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

Topic 8 companion sheet

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

This file contains the following supplementary information for Topic 8 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
clinical studies on illness development	mortality->rate (16.2%)	mortality->rate->patients (11.9%)	mortality->rate->patients->risk (1.8%)

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	0.716426	0.009364	0.000421	0.078761	0.626357
rank	85	106	80	78	111

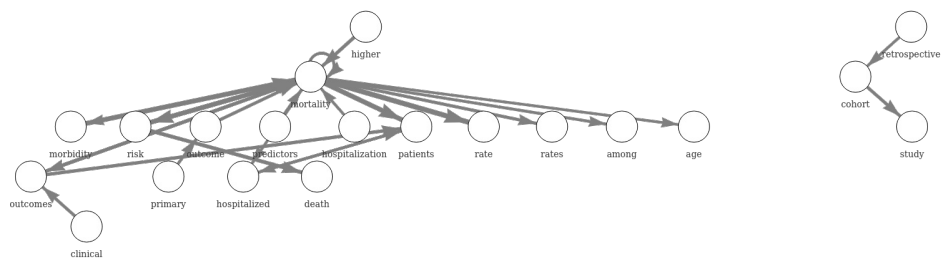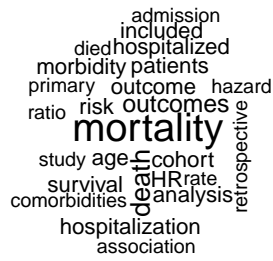Based on the aforementioned measures, Topic 8 has been classified as a SPECIALIZED topic.
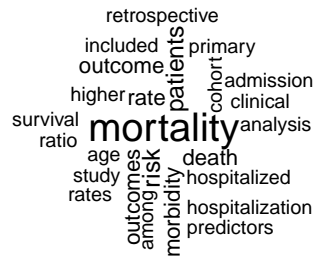
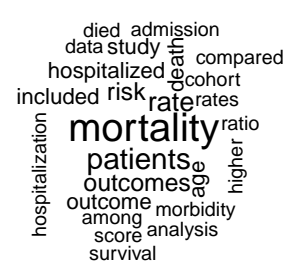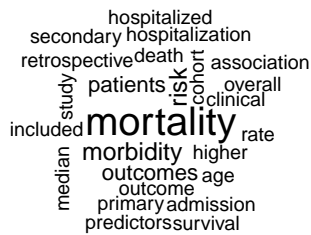Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

LDA probability

admission
included
died hospitalized
morbidity patients
primary outcome hazard
ratio risk outcomes
mortality
study age cohort
survival HR rate
comorbidities death analysis
hospitalization
association

Degree

retrospective
included primary
outcome patients
higher rate admission
clinical
survival mortality analysis
ratio
age death
study hospitalized
rates outcomes
among risk morbidity
hospitalization
predictors

In–degree

died admission
data study death compared
hospitalized cohort
included risk rate rates
mortality ratio
patients
outcomes age higher
outcome morbidity
among
score analysis
survival

Out–degree

hospitalized
secondary hospitalization
retrospective death association
study patients risk cohort overall
clinical
included mortality rate
morbidity higher
outcomes age
median outcome
primary admission
predictors survival

Betweenness

outcome
morbidity
cohort
study age admission
included
risk died data
analysis rate rates
mortality higher
death identified score
patients
hospitalized variables
march

PageRank

morbidity
among
study rates
clinical risk hospitalized data
death rate died age
mortality
patients
included higher
outcomes
analysis cohort hospitalization comorbidities
compared

Figure 2: Top 25 unigrams (i.e., nodes) by measure.
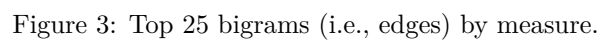
3

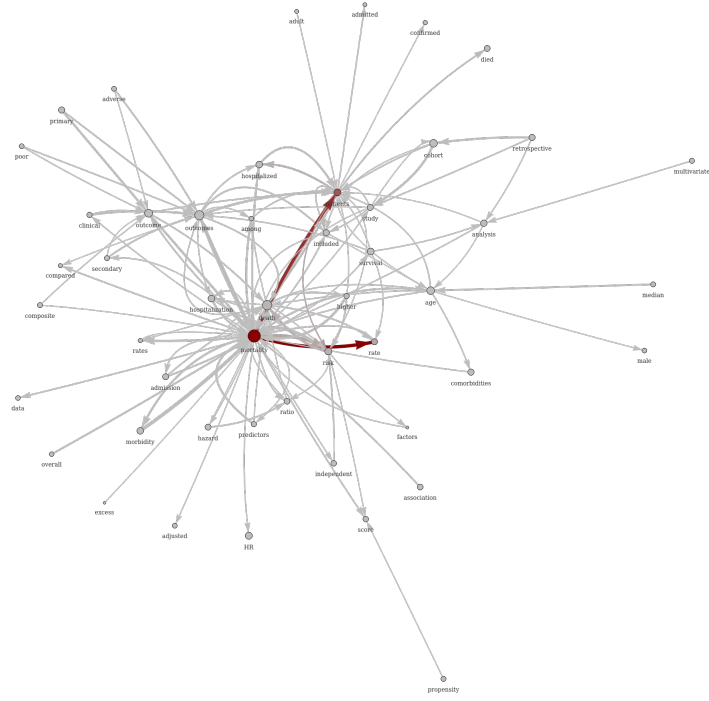Figure 3: Top 25 bigrams (i.e., edges) by measure.

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.