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

Topic 15 companion sheet

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

This file contains the following supplementary information for Topic 15 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
reviews on recent therapeutical approaches	the rapeutic->approaches $(9.7\%)$	future->strategies->various (8%)	future->strategies->various->approaches (5%)

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	0.667622	0.008989	0.000294	0.092629	0.579901
rank	65	92	56	82	65

Based on the aforementioned measures, Topic 15 has been classified as a CROSS-CUTTING topic.

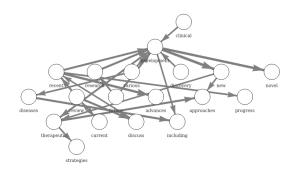
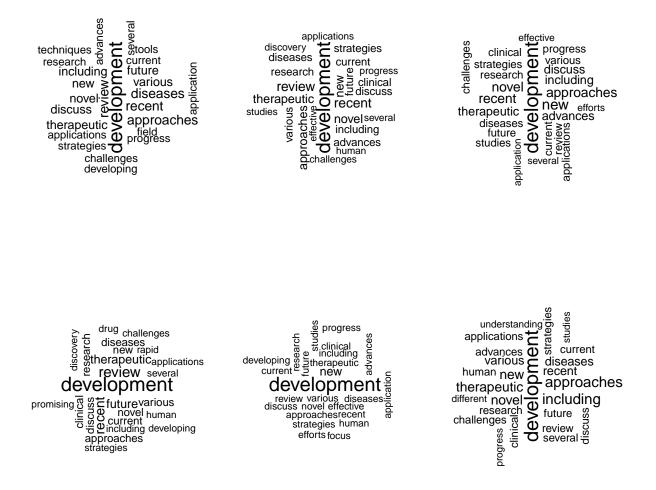


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.



Out-degree Betweenness PageRank

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

treatment-COVID-19 and social-distancing 4 and social-distance 4 and social-distancing 4 and social-di

development development
development proches
development development
development proches
development development
development proches
development development

discovery>development
development>various >
development>therapeutic to the cent-progress A
current-review review>recent to research>development to various-diseases including to new>approaches

development>new
development>new
development>novel
review>discuss review>current
therapeutic>approaches discuss>recent
future>development therapeutic>strategies
recent>development hew>therapeutic
turueversearch
therapeutic>development
clinical>development
developmentsincluding

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

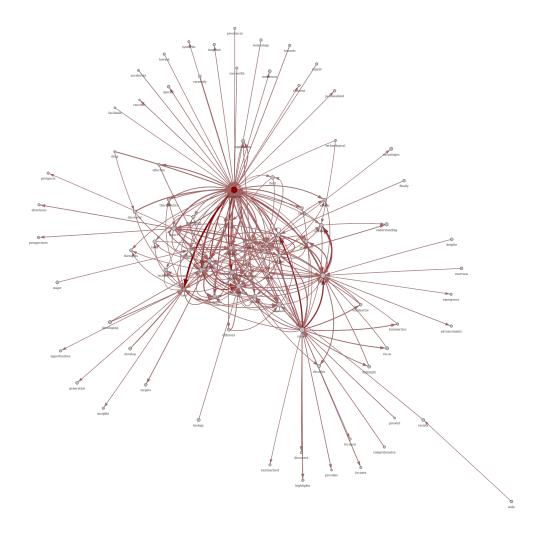


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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