

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

## Topic 17 companion sheet

G. Minello

C.R.M.A. Santagiustina

M. Warglien

This file contains the following supplementary information for Topic 17 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
NA	important->key (12.2%)	important->key->critical (7.2%)	important->key->critical->crucial (2.8%)

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	0.761353	0.007746	0.000104	0.000000	0.564414
rank	97	33	6	9	44

Based on the aforementioned measures, Topic 17 has been classified as a CROSS-CUTTING topic.

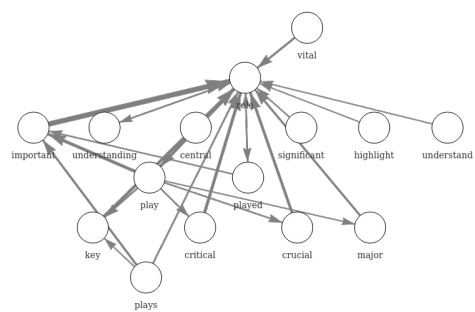


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

A word cloud of terms related to the study. The most prominent words are 'important', 'role', 'critical', 'highlights', 'understanding', 'several', 'played', 'vital', 'key', 'play', 'major', 'plays', 'various', 'crucial', 'central', 'part', 'significant', 'especially', 'determining', 'particularity', 'importance', 'explore', 'development', 'understand', and 'highlights'. The words are arranged in a circular pattern, with 'important' and 'role' being the largest and most central.

A word cloud of terms related to the importance of play. The word 'important' is the largest and most central. Other prominent words include 'play', 'role', 'crucial', 'understanding', 'significant', 'fundamental', 'highlighted', 'discuss', 'key', 'critical', 'several', 'central', 'increasingly', 'explore', 'highlights', 'highlight', 'vital', 'especially', 'particularity', 'importance', 'understand', 'major', 'plays', 'crucial', and 'role'. The words are arranged in a circular pattern around the central word 'important'.

development  
understand  
importance  
providing  
central  
various  
specific  
particular  
present  
implications  
controlling  
understanding  
major part  
important  
played  
plays  
vital  
key  
critical  
including  
several  
development

**role**

developing  
determining  
development  
vital  
understanding  
important  
major  
key role plays  
play played present  
critical various  
importance specific  
crucial providing  
studies especially  
including  
different  
controlling  
several

### Out-degree

## Betweenness

## PageRank

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



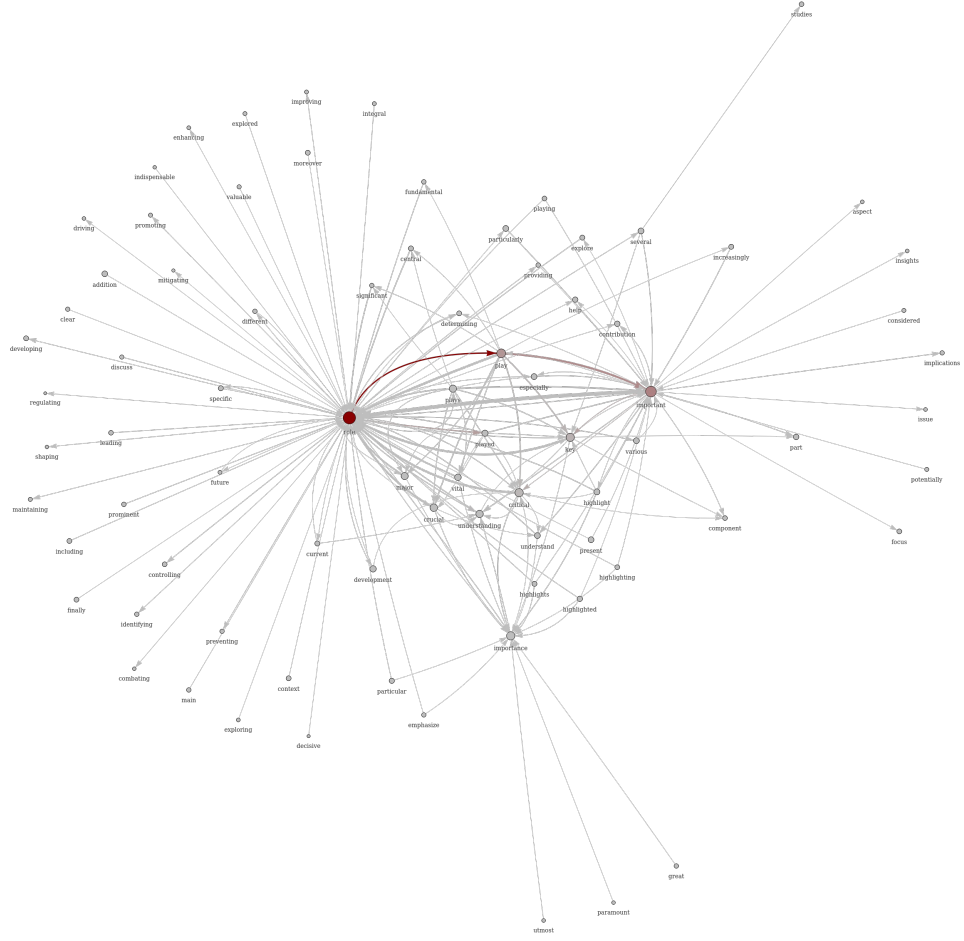


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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