

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

Topic 99 companion sheet

G. Minello

C.R.M.A. Santagiustina

M. Warglien

This file contains the following supplementary information for Topic 99 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
cellular mechanisms	gene->expression (25%)	gene->expression->genes (13.3%)	gene->expression->genes->involved (7.4%)

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	0.523734	0.010199	0.001023	0.000664	0.516244
rank	10	115	116	50	9

Based on the aforementioned measures, Topic 99 has been classified as a SPECIALIZED topic.

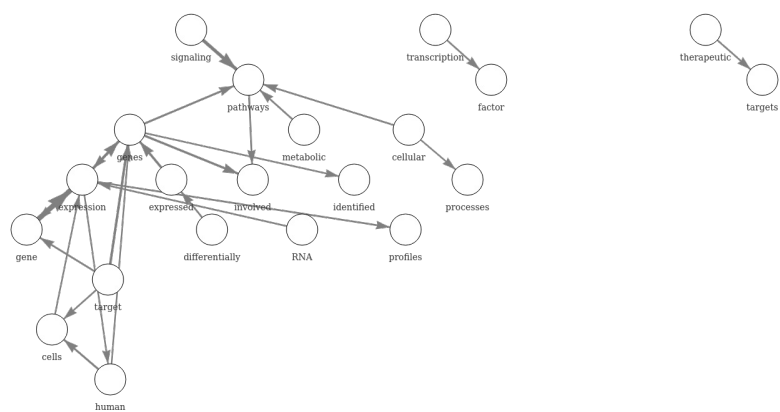


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

Out-degree

Betweenness

PageRank

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

transcription>factor
 patients>COVID-19
 differentially>expressed
 present>study viral>infection
 signaling>pathways
 cancer>cells COVID-19>patients
 novel>coronavirus respiratory>syndrome
 COVID-19>pandemic
 acute>respiratory
 coronavirus>disease
 gene>expression COVID-19>infection
 severe>acute viral>infections
 syndrome>coronavirus mental>health
 public>health epithelial>cells
 viral>replication expression>levels
 oxidative>stress
 severe>COVID-19

expression>cellular
 pathways>genes
 cellular>expression gene>genes
 cells>cells cells>genes
 pathways>expression
 expression>gene
 gene>expression
 cells>expression
 expression>genes
 genes>expression
 expression>cells gene>cells
 expression>pathways cells>gene
 genes>cells genes>gene
 expression>signaling gene>gene
 signaling>expression
 genes>pathways cells>pathways
 pathways>cells

human>genes
 cellular>processes
 transcription>factor
 pathways>involved
 cellular>pathways
 signaling>pathways
 expression>genes
 target>genes
 human>cells
 genes>expression
 therapeutic>targets
 target>cells
 gene>expression
 differentially>expressed
 cells>expression
 genes>identified
 target>gene
 genes>pathways
 expressed>genes
 genes>involved
 expression>profiles
 metabolic>pathways
 expression>human
 RNA>expression
 expression>gene

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

