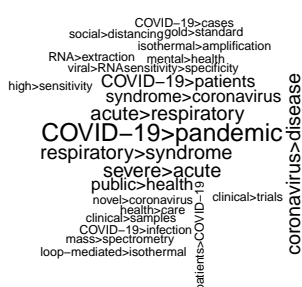


Figure 2: Top 25 bigrams (i.e., edges) by measure.

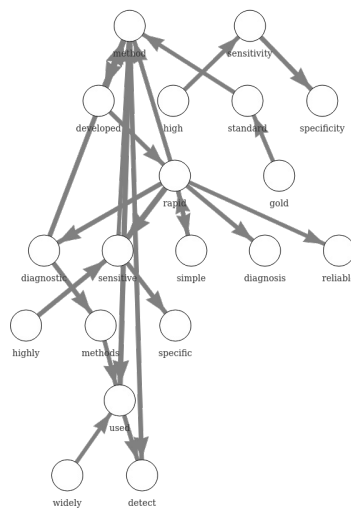


Figure 3: Network of top 25 bigrams (i.e., edges) by weight.