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

Topic 85 companion sheet

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

This file contains the following supplementary information for Topic 85 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
damage mechansims	nervous->system (29%)	nervous->system->involvement (6.1%)	nervous->system->damage->including (3.2%)

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	0.706103	0.008178	0.000177	0.000093	0.575974
rank	80	58	25	44	60

Based on the aforementioned measures, Topic 85 has been classified as a CROSS-CUTTING topic.

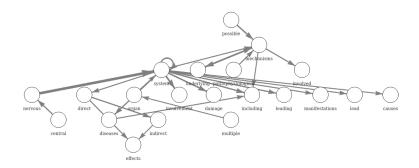


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

manifestations mechanism leading multiple lead several spossible lead several diseases central mechanisms review **System** involved conditions direct a caused effects underlying involvement in indirect in indirect

systemic causes multiple possible several effects systemic possible several effects system in possible several effects effects

w mechanism

w action

w action

w action

lead

lead

nervous

nevous

mechanism including conditions mechanisms

mechanisms

mechanisms

mechanisms

findluding conditions

mechanism

mechanism

findluding conditions

effects something
central no
causes of the centra

processes pespecially in the service of the service

related of possible central of some central of

Out-degree Betweenness PageRank

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

angiotensin-converting-enzyme
distress-syndrome
severe-COVID-19
severe-COVID-19
viral-infection COVID-19-infection
health-care public-health
COVID-19-patients
respiratory-syndrome
coronavirus-disease
COVID-19-pandemic
acute-respiratory ounited-states
clinical-trials Severe-acute
syndrome-coronavirus
nervous-system
novel-coronavirus
respiratory-distress
respiratory

central-system leading-system
system-possible
system-central diseases-system
system-causes system-diseases
system-damage
system-damage
system-leading
system

mechanisms-including system-causes indirect-effects system-mechanisms direct-effects platforbysological-mechanisms direct-effects platforbysological-mechanisms direct-effects system-diseases in system-posses of system-diseases in control of the system-including system-including

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

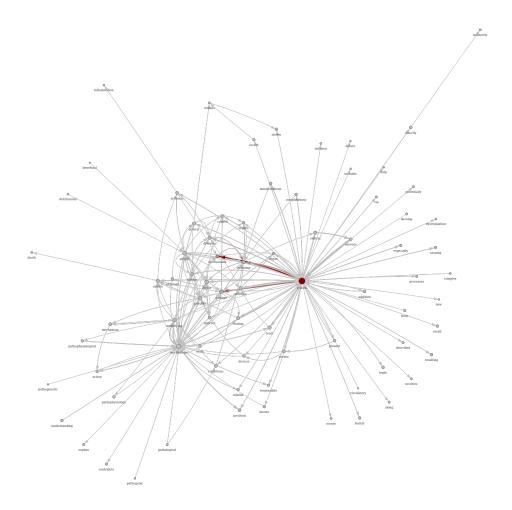


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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