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

Topic 92 companion sheet

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

This file contains the following supplementary information for Topic 92 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
predictive models	predictive->value $(20.4\%)$	predictive->value->model $(8.1\%)$	predictive->value->model->used (3%)

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	0.651818	0.007443	0.000384	0.007943	0.596798
rank	60	17	75	59	88

Based on the aforementioned measures, Topic 92 has been classified as a SPECIALIZED topic.

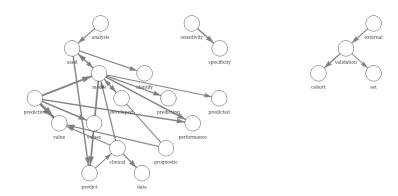


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

independent sensitivity validation of identify area of predicted developed predictive of paralysis value clinical of predicting performance predicting respectively of paralysis value clinical of performance predicting respectively of performance predicting performance performa

area predicted
validated set
validated set
validated set
validation prognostic
performance set
specificity
analysis model cohort
predictive dopto
predictive dopto
predict set
score set
analysis model cohort
predictive dopto
score set
analysis model cohort
predictive dopto
score set
analysis model cohort
predictive dopto
set
analysis model cohort
dopto
specificity
analysis model cohort
dopto
specificity
analysis model cohort
dopto
specificity
analysis model cohort
dopto
set
analysis model cohort
dopto
specificity
analysis

accuracy
variables study
validation analysis
cohort Predicting 32
values model
set Value 6
area predict clinical 6
performancedata
predictive identify
sensitivity validated
predicted

predicted

producted

productive

analysis validation

performance model

data

analysis validation

performance model

develop used

validated clinical light score

prognostic:

external

predict

predict

predict

productive

predictive
value predicting
validated predicted
analysis predict
values
identify best
cohort MOGE
developed Die
dan data
developstudy Die
score
pembelajaranset

pedolepset sensitivity

padolepset sensitivity

pedolepset sensitivity

pedolepset sensitivity

productive

performance is values

perfor

Out-degree Betweenness PageRank

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

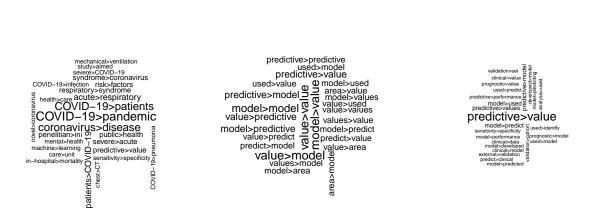


Figure 3: Top 25 bigrams (i.e., edges) by measure.

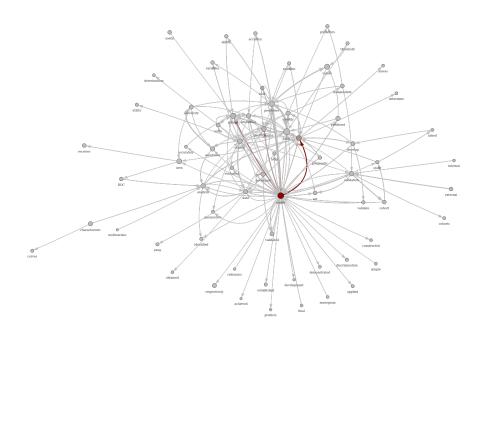


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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