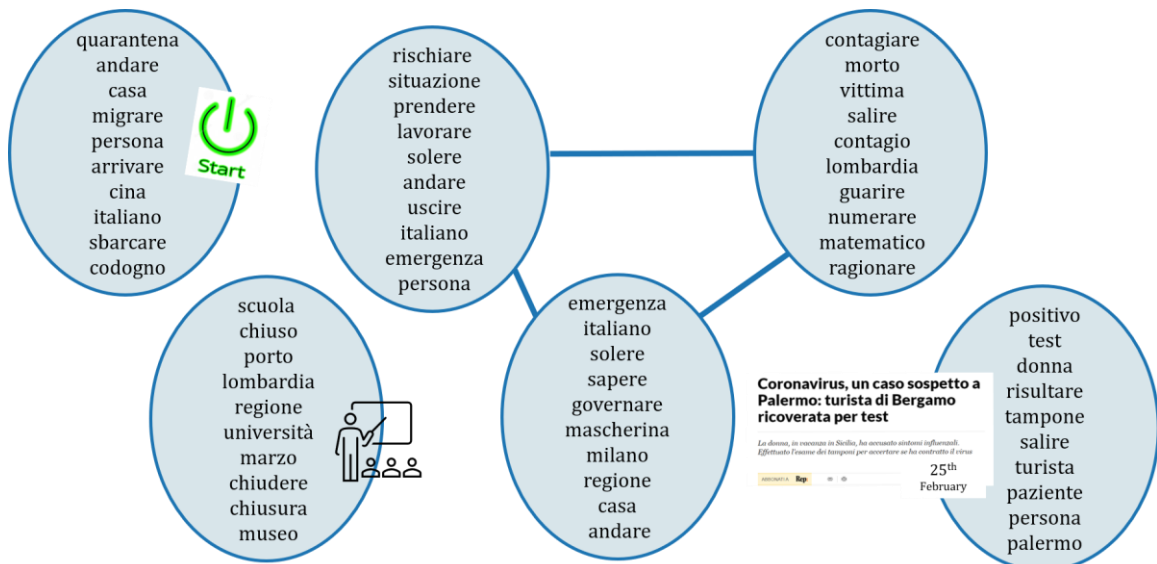


## k-means ++ clusters

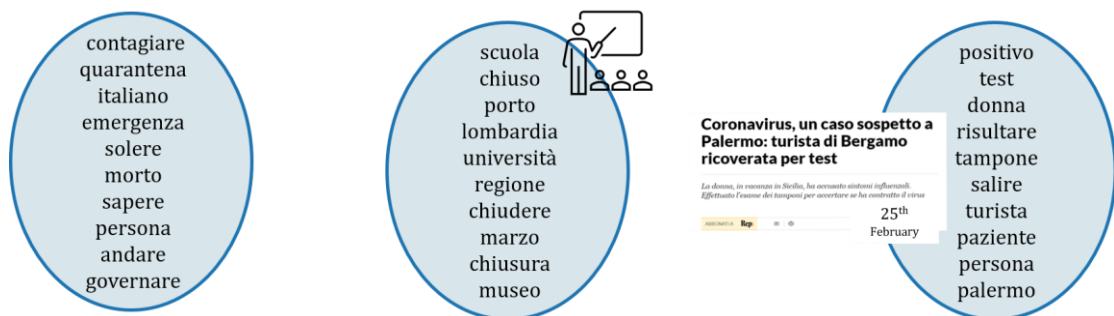
### First week

The first week goes from **Sunday 23 February - Sunday 1 March**, so at the begin of the emergency. Looking at the coherence score graph, 6 has been chosen as the number of clusters. As you can see from the follow figure for some clusters there is a clear topic.

- There is a cluster with words *codogno*, *cina*, *sbarcare* that can be interpreted as the cluster of the start of the pandemic since the virus arrived from Cina and the first Italy's hotbed was Codogno.
- Another cluster clearly refers to the news about an infected woman from Bergamo that goes on holidays in Sicily.
- Then a cluster concerning the closing of universities, schools and museums has been identified.
- Finally, there are some clusters that contain words about the "general" emergency.



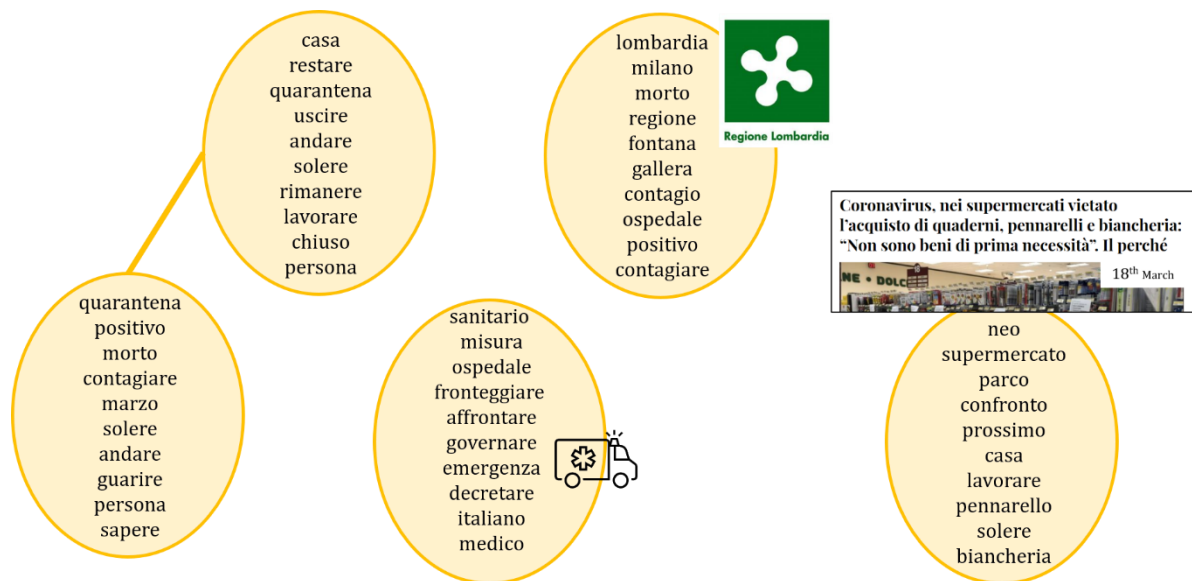
Considering the last observation, I have tried to reduce the number of clusters, to see if it was possible to merge the 3 similar ones. Surprisingly, this happened, as you can see from the following figure.



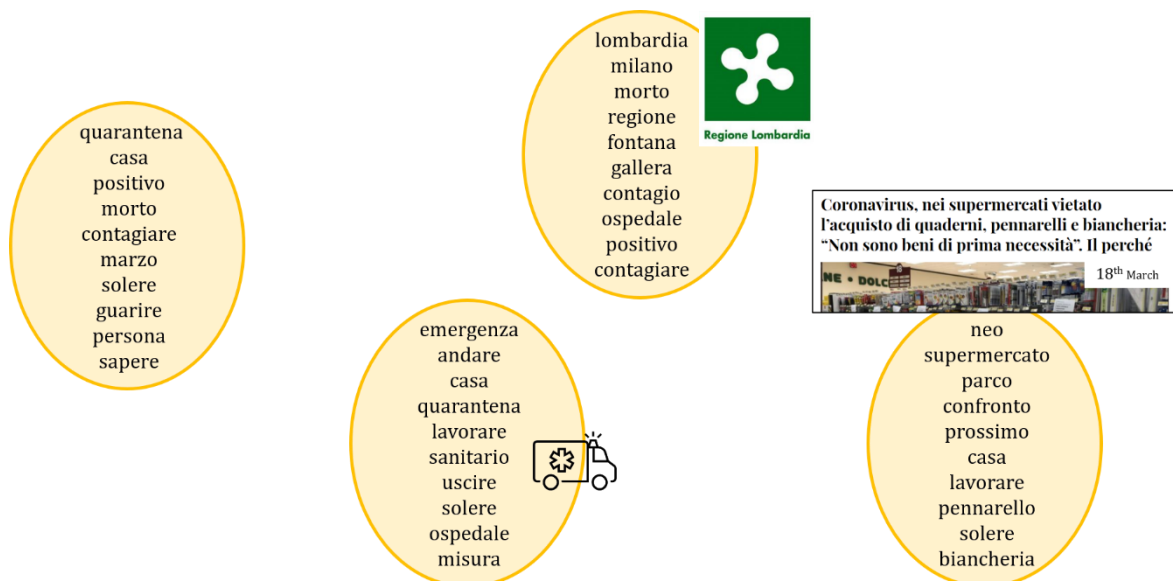
## Second week

The second week goes **from Sunday 15 March - Sunday 22 March** during the lockdown and include Easter holidays. Looking at the coherence score graph, **5** has been chosen as the number of clusters and clear topic can be identified:

- One cluster is about Lombardia, with Fontana and Gallera as protagonists
- Another cluster is about the law decree of 18<sup>th</sup> of March that forbids to buy exercise books, pens, markers, and clothes at the supermarket, since they were not considered necessities goods.
- A cluster about the hospital emergency has been found.
- Then, as before, there are 2 cluster concerning the Covid emergency in “general”.

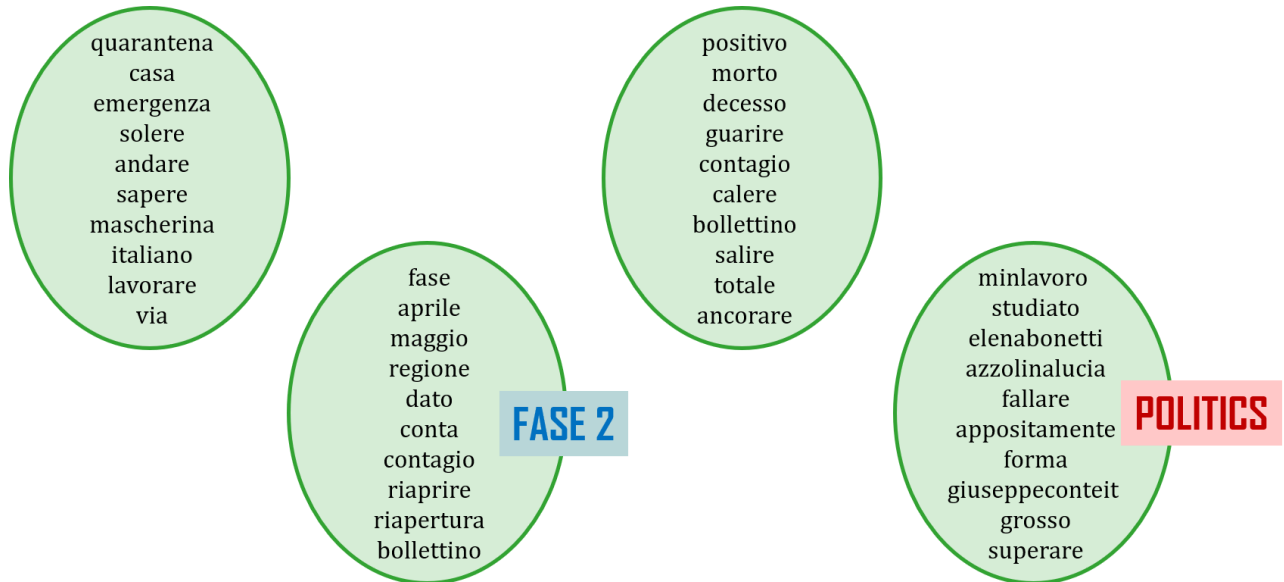


As before I have tried to reduce the number of cluster and as before it worked.



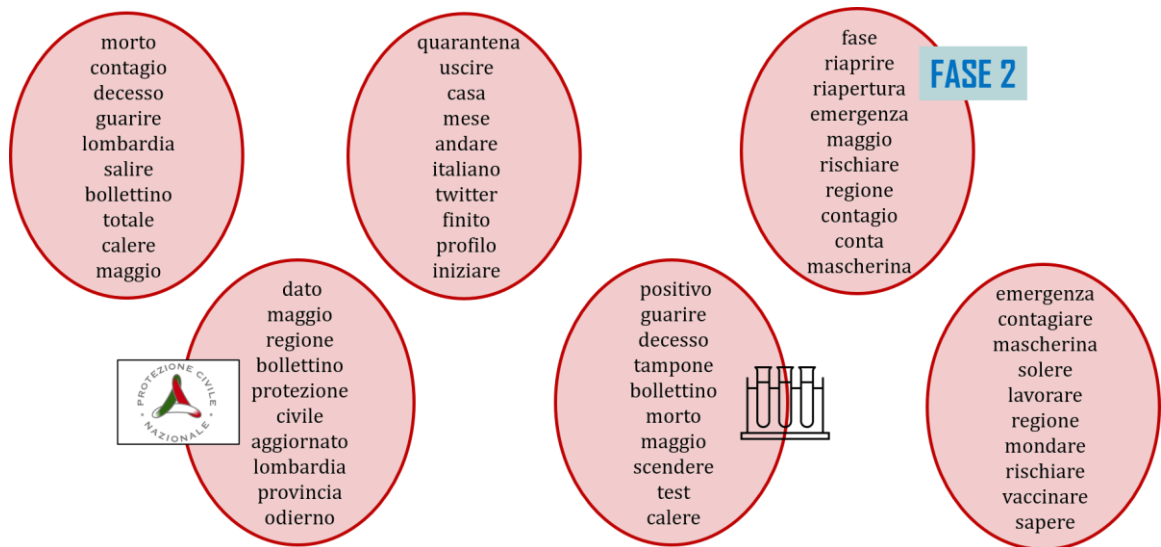
### Third week

The third week goes from **Sunday 19 April - Sunday 26 April**, so after more than 1 month of lockdown and near to the begin of Phase 2, with less restrictions. A cluster concerning the politic with Conte, Azzolina and Bonetti as protagonists, can be identified. Another cluster with terms concerning the re-open and thus the Phase 2 start is of course present (since at the end of April everyone wanted to get out of home). Finally, there are two clusters with “general” terms regarding the emergency with words like *positivo*, *morto*, *decesso*.



## Fourth week

The fourth week goes from **Sunday 17 May - Sunday 24 May**, so when the Phase 2 started and shops, restaurants, and activities re-open. For this week cluster about the Phase 2, one about the Protezione Civile daily updates and one about the test have been identified.



In one cluster appears the word “twitter” and I was curious to find out what it was about. Thus, I increased the number of clusters up to 9 and other interesting clusters have been created. One about the fake news on Twitter and another one about the real estate market.

