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

Topic 51 companion sheet

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

This file contains the following supplementary information for Topic 51 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
age and gender	older->people (14.6%)	older->people->individuals (5.7%)	older->people->individuals->higher (2.5%)

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	0.794482	0.007989	0.000197	0.000000	0.598469
rank	103	45	32	16	91

Based on the aforementioned measures, Topic 51 has been classified as a CROSS-CUTTING topic.

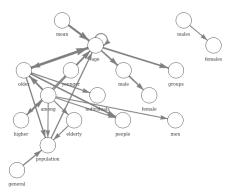


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

individuals
males men
male study
people of among frailty
data
higher oage status
population of mean
population of general
aging female
elderlyyounger
females
prevalence

differences
syyoung general
page of male data
page of male data
page of groups
people ged of meanadult
femaleamong males
younger higher
females study
lower

higher study
malesgroups
elderly propulation
younger E Q Propulation
aged
females older 0 propulation
individuals differences
adult
comorbidities

young men
aged individuals
higheryounger
elderly age to see the people meangeneral
soamong females
adult lower

malesstatus comorbidities individuals men among aging data good aged mean Eage in people by the people of included differences range

comorbidities individuals lower higher agedamongless males population female age of the second populat

Out-degree Betweenness PageRank

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



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

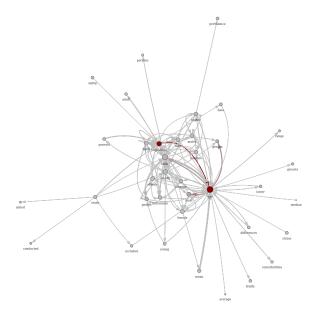


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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