

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

Topic 106 companion sheet

G. Minello

C.R.M.A. Santagiustina

M. Warglien

This file contains the following supplementary information for Topic 106 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
spanish	de->la (21.6%)	de->la->pandemia (15.1%)	de->la->pandemia->esta (4.4%)

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	0.414684	0.008650	0.002981	0.001459	0.448345
rank	2	81	120	52	2

Based on the aforementioned measures, Topic 106 has been classified as a LINGUISTIC OUTLIER topic.

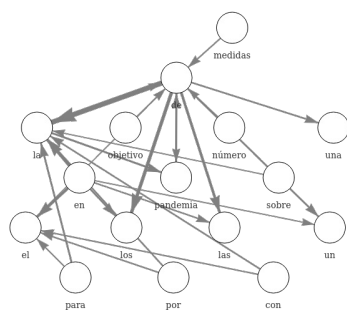
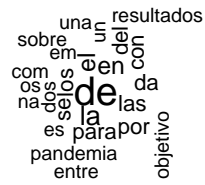
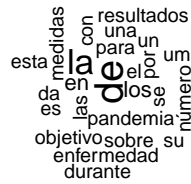


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

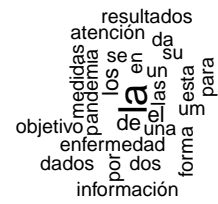
LDA probability



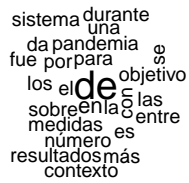
Degree



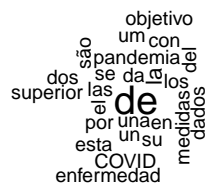
In-degree



Out-degree



Betweenness



PageRank



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

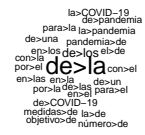


Figure 3: Top 25 bigrams (i.e., edges) by measure.

