Social Media Analytics

Fabrizio Cominetti, 882737

Ruben Agazzi, 844736

Analysis of Italian Tweets Smart Working

The topic related to smart working has taken a prominent place in work-related discussions, especially following the latest changes on the issue from the government. So, we decided to analyze tweets about *smart working* in Italy.

Data Collection

Using the API provided by Twitter and the Python Tweepy library, we performed a download of Tweets containing one or more of the following keywords:

- smartworking;
- remotework;
- lavoroagile;
- smart working;
- remote work;
- lavoro agile.

Data Pre-Processing

- ✓ No Duplicates
- ✓ Change Date format
- ✓ Remove Numbers
- ✓ Lower Text
- ✓ Lemmatization
- ✓ Remove Punctuation
- ✓ StopWords Deletion
- ✓ Tokenization

Social Network Analysis

```python
for index, row in author_mentions.iterrows():
  mentions_array = json.loads(row['mentions'])
  for item in mentions_array:
    Graph.add_edge(row['author'], item['username'])
```

Graph Creation

In order to create the graph of the interactions, we took for each tweet created by a user its mentioned users, these last will be nodes of the graph.

For every user mentioned inside a tweet we created an edge between the two nodes.

For graph creation and graph analysis we used the python library NetworkX.

Social Network Analysis

- Nodes Degree
- Average Degree
- Assortativity
- Community Detection

Nodes Degree and Average Degree

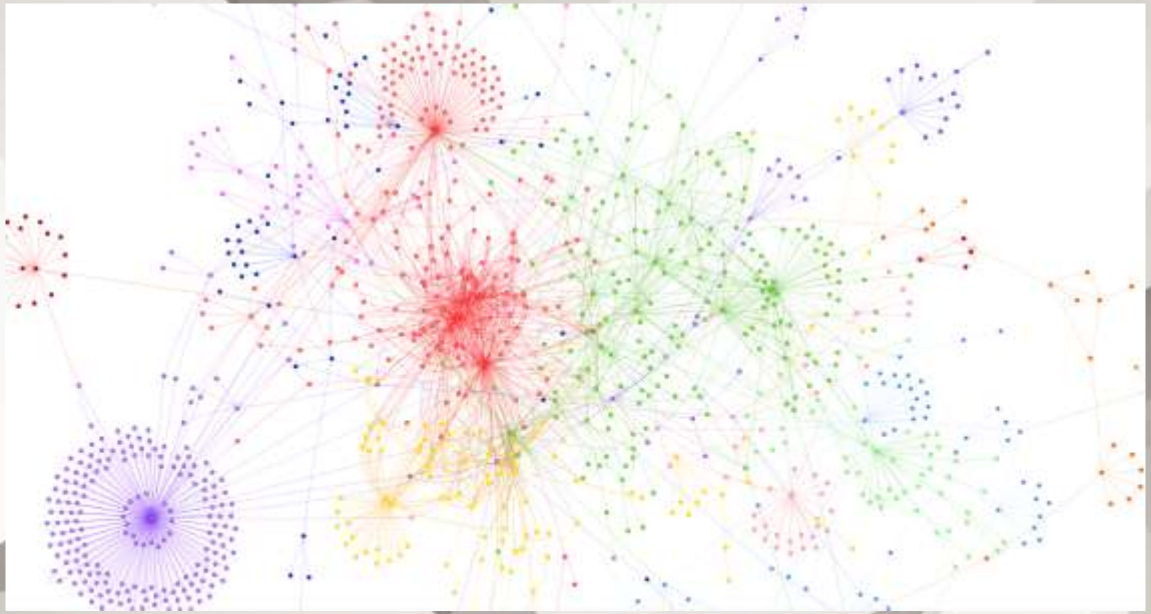
Node name	Degree
BonomiAllegra	210
mrmcphisto	145
GiorgiaMeloni	141
AlbertoBagnai	118
ClaudioDurigon	82

For each node, we calculated its degree (the number of incident edges that the node have). After that, we calculated the average degree between all nodes, which is equal to 3.704.

Community number	Number of nodes
Community 0	234
Community 1	210
Community 2	179
Community 3	139
Community 4	87

Community Detection

For the community detection part, we used the library NetwokX with *modularity maximization* in order to find the communities. The algorithm found 197 different communities, where the bigger community contains 234 different nodes.



Assortativity

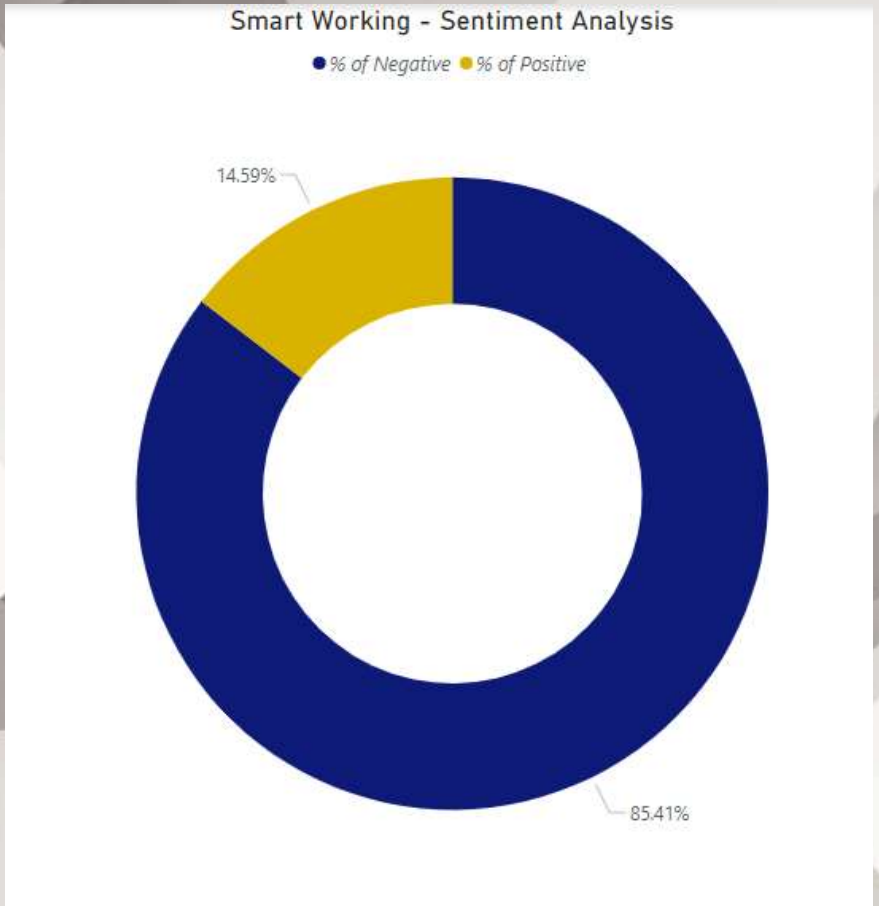
We also calculated the assortativity of the interaction network. We found an assortativity coefficient equal to -0.244, which makes our network *disassortative*.

In this case, smaller nodes tends to connect to bigger nodes in terms of degree.

Social Content Analysis

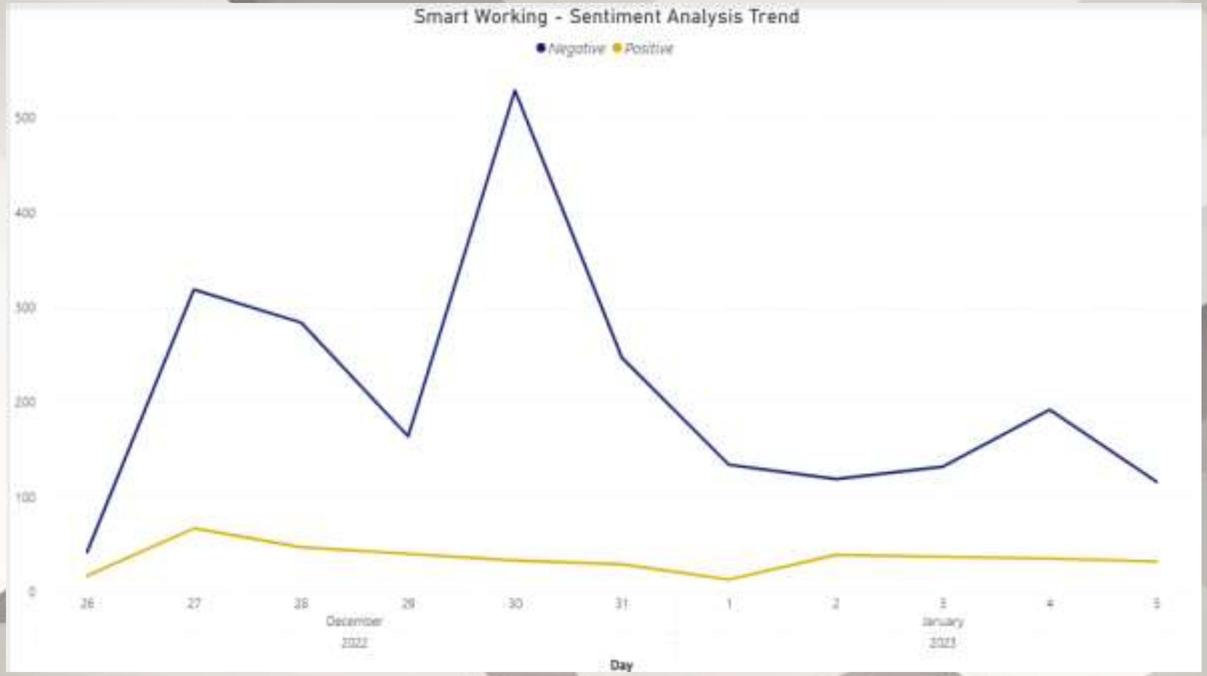
Social Content Analysis

- Sentiment Analysis
- Emotion Recognition



Sentiment Analysis

With the use of FEEL-IT, based on (Um)BERT(o), we calculated the sentiment score of the tweets in the dataset. As we can observe, the great majority of tweets was labeled as ‘negative’.

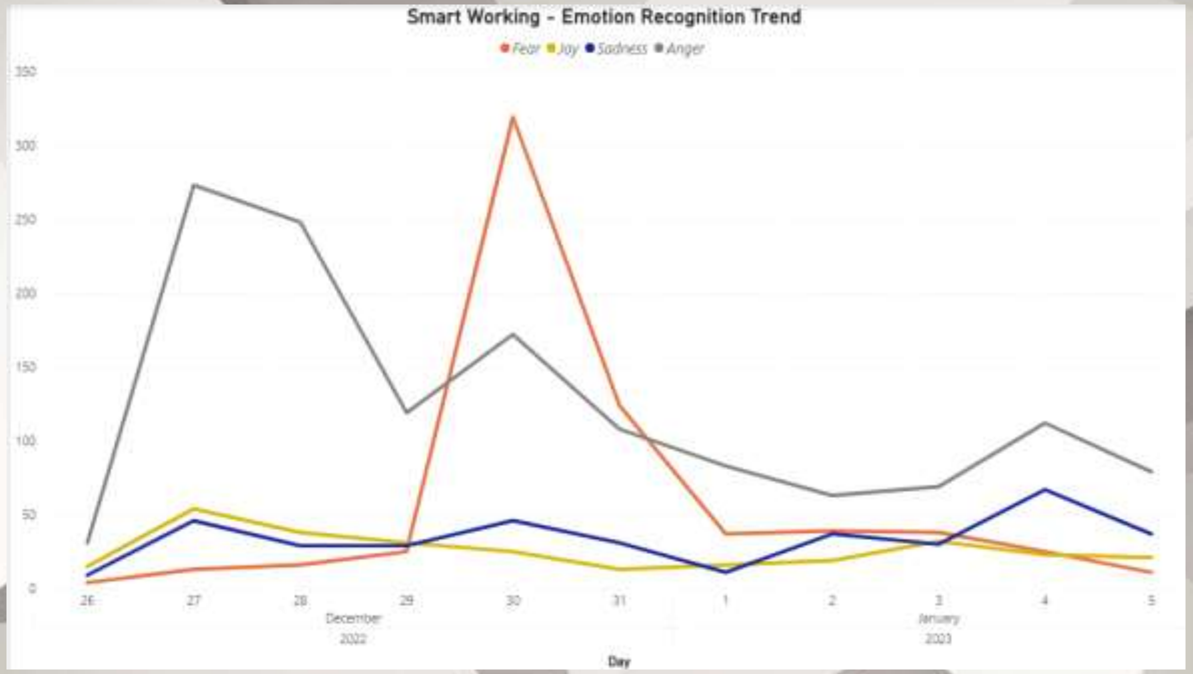


Sentiment Analysis

On the left, we can observe the time series related to the sentiment of the tweets considered.

30/12 was the day when a large majority of tweets were made, and also the day with a high level of negative sentiment.

Emotion Recognition



Through FEEL-IT, we identified the emotions present in the various tweets: anger, fear, joy, and sadness.

As observable, the emotion *anger* is the most prevalent, while the emotion *fear* has a peak on the day of 30/12.

Findings

With this social network analysis, we obtained some interesting information from both the social network analysis and from the sentiment analysis tasks:

- We found out that even with a relatively small number of tweets, users create large interaction networks and communities
- We found out that the relevant emotion about the topic of smart working, in the recent past, is mostly "anger", probably due to the limitations that were announced.