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

Topic 59 companion sheet

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

This file contains the following supplementary information for Topic 59 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
technologies	surface->structure (3%)	process->storage->efficiency (1.3%)	process->storage->parameters->ph (0.9%)

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.
value	0.481577	0.009814	0.001094	0.252551	0.475636
rank	3	111	117	110	4

Based on the aforementioned measures, Topic 59 has been classified as a SPECIALIZED topic.

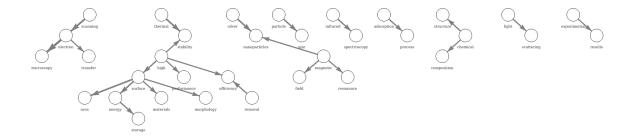


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

investigated
applications
appli

applications Eprocess structure chemical energy high stability electron sused Surface light examples and particles organic microscopy ph different efficiency performance

spectroscopy
thermal to different
chemical
chemi

material
silver a scanning
silver a scanning
used imaterials
light high energy
sourface solution
ochemical a b b
thermal metal
structure membrane
adsorption

characterization
morphology
synthesis
synthesis
chiadren composition
morphology
synthesis
characterization
characterization
morphology
synthesis
characterization
characterization
morphology
synthesis
characteri

performance
the chemical sychemical sychemic

Out-degree Betweenness PageRank

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



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

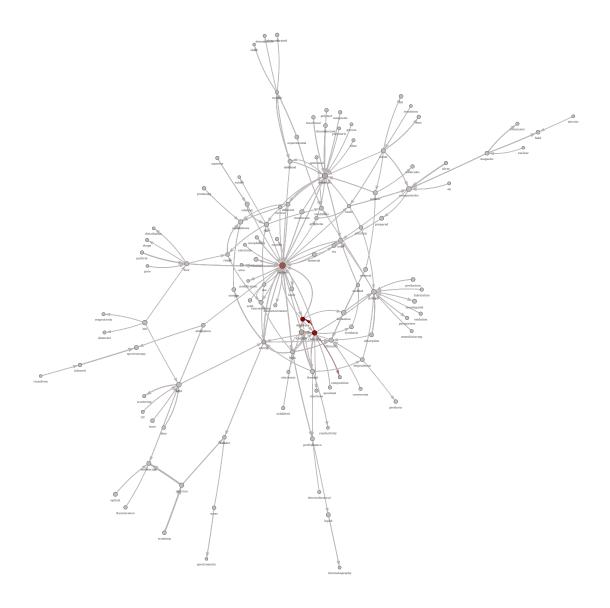


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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