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

Topic 41 companion sheet

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

This file contains the following supplementary information for Topic 41 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
impact on life quality & sexual health	quality->life (50.9%)	quality->life->expectancy (18.2%)	quality->life->expectancy->health (17.9%)

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.
 0.661186	0.006953	0.000179	0.012208	0.630961
62	6	26	60	114

Based on the aforementioned measures, Topic 41 has been classified as a CROSS-CUTTING topic.

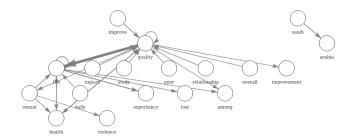


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.



conducted including expectancy £daily to be several garabia of the property of

relationship
conducted people
violence
with a mongstatus
aimsprace in a mongstatus
aimsprace in

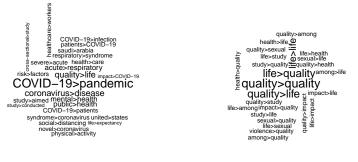
increase
people improving
relationship
relatedpoor improve
impact study data
for a point of the first of the

Some conducted pimprovement sexual quality poor impact data including life among health study saudi expectancy people violence related

status
men significant
minact
minact
men sexpectancy
men significant
minact
min

Out-degree Betweenness PageRank

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



ille-among quality-among health-life to the health-life to the health mily sexual-health and the health-quality sexual-health and the health-quality sub-yule to the health-quality left to the health-quality lef

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

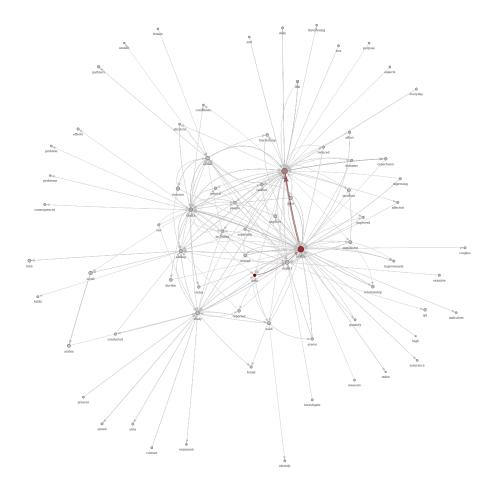


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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