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

Topic 55 companion sheet

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

This file contains the following supplementary information for Topic 55 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
pandemic management	pandemic->response (22.5%)	pandemic->response->crisis (9.4%)	pandemic->response->crisis->management (4.3%)

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.
_	0.706493	0.009293	0.000265	0.177890	0.536143
	82	103	49	101	19

Based on the aforementioned measures, Topic 55 has been classified as a CROSS-CUTTING topic.

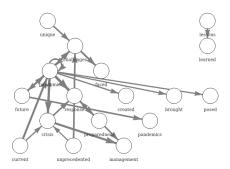


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.



managing planning opportunities planning opportunities plan current some plan curren

situations are of some situations are of some satisfied and some stands of some s

article temergency managing unique times of faced face crisis facing pandemic challenges lessons planning future management unprecedented preparedness strategies

presents

sequence of the control of

unprecedentedpandemics
unprecedentedpandemics
unprecedentedpandemics
unprecedentedpandemics
pandemicmust
Challenges
article response plan. Elearned future lessons a strategies current managing

Out-degree Betweenness PageRank

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

future>challenges
pandemic>management
pandemic>pandemic>future
pandemic>pandemic>future
crisis>challenges
crisis>challenges
crisis>challenges
challenges>challenges
challenges-challenges
challenges-c

unprecedented-crisis of future-complete services and future-complete services and future-complete services and future-pandemics and pandemic-challenges and pandemic-challenges and pandemic-challenges and pandemic-pandemic current-pandemic pandemic-prepared services pandemic-prepared services and pandemic-prepared services and pandemic-prepared services and pandemic-pandemic pandemic pand

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

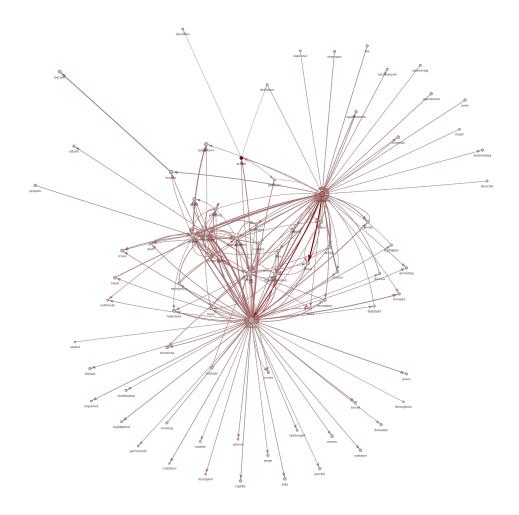


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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