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

Topic 52 companion sheet

G. Minello C.R.M.A. Santagiustina M. Warglien

This file contains the following supplementary information for Topic 52 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
interview studies	themes->emerged $(9.2\%)$	themes->emerged->three $(5.4\%)$	themes->emerged->three->key (1.6%)

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	0.615111	0.008247	0.000330	0.219400	0.601474
rank	46	61	66	108	95

Based on the aforementioned measures, Topic 52 has been classified as a CROSS-CUTTING topic.

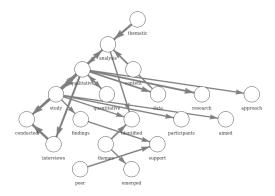


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.



included key content striers peer interviews identified quantitative graph of the peer interviews identified quantitative graph of the peer interviews approach graph of





context quantitative
challenges interviews
identifiedprovided
data conducted
themes research experienced
regarding SUPPORT included
analysis po participants
including supproach includings
approach includings
findings

explore

property interviews
barriersfacilitators
beindentifiedemerged
keyn qualitative used
approach Support indings
Study challenges
analysis of included
research of three
participants occoncerns
including

Out-degree Betweenness PageRank

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

COVID—19-infection interviews-conducted patients-COVID—19 social-distancing patients-COVID—19 social-distancing social-distancing primary-care COVID—19-patients health-acrewides public-health health-acrewides public-health shelth-care COVID—19-pandemic mental-health study-aimed health-care acute-respiratory syndrome-coronavirus thematic-analysis social-support severe-acute semi-structured-interviews physical-activity physical-activity

support-identified barriers-support support-barriers support-barriers support-barriers support-barriers support-barriers support-support identified-supportqualitative-interviews qualitative-support-study analysis-yqualitative-support support-support support-support support-analysis-support support-support support-support support-support support-support support-sup support>conducted interviews>interviews study>qualitative conducted>qualitative

ta-analysisthematic-analysis auditative-study application of the property of t

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

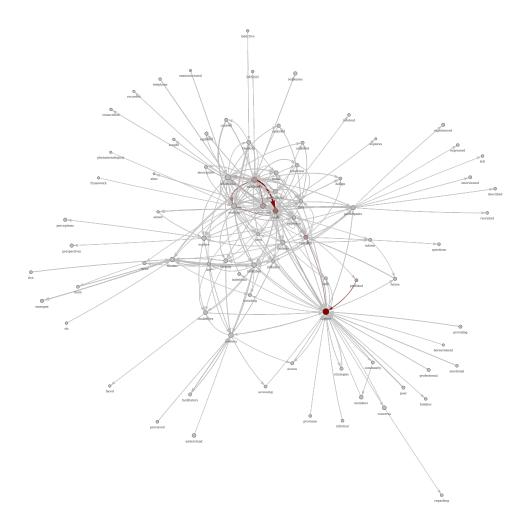


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness centrality. Isolated nodes have been removed after filtration.