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

Topic 64 companion sheet

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

This file contains the following supplementary information for Topic 64 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
pediatry	children->adolescents (12%)	children->adolescents->young (5.3%)	children->adolescents->young->families (2.4%)

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	0.680914	0.007106	0.000243	0.000000	0.599698
rank	70	9	43	30	93

Based on the aforementioned measures, Topic 64 has been classified as a CROSS-CUTTING topic.

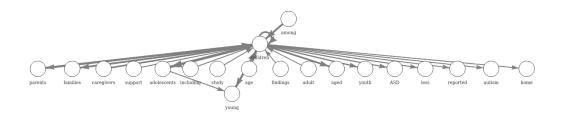


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

adolescent findings
so of the control of the contro

adolescent
ADM autism surport
adult. reported
home DE among
ASD DE young age
included Children Study parents
study parents
caregivers who was a caregiver who was a caregi

developmental among adult so home reported so home report

especially
Page 1

Supply Supp

among
p fathers home support
youth
So fincluding reported
So ASDadolescents
Children
ageprovided sata a
aged by selection agebrovided satual
study on the selection of the selec

Out-degree Betweenness PageRank

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



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

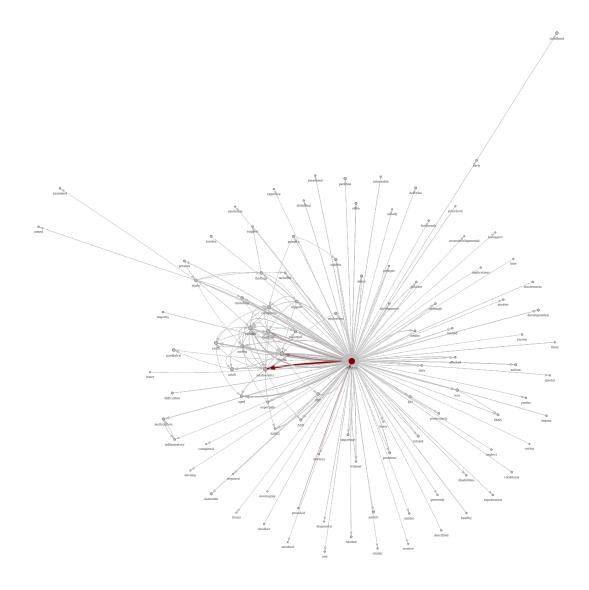


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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