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

Topic 115 companion sheet

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

This file contains the following supplementary information for Topic 115 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
telemedicine	patient->care (10.5%)	telemedicine->visits->telehealth (3.6%)	telemedicine->visits->telehealth->access (1.1%)

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	0.612733	0.008790	0.000565	0.000000	0.578154
rank	43	87	100	25	62

Based on the aforementioned measures, Topic 115 has been classified as a SPECIALIZED topic.
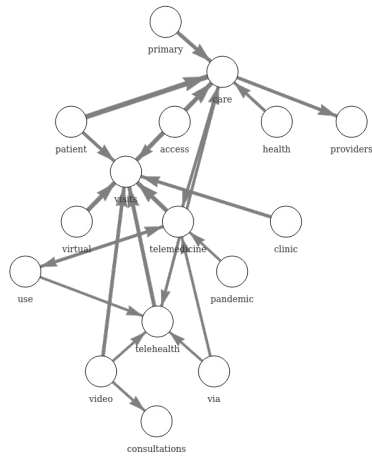
Figure 1: Network of top 25 bigrams (i.e., edges) by weight.

LDA probability

Degree

In–degree

Out–degree

Betweenness

PageRank

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

patient>telehealth
telehealth>telemedicine
visits>telemedicine
care>telehealth
patient>patient
telemedicine>patient
care>patient
visits>patient
care>virtual
virtual>care telemedicine>care
patient>visits care>care
care>telemedicine
patient>care
telemedicine>telemedicine
care>clinic
clinic>care visits>care
telehealth>care
telemedicine>visits
telemedicine>telehealth
telehealth>patient
care>visits
patient>telemedicine

COVID-19>infection
patients>COVID-19
patient>care
care>patients respiratory>syndrome
COVID-19>patients
united>states acute>respiratory
video>visits public>health virtual>visits
health>care telehealth>visits
COVID-19>pandemic
coronavirus>disease
use>telemedicine mental>health
primary>care
social>distancing virtual>care
severe>acute
in-person>visits
syndrome>coronavirus
telemedicine>visits
novel>coronavirus

pandemic>telemedicine
video>telehealth
in-person>care
use>telemedicine
clinic>visits access>care
care>visits virtual>visits
patient>visits in-person>visits
telemedicine>visits
patient>care
virtual>care
telemedicine>care telehealth>visits use>telehealth
visits>telemedicine primary>care health>care
video>visits video>consultations
care>providers
telehealth>care
care>telemedicine
telemedicine>use
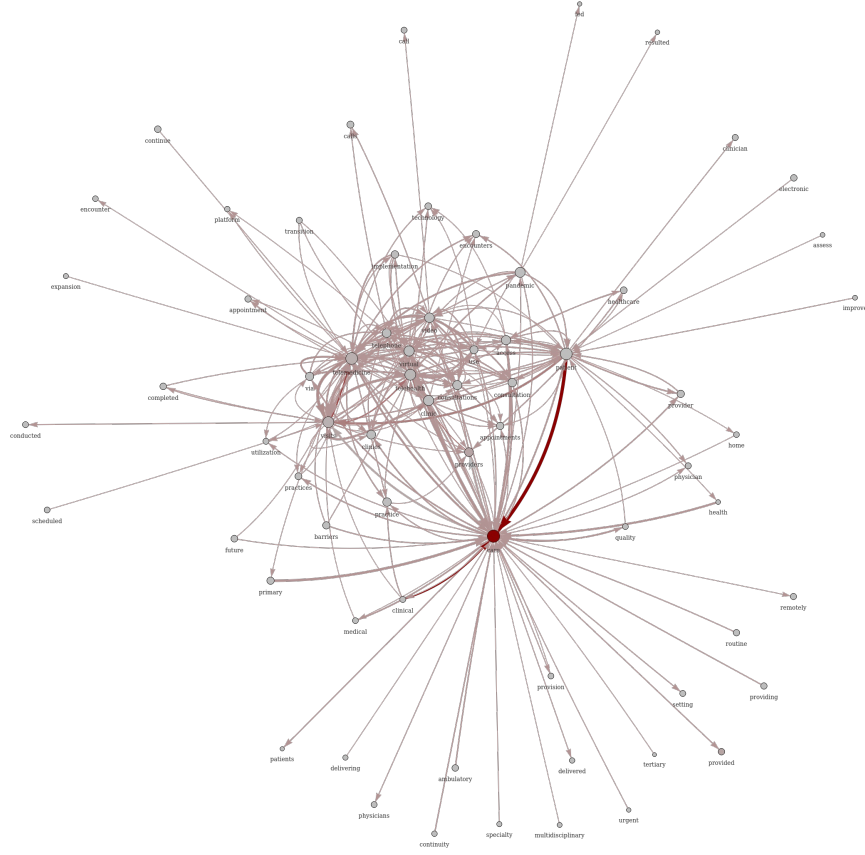
Figure 3: Top 25 bigrams (i.e., edges) by measure.

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.