**Data extraction**

We used two corpuses: the first one included all the 324 facebook posts of Bernie Sanders since Jan 2018; the second one is made of all the respective comments to the posts.

**Topic clusters and semantic networks**

After parsing the two corpuses, we extracted the 500 most frequent noun phrases and created a semantic network map with 200 nodes to get a first impression of the data. To create our own categories and clean the csv files, we downloaded the cluster categories.

We realized that the names of the default topics created by Cortext were not very accurate, e.g. the topic *richest country in the history & history of the world*. Moreover, the terms do not always correspond to the default cluster. In some cases the mismatch was obvious, for instance, the noun phrase “newspapers and television” was associated with *richest country in the history & history of the world.* In some other cases, it was necessary to double-check the content of the posts using the corpus explorer to see how the terms are used in context. Also, some terms such as “next year” or “new report” were too general to be related to specific clusters and we deleted them manually. Based on these findings, we created a more refined list of topics.

We identified the following topics:

1. Trump administration
2. Healthcare
3. Education
4. CC [climate change] and sustainable energy
5. Corporate welfare (e.g. terms such as “tax breaks” or “tax cuts”)
6. Minimum wage
7. Socio-economic justice
8. Criminal justice
9. Gun control
10. Supreme Court [Supreme Court’s decisions]
11. Support for candidates [Iowa and Illinois elections]
12. Events [at which Sanders appears]
13. Negative comparison [Sanders often compares the U.S. with other countries to highlight the deficits of the U.S.]

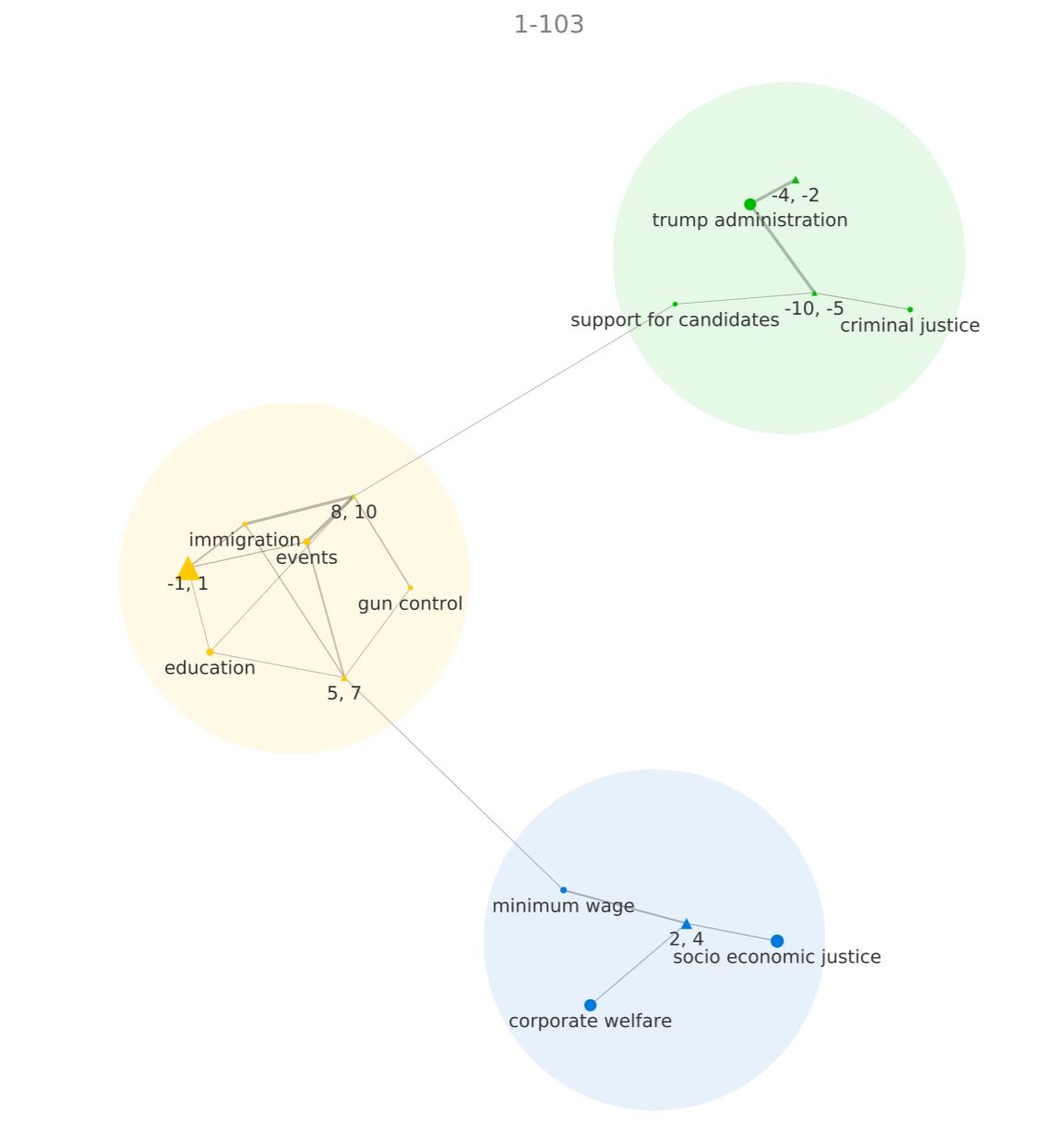
We uploaded the customized list as a terms list which we indexed to the 324 posts.

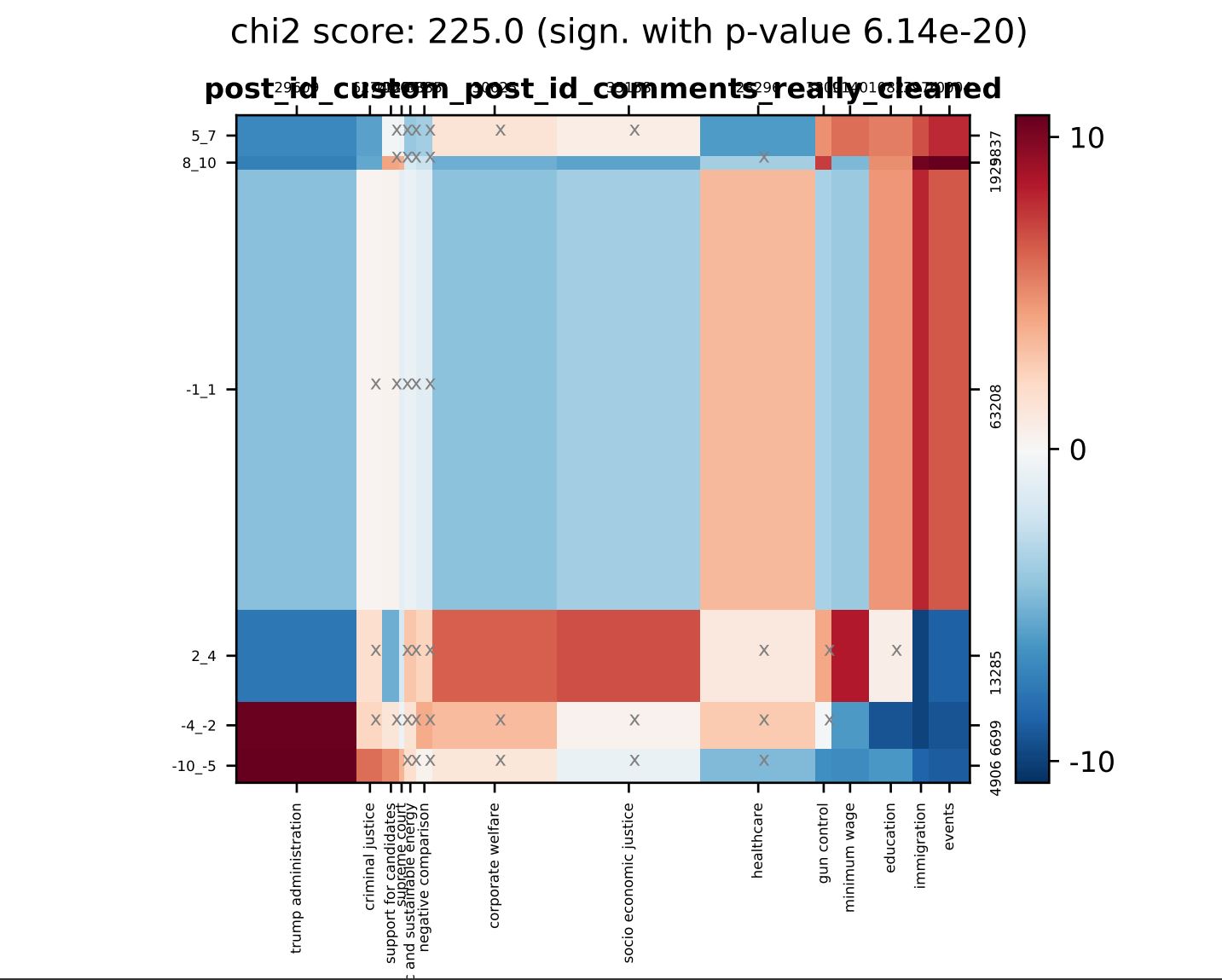
Not all of the posts had an assigned topic. In total 59 posts were not labelled. We created a new terms list that did not include posts that had not be tagged to any of the topics.

To link the posts and their topics to the respective comments, we built another corpus list which we indexed to the comments.

Meanwhile, we ran a sentiment analysis of all the comments. We decided to use Textblob and to focus on sentiment polarity which varies between -10 (very negative) and +10 (very positive). We sliced the data into the following six bins for better visualization: -10 to -5, -4 to -2, -1 to 1, 2 to 4, 5 to 7, and 8 to 10.

After having a first look at our data (showing post\_id, post message, comment message, associated topic and associated sentiment polarity), using corpus explorer, we created a semantic network and a contingency matrix for visualization:





To have an in-depth analysis of these results, we created sub-corpuses for each cluster by running queries. Using the corpus explorer, we could do a qualitative analysis of the most interesting boxes from the contingency matrix.