



Figure 5: Causal clusters, or causal topics, are shown to the right of the produced causal claim network. Each topic consists of a set of keywords that describes the cluster.

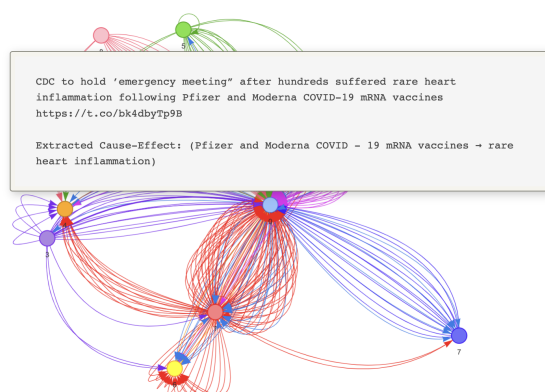


Figure 6: Hovering over an edge in the causal claim network displays the document and extracted causal claim that constitutes that edge. The document is shown at the top of the box, and the extracted cause claim is at the bottom.

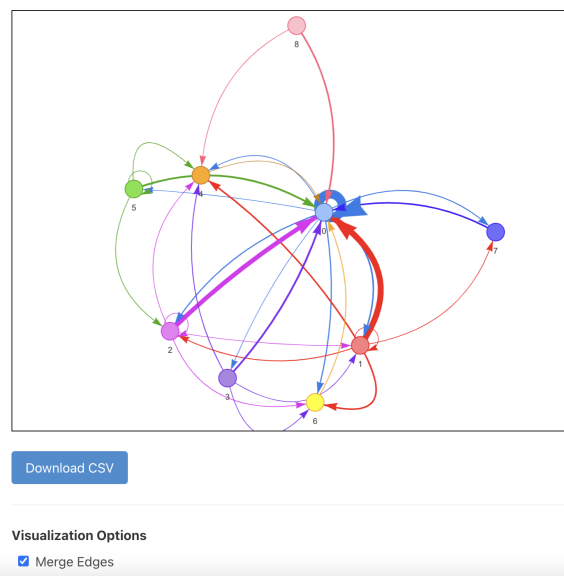


Figure 7: Causal claim network with merged edges, where edge weights equates to the number of documents linking two clusters. Merging edges is useful to quickly assess degree of linkage between causal clusters (nodes) in the network.

Cluster Options

Number of Clusters

After training the topic model, the number of topics that will be reduced. For example, if the topic model results in 100 topics but you have set `nr_topics` to 20 then the topic model will try to reduce the number of topics from 100 to 20.

Setting this value to 0 will automatically reduce topics using HDBSCAN. Setting this value to -1 will not perform topic reduction. Default value is 0.

N-gram range

It relates to the number of words you want in your topic representation. For example, "New" and "York" are two separate words but are often used as "New York" which represents an n-gram of 2. Thus, the `n_gram_range` should be set to (1, 2) if you want "New York" in your topic representation. Default value is (1, 2).

Top N words

Refers to the number of words per topic that you want to be extracted. Default value is 10.

Figure 8: A user can specify parameters when running the pipeline to engage with exploratory data analysis. Users can specify the number of clusters, the n-gram range used during processing, and set the number of words to describe each topic.