## Topic 5 companion sheet

Digging under the surface of COVID-19 topics in scientific literature

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

10 November, 2021

This file contains the following supplementary information for Topic 5 of the manuscript *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 25bigrams (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 (99 percentile) topic network (Figure 4)
- Automatic n-gram label proposals of subtopics, if multiple subtopics of large size exist (Table 4)
- Network of top 25bigrams (Figure 5)

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.

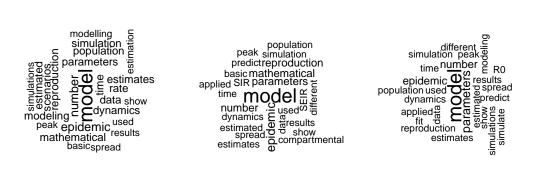
Human label	2-gram label	3-gram label	4-gram label
epidemic models	mathematical->model (10.3%)	epidemic->used->model $(6.7\%)$	reproduction->number->R0->model $(2.9\%)$

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	0.637157 $52$	0.010230	0.000681	0.000000	0.599124
rank		116	105	27	92

Based on the aforementioned measures, Topic 5 has been classified as a SPECIALIZED topic.



population
education
educa

simulations developed simulations developed spredict show presults a number used of data R0 spread spread spread spread spread applied estimated estimates dynamics simulate different

reproduction predict production of predict production of dynamics population of used time Roparameters appread MOGE data depidemic results cestimated scenarios obtained of basic simulations

Out-degree Betweenness PageRank

Figure 1: Top 25 unigrams (i.e., nodes) by measure.

covided-tracing unide-states with the control of th

model-scipulation
model-state data-model
population-model imodel-sdynamics
model-state data-model
population-model imodel-sdynamics
model-sdata model-spetameters
model-sdata model-spetidemic
scenarios-model numbers-model
model-smodel-scipulation-model-imodel-scipulation-model-imodel-scipulation-model-imodel-scipulation-model-imodel-scipulation-model-imodel-scipulation-model-imodel-

schatic-model of model-optamics of the model-optamics optamics of the model-optamics optamics optamics

Figure 2: Top 25 bigrams (i.e., edges) by measure.

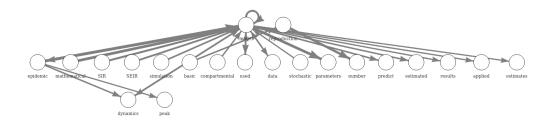


Figure 3: Network of top 25 bigrams (i.e., edges) by weight.