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

Topic 46 companion sheet

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

This file contains the following supplementary information for Topic 46 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
correlation with heart disease	heart->failure (15.6%)	cardiac->arrest->patients (5.8%)	heart->failure->arrhythmias->acute (1.4%)

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	0.543417	0.007510	0.000468	0.007667	0.584159
rank	16	21	87	58	71

Based on the aforementioned measures, Topic 46 has been classified as a SPECIALIZED topic.

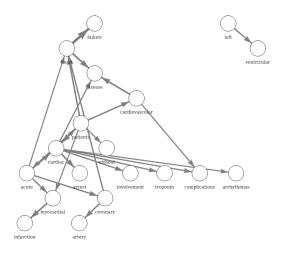


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

pressure
events troponin
le ventricular
otto diccoronary

elevation
involvement
arcardiovascular
patients
patients
cardiacarrhythmias
patients
cardiacarrhythmias
patients
cardiacarrhythmias
patients
acute of arrest to arrest

underwent arthythmias underwent arthythmias underwent arthythmias no cardiovascular complication of the control of the cardiovascular complications of the cardiovascular complications of the cardiovascular arthythmias of the cardiovascular complications of the cardiovascular of the car

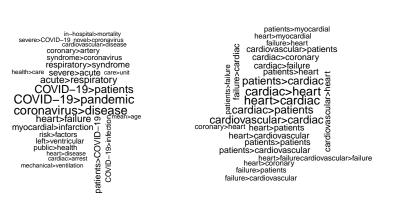
percutaneous adverse elevation cardiovascular cardiovascular stem cardiac coronary without left patients ventricular infarction elevated significant

events death acute qtcinvolvement seincht patients of the pati

shock elevation
events coronary
cardiovascular
troponindisease
acute cardiac
artery patients
clinical patients
involvement heart owithout
myocardial owithout
infarction
complications or arrhythmias
arrhythmias

Out-degree Betweenness PageRank

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



cardiac-complications
cardiac-complications
cardiac-complications
patients-acute patients-bard
patients-acute patients-heart
myocardial-infarction
cardiac-disease heart-disease
patients-cardiac-cardiac-patients
cardiovascular-disease acute-heart
cardiovascular-disease acute-heart
cardiac-patients
cardiac-function patients-without
cardiac-function patients-without
cardiac-function patients-without
cardiac-function patients-without
cardiac-function patients-without

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

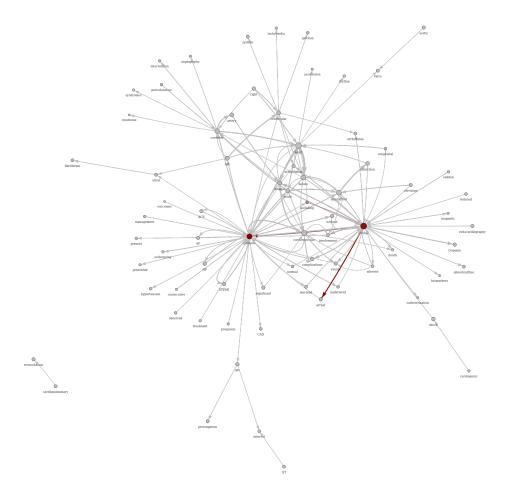


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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