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

Topic 26 companion sheet

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

This file contains the following supplementary information for Topic 26 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
differentiated access to health realated to socio-economic groups	household->income (13.6%)	household->income->inequality (6.2%)	population->derived->nonprobability->sample (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	0.601686	0.008257	0.000378	0.362758 118	0.571781
rank	38	62	73		54

Based on the aforementioned measures, Topic 26 has been classified as a SPECIALIZED topic.

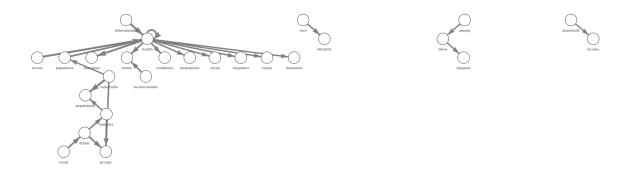


Figure 1: Network of top 25 bigrams (i.e., edges) by weight.



ethnicity
hispanic determinants
communities access
population among or race vulnerable or racial conditions

socioeconomic individuals reacthnicity minority black access ethnic ethn



racial socioeconomic populations 9 people access inequalities inequalities including heart populations of the population of the population hispanic population hispanic

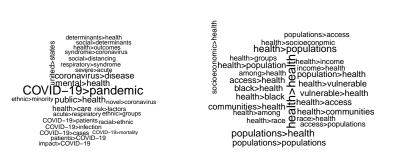
disproportionately ethnicity conditions
including volumerable communities

work to the communities of communities of communities

work to the communities of comm

Out-degree Betweenness PageRank

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



household-income health-incuments health-insurance racial-ethnic health-insurance racial-ethnic health-outcomes determinants-health health-outcomes determinants-health vulnerable-populations

gaccess-health health-inequities health-status health-scale health-status health-scale health-status health-scale health-status health-scale with ethnic-groups race-ethnicity ethnic-groups race-ethnicity ethnic-minority-groups minority-propulations vulnerable-population

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

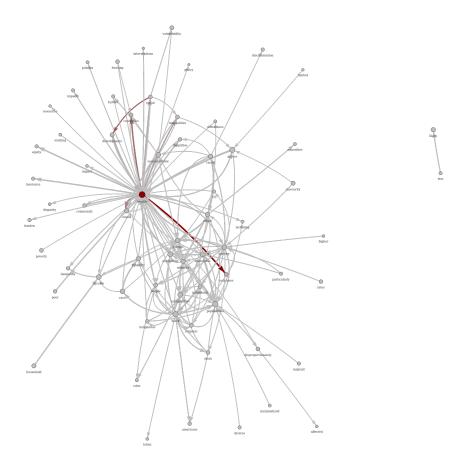


Figure 4: Filtered topic network (by weight). Layout based on Fruchterman-Reingold algorithm. Node size is proportional to topic-specific word probability provided by LDA. Edge width is proportional to topic-specific bigram weight provided by LDA2Net method. Node and edge color represent their betweenness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