

Topic 1 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 1 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
protective devices	protection->equipment (13.9%)	protection->equipment->used (5.5%)	filtration->protection->equipment->used (1.8%)

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	0.523499	0.008676	0.000724	0.243170	0.521774
rank	9	82	107	109	11

Based on the aforementioned measures, Topic 1 has been classified as a SPECIALIZED topic.

A word cloud of COVID-19 related terms. The most prominent words are 'mask', 'face', 'respirators', 'protection', 'airborne', 'particles', 'wearing', 'filtration', 'N95', 'surfaces', 'droplets', 'aerosol', 'equipment', 'used droplet', 'disinfection', 'reduce', 'contamination', 'aerosols', 'room', 'waste', 'use', 'particle', and 'exposure'. The words are arranged in a circular pattern, with 'mask' and 'face' being the largest and most central.

A word cloud of COVID-19 related terms. The most prominent words are 'face', 'mask', 'respirator', 'equipment', 'airborne', 'droplets', 'aerosols', 'use', 'aerol', 'partides', 'wear', 'air', 'droplet', 'N95', 'surgical', 'respirators', 'surfaces', 'contamination', 'shield', 'filtration', 'disinfection', 'efficiency', 'protection', and 'wearing'. The words are arranged in a circular pattern, with 'face' and 'mask' being the largest and most central.

A word cloud of COVID-19 related terms. The most prominent word is 'mask', followed by 'use', 'equipment', 'protection', 'respirators', 'surfaces', 'air', 'efficiency', 'contamination', 'disinfection', 'droplets', 'N95', 'face', 'aerosol', 'particles', 'respirator', 'filtration', 'surgical', 'shields', 'coverings', 'shield', and 'wearing'. The words are arranged in a circular pattern, with 'mask' at the center.

A word cloud of terms related to COVID-19 and face masks. The words are arranged in a circular pattern, with 'face' being the largest and most central word. Other prominent words include 'mask', 'respirator', 'equipment', 'cloth', 'air', 'protection', 'surgical', 'use', 'droplets', 'aerosol', 'N95', 'particles', 'wearing', 'without', 'reduce', 'wear', 'airborne', 'filtration', 'aerosols', 'contaminated', 'respirators', and 'respirator'. The words are in various shades of gray and sizes, creating a dense, circular composition.

A word cloud of COVID-19 related terms. The most prominent word is 'face', followed by 'mask', 'respirator', 'fit', 'virus', 'respirators', 'materials', 'coverings', 'equipment', 'protection', 'droplet', 'aerosol', 'N95', 'filtration', 'wearing', 'used', 'surgical', 'particles', 'aerosols', 'droplets', 'spread', 'covering', 'contamination', 'generated', and 'respirator'. The words are arranged in a circular pattern around the central 'face' and 'mask' words.

A word cloud of COVID-19 related terms. The most prominent words are 'mask', 'face', 'respirators', 'aerosols', 'droplets', 'surgical', 'efficiency', 'via', 'airborne', 'disinfection', 'surfaces', 'contamination', 'spread', 'wearing', 'use', 'particles', 'protection', 'equipment', 'N95', 'air', 'filtration', 'generated', 'droplet', and 'equipment'. The words are arranged in a circular pattern, with 'mask' and 'face' being the largest and most central.

2

counts

probs

weights (probs*counts)

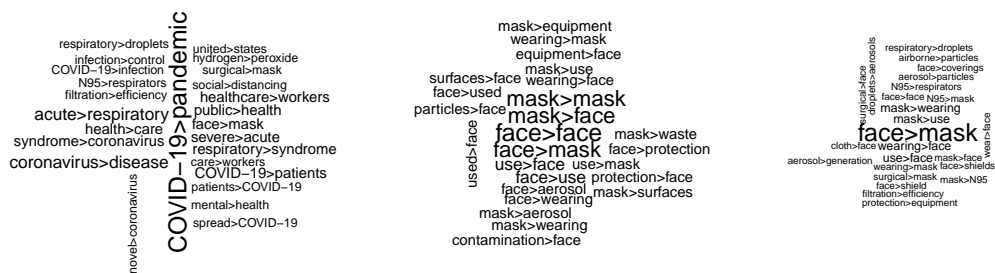


Figure 2: Top 25 bigrams (i.e., edges) by measure.

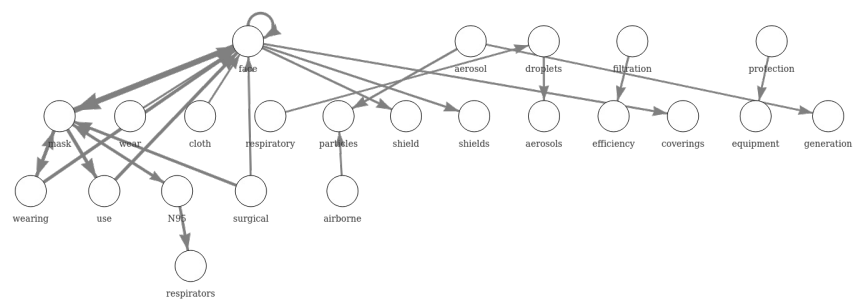


Figure 3: Network of top 25 bigrams (i.e., edges) by weight.