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

Topic 43 companion sheet

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

This file contains the following supplementary information for Topic 43 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
effects of alcohool & addictions	alcohol->medications (10.5%)	alcohol->medications->treatment (8.5%)	alcohol->medications->treatment->cannabis (2.5%)

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	0.724955	0.007183	0.000198	0.000000	0.645955
rank	88	12	33	34	117

Based on the aforementioned measures, Topic 43 has been classified as a CROSS-CUTTING topic.
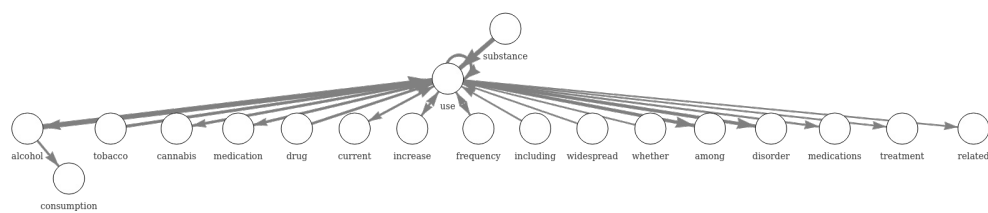
Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

LDA probability

frequency
prescribed
including
alcohol
consumption
effects
usage disorder
drinking
medications
medication
used
use
harm increase
current
significant
smokers among
related
addiction substance
tobacco
prescription
reduce

Degree

including
prior used reduce
drinking
treatment
medications
increase
drug
medication
substance
alcohol
data
use consumption
frequency
among
tobacco
cannabis
related
disorder
current
whether
significant
widespread

In−degree

significant patterns
including frequency
prior increase
smokers consumption
used among
current data
medication use tobacco
disorder related
alcohol
substances medications treatment
cannabis drinking
substance electronic

Out−degree

significant
consumption
whether widespread
marijuana
drug frequency
medications
including medication
substance
alcohol likely
use
cannabis
current used
increase tobacco prior
reduce
regarding among
problematic cigarette

Betweenness

substance
including
medication
related
data
medications consumption
treatment provided
prior substances
electronic use tobacco
among used
alcohol
disorder frequency drug
usage cannabis
increase
current
patterns

PageRank

consumption
frequency cannabis
effects substance
increase smokers
among treatment
alcohol related
drug use tobacco
data
including used reduce
medication current
significant medications
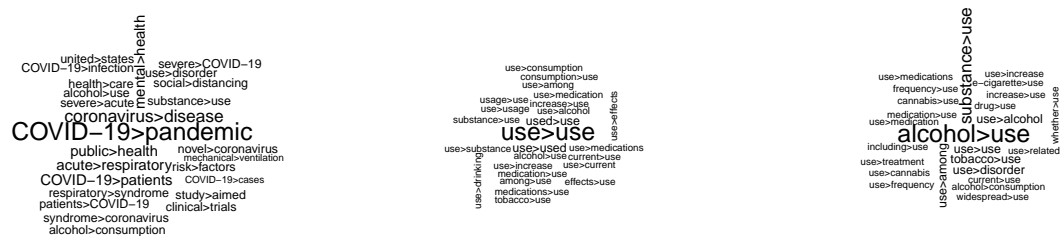prior individuals
disorder

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.