POLITOK

How do Italian politicians use TikTok as tool to promote their political ideas and influence the young generation during the 2022 elections?

Network Science Students:

Colussi Davide, De Masi Michele, Elfes Jan, Grilli Alessandro, Sbarbati Riccardo

Social Network Analysis Students:

De Angelis Martina, Hepaj Anjeza, Jaja Raghda, Pali Klarissa, Valenza Jessica

Abstract

A hot topic during the weeks prior to the last national elections in Italy, September 25 2022, was the advent of some of the most important figures in the political panorama on the social media platform TikTok. The peculiarity of this event lies in the fact that, compared to the other social networks, TikTok has always been mostly populated by the younger generations, and only in recent times, as its popularity is rising, it has gained the interest of a wider audience. Its natural format based on short videos, generally matched with music, may offer a good opportunity to send compact and focused messages to the public, and it may seem that many politicians have tried to exploit it for their political campaign. In this work we explore this possibility, analyzing materials posted in the weeks before September 25 by both politicians and influencers, to look for similarities between the way they present themselves and their ideas. We analyse if politicians are trying to emulate other influencers, and if this choice was mostly a tactic aimed at the short-term electoral campaign or if politicians decided to settle on TikTok and use it as a new way of conveying to young people. The analysis is based on methods from Network Science and Natural Language Processing. Using networks allows us to showcase the complex aspects hidden behind the interpretation of speech behaviors. As an additional comparison, the data has been interpreted with a software for sentiment analysis, to gain additional insight w.r.t. ana anlysis purely based on semantic content. The main method of comparison between all the data representations developed has been the Pearson correlation parameter, applied to embeddings extracted from different collections of either posts or comments. The final results obtained across all the different methods were not very robust, but they provided interesting first insights into the different approaches. The correlations obtained were in some cases in line with what was expected and with the final outcome of the election, underlying some clear similarities between the material posted by the politicians and the influencers observed, especially when the analysis was focused on some of the preferred topics brought by the politicians and their parties. Finally, the sentiment analysis on both the posts and the comments managed to highlight many of the expected characteristics of the individuals in exam and the reactions of the public that follows them.

1. Introduction

The global use of social media platforms has fundamentally transformed contemporary politics. Discussions and election campaigns are increasingly held on social media [1].

Traditional media such as television broadcasting or newspaper has been the predominant source of information during the political campaigns, although, the recent years and the wide use of the new media has brought the candidates to opt upon new communication strategies for the respective campaigns. Scholars predicted an increased and targeted web use by political campaigns, this includes use of Social Network Services (SNSs), which allow candidates to build profiles and showcase connections within a delimited system[2].

Social media has radically changed the connection between politicians and voters. In 2008 former president of the United States of America, Barack Obama, for the first time during a political campaign used digital communication; the wise and coordinated use of social media showed that these strategies could directly affect the mobilization of the voters. From that moment the use of social media during the elections became crucial to the campaign. In Italy, Movimento 5 Stelle and Lega started to use the digital communication following the wave of Trump after Obama[3]. It started with the use of Facebook, then Twitter where 80% of the most important politicians worldwide

had a profile; in the last five years Instagram became crucial for the communication and the past two years TikTok got the attention of the politicians.
During the pandemic of Covid-19 politicians started to use social media with consistency and gained popularity due to the digitalization of everyday life. During the last elections in Italy, politicians started to use strategically every social media. TikTok, the cyber space of the younger generations, started to welcome politicians, even the oldest ones like Silvio Berlusconi[4]. The intention is to raise awareness among the young potential voters on the importance of the voting act, although many international media read the subscription on TikTok as a strategy to gain more votes omitting the importance of getting informed before making the decision. In this paper we want to analyze how the politicians used this new platform and if it influenced the mobilization of the young voters.

According to statistics from 2022, the majority of internet users in Italy use YouTube (88%), followed by Meta platforms (64%) and TikTok (54%), while only 7% use Twitter [5].
As a team, we choose TikTok because it played an important role in influencing the last Italian elections since most politicians used it to deliver their political camping and message. Furthermore, this platform is a useful tool to conduct a well-constructed social network analysis.
Social media has developed into a powerful tool for marketers to communicate with and advertise to consumers. More crucially, marketers use social media to increase brand loyalty and recognition [6]. Although numerous studies have looked at the purposes and pleasures of social media use, very little study has looked specifically at the pleasures of TikTok, a rapidly expanding social media platform and mobile video app known in mainland China as Douyin. With 400 million users in China and more than 800 million active users worldwide, TikTok has emerged as one of the most widely used social media sites [6]. TikTok was named marketer of the year by the American publication Advertising Age in 2020. [6] The only social networking services in the world that are accessible in both China and the United States and are extremely well-liked on a worldwide scale are TikTok and Douyin. (As of now, India is the only nation that has outlawed TikTok.)
China is the main market for firms operating internationally due to its fast economic expansion and vast population. However, due to the severe regulation of international social networking sites, China's use of social media differs from that of the rest of the world. The video-sharing website TikTok/Douyin is distinctive. The majority of videos are under 60 seconds long, and choosing background music from its music library is an easy process for video creators, making the videos more enjoyable [6]. Users can record numerous live events or create product testimonial videos. They are urged to copy others and create videos that are similar to memes or trends. Videos from TikTok can be easily shared on other social networking and messaging sites. Following a user's selection of a video to view, the platform will play additional films depending on complex algorithms. This keeps consumers interested and frequently causes them to lose track of time. The "For You" function provides a stream of videos that are continuously tailored to users' preferences depending on their prior viewing [6]. The TikTok app is free to download on both the Apple and Android app stores, and it's a start-creating tab that includes simple video editing features that anyone who has little to no technical expertise can use to make videos. This makes it very simple for any user to use the app's link capacity features. It is socially acceptable to participate as an audience member, a performer, or both, and one can decide how much interaction with other members they want to have.
Recently, TikTok has drawn notice in pieces in the general press on how teenagers are using SNSs to express their ideas on social topics that are important to teenagers [7]. Hence, the TikTok app makes use of a range of "socio-technical elements" that let young teenagers build and maintain a cohesive network of connections that facilitate the emergence of activist communities [7]. The TikTok app has a wide range of connection options, and they stand out from other SNSs in that the communication is based on young teens' shared interests and is presented in a video format, allowing them to express themselves without worrying about connecting with friends or friends of friends through stories.

2. Research Questions

The interest in how the Italian politicians communicated to young generation people before the election periods is what led us to a deeper analysis developing our research question, which is: "How did Italian politicians use TikTok as tool to promote their political ideas and influence the young generation during the 2022 elections?".
This curiosity came up to our minds because of the hype the politicians had in a specific social network our generation commonly used, or rather, "TikTok", indeed, it is not surprising to find that political parties and leaders are re-assessing the opportunities TikTok can offer to shape people's views and opinions as they reach adulthood and start voting, or even before then. Talking to young audiences has posed a challenge to political leaders, and yet adolescence is the time when individuals firm up ideas about their identities and start developing their opinions about society.[8]
We could have never expected such a thing and the unexpected made us questioning several curiosities, such as: What the politicians did has to be considered a causality or a strategy? If they strategically choose to communicate

through this social network to the young generation, has it worked efficacy or not? What kind of communication did they use for the video as a winning strategy?
In our research questions we want to deeper analyze the politicians belonging to the right party, who specifically made this Social Network as the most important tool for their pre-election communication understanding better how the young generation have been considering them.
For a more complex analysis we will have a focus more on the Community Detection aspects which may have an important function of control and stability within the group; vertices lying at the boundaries between modules play an important role of mediation and lead the relationships and exchanges between different communities [9],and on the eventual Hemophilia's cases: that is all of them were conducted in a physical world scenario by surveying a group of human subjects.
Their ties were subject to social influence [10] and to polarizations between the subjects of the target chosen, for this reason we will need: comments, video-answers and duets on TikTok. In the following section we will specify the different types of analysis we used.

2.1. Type of analysis

We structured our analysis in the following way:

- Firstly, by finding similarities on the communications used by the politicians according to the different algorithms and topic founds.
- Secondly, we selected seven main topics: economy, environment, young, rights, politics, elections, cultural educations to be able to select the word's categorization.
- Thirdly, to build semantic network analysis we built adjency matrix with words and people and with politicians and influencers to create bipartite graphs.
- Sentiment analysis: analyzing words from the comments and video texts to understand how the audience react, if in a positive or negative ways.

2.2. Sub-Network comparison

We decided to compare the videos and the comments separately because the network used on the videos were made to compare the type of communication politicians used on the platform and the comments, instead, were used only to make an average on how most of the people were reacting to those videos analyzing if those comments are positive, negative, or ironic.

2.3. Sentiment analysis

LIWC: it is a tool used for the sentiment analysis in which the words can be associated to classes and those classes can be associated to words that belongs to emotions. This sentiment analysis will be done to see users' reactions. It will be explained in details in sec. 9.1.

3. Database

This section describes procedures carried out to gather the videos from both politicians and internet figures on TikTok. We first defined a time frame that goes from 21st July 2022, the day the President of the Republic, Sergio Mattarella, announced the dissolution of the chambers [1], to the 24th September 2022, the day before the elections. Then we created a list of the leading politicians running for the position of Prime Minister.

- Giorgia Meloni (https://www.tiktok.com/@giorgiameloni_ufficiale?lang=it-IT)
- Matteo Salvini (https://www.tiktok.com/@matteosalviniufficiale?lang=it-IT)
- Silvio Berlusconi (https://www.tiktok.com/@silvio.berlusconi?lang=it-IT)
- Giuseppe Conte (https://www.tiktok.com/@giuseppeconte_it?lang=it-IT)
- Matteo Renzi (https://www.tiktok.com/@matteorenziufficiale?lang=it-IT)
- Carlo Calenda (https://www.tiktok.com/@carlocalendaofficial?lang=it-IT)
- Enrico Letta (https://www.tiktok.com/@partitodemocratico?lang=it-IT)

We proceeded to look through the electoral programs of each politician to select the themes of interest to young adults: among these *school*, *work*, *environment*, *human rights*, *economy*, *young generations*, and *safety*, were detected. These key-themes became keywords, that paired with the hashtags *#elezioni* and *#elezioni2022*, would allow us to conduct the research on influencer's videos. Thus began the research of the videos on TikTok and the consequent categorization on Excel[2].

[1] https://www.quirinale.it/elementi/70558
[2] https://docs.google.com/spreadsheets/d/1HNwzxmCR_oE3o6AbSpaZvHe-b9DCk8QTMcqdD2431C8/edit#gid=0

3.1. Politicians

We first began our research sorting through the politicians' profiles. The part of video selection was conducted manually: we carefully viewed and selected videos of all chosen politicians that matched the identified topics, in the specific time frame of the election campaign period. The videos were gathered from the official TikTok profile of said politicians. Every politician presented a personal profile, with the exception of Enrico Letta, who was represented by the *@partitodemocratico* profile, that is, that of the party to which he belongs; however, the profile of the party was taken into consideration in order to have a greater representation of the left coalition.
In the appendix B is shown a table 6, which is referred to the used Politicians.

3.2. Influencers

As previously said, to select videos from influencers we first identified some keywords that matched the key-themes; these words would be researched on the "Research" section of TikTok along with the hashtags #elezioni and #elezioni2022. The keywords in question are:

- **School:** scuola, istruzione, educazione
- **Work:** lavoro, salario, stipendio
- **Environment:** crisi climatica, clima, inquinamento, energia
- **Human Rights:** diritti, LGBTQ+, immigrazione, salute
- **Economy:** economia, crisi, tasse, RDC
- **Young Generations:** giovani, futuro
- **Safety:** sicurezza, criminalità

Through this research we were able to collect 71 videos, from a wide variety of accounts, which include *influencers*, *micro-influencers*, *info pages*, *celebrities/public figures* and *personal accounts*.[3]
In the appendix B is shown a table 7, which is referred to the used Influencers.

3.3. Database

We selected 171 videos in total, 100 from *politicians* and 71 from *influencers/public figures/ info pages*. Once the selection of the videos was concluded, we were able to identify micro-themes that would be associated with the initial macro-themes, as it follows:

- **School / Education** (compulsory schooling, school investments, education level, knowledge and culture, from school to workplace, university acceptance, closed number access to university)
- **Work / Jobs** (RDC [4] , from school to workplace, quote rosa, minimum wage, dismissal, worker protections)
- **Environment** (climate change, renewable energy, sea pollution, nuclear energy, gas)
- **Human rights** (LGBTQ+ rights, healthcare, euthanasia, abortion right, immigration, freedom of speech)
- **Elections and the right to vote** (*fuorisede*[5] voting, voting abroad, how to vote, right to vote at 16 years old, electoral programs)
- **Economy** (bills, taxes, poverty, *flat tax*, *RDC*, *patrimoniale*[6], speculation, IVA[7] , PNRR[8])
- **Young generations** (jobs, military, future, 18App[9], schooling, culture, education)
- **Safety** (criminality, hate crimes, rape, drugs)

These micro-themes were then labeled under narrow categories, named *Topic_Label*, that allowed the conduction of the search in a more selective and orderly manner. The database follows the FAIR principles[10]: all the videos selected are *Findable*, *Accessible* and *Reusable* as they are all public in the TikTok platform; the data is **Interoperable** as well, since the data adopts a standardized language for knowledge representation that can be easily understood, shared and widely applicable.

[3] We have identified as *influencers*, accounts with over 100K followers; as *micro-influencers*, accounts with more that 1000 followers but less than 100K; as *personal profile*, accounts with less than 1000 followers. The categorization *info page* refers to those accounts which purpose is information on current affairs

[4] RDC (Reddito di Cittadinanza) – Citizen's Income in English – is a measure to combat poverty that came into effect from 6th March 2019

[5] The term fuorisede refers to all those people – in particular students and workers – who have moved from their hometowns to peruse their careers.

[6] Patrimoniale is a tax that affects the wealth (patrimony) of a subject. It is a tax that does not cover the income received by the taxpayer but focuses on the capital held, both in Italy and abroad. (Source: https://fiscomania.com/imposta-patrimoniale/)

[7] IVA (Imposta sul valore aggiunto)– VAT in English - is the Value added tax, that is the consumption tax that is applied to goods and services in Italy (Source: https://www.agenziaentrate.gov.it/portale/web/english/nse/business/vat-in-italy)

[8] PNRR (Piano Nazionale di Ripresa e Resilienza) – National Recovery and Resilience Plan In English – is the program for economic recovery after the Covid-19 pandemic crisis (Source:https://www.mef.gov.it/en/focus/The-National-Recovery-and-Resilience-Plan-NRRP/)

[9] 18app is a program provided for by law through which 500 euros are awarded to all eighteen-year-olds (Source: https://medium.com/team-per-la-trasformazione-digitale/18app-cultural-bonus-eighteen-year-olds-spid-digital-identity-service-bug-67e084077b58)

[10] https://library.cumc.columbia.edu/insight/what-are-fair-data-principles

4. Data collection

The extraction of the data is done through two different APIs, both from third-parties, i.e. they are not official TikTok software.
Starting from the database, built as explained in section 3, it is possible to extract the video ID from the url address. This is done implementing a PYTHON script. An example of url address for a TikTok video is: WWW.TIKTOK.COM/@USERNAME/VIDEO/1234567890123456789.
In the next two sub-sections,4.1 and 4.2, we explain how to download the data using the APIs.

4.1. Videos - TikTokAPI

All the documentation of how this API works can be found in [11].
This software is used to download the videos, and two python libraries need to be downloaded: TIKTOKAPI and PLAYWRIGHT. It can be done through pip, conda or anaconda software. Some other libraries may be required for the correct working of the PLAYWRIGHT library.
The first thing that one needs is the **custom_verify_fp**, a TikTok parameter used to get access, which can be found in the cookies folder of the browser, usually under the name *s_v_web_id*.
Having all the necessary libraries and the key it is possible to proceed to the video acquisition.
Through a implemented python script, which takes as input the database and the **custom_verify_fp**, it is possible to extract the text from the videos in four main steps:

- **Video ID:** from the database the video ID is extracted from the url address;
- **TikTokApi:** for each ID the software, which requires the custom key, will download the video in the *.mp4* format;
- **Video to Audio:** using the MOVIEPY library all the downloaded videos are converted into audio files, in the *.wav* format (this format is required for the next step to extract the text);
- **Text Extraction:** through the **speech_recognition** library the text for each file is extracted and saved into an appropriate file. Since for some files the audio track contained too much noise, this library did not work, an external software was used to extract the text.

In the appendix C.1 there is a user guide explaining how to use this software. As the API was not able to access comments we needed to proceed with another second API.

4.2. Comments - TikAPI

All the documentation of how this API works can be found in [12].
This software, unlike the other, is proprietary. The use of it requires two different keys: **TikApiKey** needed to connect the software and **AccountKey** which connects a TikTok account to the API. This software gives the possibility to get access to different metadata and data. In our work it was used to get the comments. Through an implemented python function one can extract the comments of a video using the the video ID. The output of this function is a nested dictionary, *i.e.* a dictionary containing sub-dictionaries, which is saved into a *.json* file.
Inside this file there are different keys, the most relevant ones are: ”cursor” which gives the number of downloaded comments and ”comments” which gives all the metadata of the comments. The latter it is composed by sub-keys, the more important ones are ”text”, ”comment_language” and others.
Finally, a file for each video, which contains all its comments, is created. In the appendix C.2 there is a user guide explaining in more detail how to use this software.

4.3. Topics

For additional analysis, each video has been assigned to a topic. This was done by hand, looking at the videos and selecting the topics they covered. The main topic has been chosen for each video.
Unlike the general topics, as shown in 3.3, these ones are used to divide the data into different classes to be applied to different algorithms.
Table 1 lists the topics and number of videos belonging to each one.

Topic List		
Topics	Politicians	Influencers
Environment	9	14
Rights	7	16
Economy	24	5
Elections	23	15
Young	11	4
Politics	17	7
Culture/Education	10	10

Table 1. Number of politician and influencer videos for each topic

5. Language Processing

For processing of the text we use *SpaCy*[13]. It relies on pre-trained language processing pipelines that are available in different languages. We use IT_CORE_NEWS_SM which is the smallest of the three available piplines for Italian. It was trained on news and media outlets and consists of a tokenizer, morphologizer, tagger, parser, lemmatizer, attribute ruler and named entity recognizer. To build the semantic networks, we rely on the output of the lemmatizer. Further, we added the emoji module to detect emojis and used the tags from the tagger to filter punctuation.

As the heavier large language models have particular needs, we adapted the processing accordingly. *LIWC* had problems with the mixture of Italian and English words in the comments. Therefore, we needed to translate all comments to English. Both, *BERT* and *BERTopic* perform their own text processing which interferes with the processing done by *SpaCy*. Therefore, these two models were applied to the raw text.

6. Exploratory data analysis

In this section we report first results on both videos and comments. This includes word count statistics, semantic networks, and topic analysis.

6.1. Videos

The first analysis is done on the text extracted from the videos, as described in section 4.1. The aim of this section is to find a good representation and a sensible way to compare the speech of different TikTok users, both from the political landscape and from the more ”standard” influencer.

6.1.1 Vocabulary

Each network is built by selecting a certain subset V of videos to analyse, and the first step is to build a vocabulary of key words from the *raw* documents in V.
After processing each document with the *NLP* pipeline, as described in section 5, the vocabulary is composed extracting the unique vector of the tokens occurring in the cleaned documents. Tokens appearing only one time throughout all the documents are filtered out as an attempt to remove some *noise*. Indeed, after this filtering, the 3767 tokens extracted from the NLP pipeline are reduced to 1555.

6.1.2 Semantic network

The first and most straightforward network is built by creating an adjacency matrix where the nodes are all the words in the documents. The edges are defined between words that appear in the same document, and are weighted by the number of documents they appear in together. The weight is normalized by the highest frequency value.
Running a centrality measure on this networks allows to highlight the most central and characterizing words for the chosen subset of documents V.

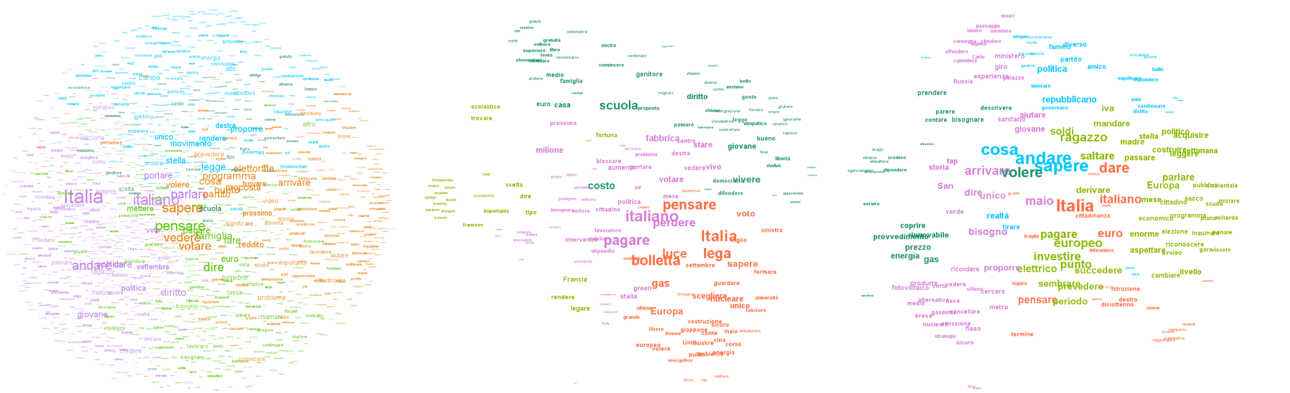
Some examples of this architecture are shown in figure 1:

- **1a** - adjacency network of all the available documents (170) in the dataset (1555 nodes, 184878 edges).
- **1b** - only documents of the 26 posts by *Salvini* (186 nodes, 4220 edges).
- **1c** - only documents of the 22 posts by *Calenda* (176 nodes, 4897 edges).

The size of the words depends on their centrality measure, obtained with *PageRank*, and to better visualize them community detection with modularity has been run, to highlight similar groups of words.
In every graph the words *Italia* and *Italiano* are always amongst the most important hubs, together with verbs that can be associated with political propositions and the election campaign, like *pensare*, *sapere*, *andare*.
Communities are more clearly visible in graphs of a single individual, and are generally related to words appearing in a single video, so that they are only closely linked to each other and to the main hubs. In the general network (1a) these spatial divisions disappear, but the coloring still shows four main communities. While containing many tokens not belonging to a specific semantic group, it is interesting to see how a close inspection of each community can identify groups of words that can characterize it in the following way:

- **Purple** ($\sim$33%) - political discussions and propositions.
- **Green** ($\sim$23%) - economics.
- **Orange** ($\sim$23%) - elections.
- **Light blue** ($\sim$21%) - healthcare/environment.

At this point it is possible to build a different adjacency network for every politician.
As a first result we can extract the most influential words ranked with PageRank from each of those networks, as one can see in table 2. From this first analysis one can already see a distinction between the topics which are present in politicians' videos. For instance Silvio Berlusconi focuses more on the elections, using words like 'Settembre' and 'Votare', mainly referring to young people with the word 'Giovane', considering that he joined TikTok (composed mainly of young people) right before the elections. On the other hand Matteo Salvini's main topic seems to be the economy, with a predominance of tax-related words like 'Bolletta', 'Pagare' and others like 'luce', 'costo' and 'gas', probably trying to address people who are vulnarable due to the high inflation. While right wing politicians tend to often cite the name of their corresponding parties and to use more modal verbs as 'potere' and 'dovere', the more moderated ones seems to call into question politicians of opposite parties, in particular Giorgia Meloni who had the majority in electoral voting forecasts.



(a) *All documents* (b) *Salvini* (c) *Calenda*

Figure 1.
Adjacency networks of posts from different subsets: all available documents (a), avg. degree 10.9; Salvini (b), avg. degree 9.0 and Calenda (c), avg. degree 12.7.

Most Influential Words	
Politicians	Top 5 words
Berlusconi	Settembre, Votare, Giovane, Dare, Prossimo
Calenda	Italia, Cosa, Sapere, Dare, Volere, Italiano
Conte	Dire, Meloni, Riguardare, Portare, Sapere
Meloni	Fratello, Vedere, Mettere, Credere, Diritto
Partito	Scuola, Parlare, Chiamare, Euro, Giorgia
Renzi	Prendere, Idea, Portare, Andare, Vedere
Salvini	Bolletta, Italia, Lega, Pagare, Pensare

Table 2. Most influential words for every Politicians, ranked with PageRank values

6.1.3 Degree Analysis

From the adjacency matrix we can conduct an analysis on the degree distribution.
We analyzed the degree distribution of the network composed by all politicians and the network composed by all influencers. First of all we can plot an histogram indicating the distribution of the degrees in each network, shown in figures 2 and 3.

We can also check if the network is scale free. To do this we have fitted the degree distribution to the power law: $p_k = Ck^{-\gamma}$, obtaining the values for k and γ showed in table 3. We can see that the networks are not scale-free, since γ is not between 2 and 3 for both networks. $kmin$ indicates the minimum value of degree k from which the fitting have been applied, since lower values of k do not follow the power law, as we can see in figures in the appendix A.1a and A.1b. A $kmax$ has not been set, so the fitting procedure was done from the degree $kmin$ to $+\infty$.

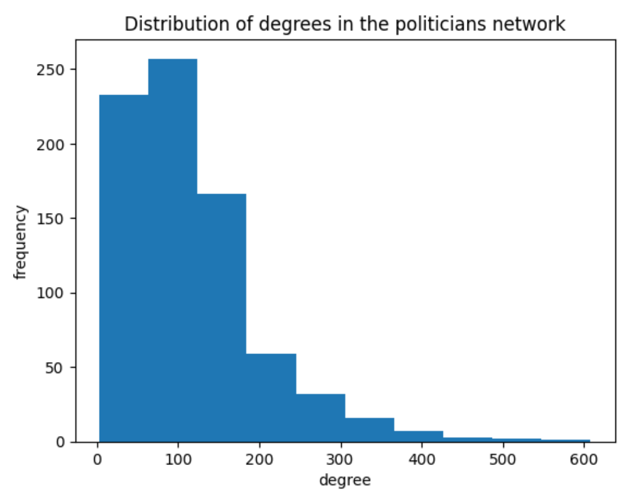


Figure 2. Distribution of node degrees in the politicians network

Network	$kmin$	γ
All politicians	126	3.51
All influencers	368	4.54

Table 3. Values found by fitting the power law to the degree distributions of the networks.

6.1.4 Assortativity

In this section we discuss the assortativity of the networks. The first network analysed is the one shown in figure 1a, with the words from all the documents in the dataset connected by their adjacency and ranked by their centrality. Both the degree correlation matrix in figure 4a and the assortativity plot (4b) show a disassortative trend, with a

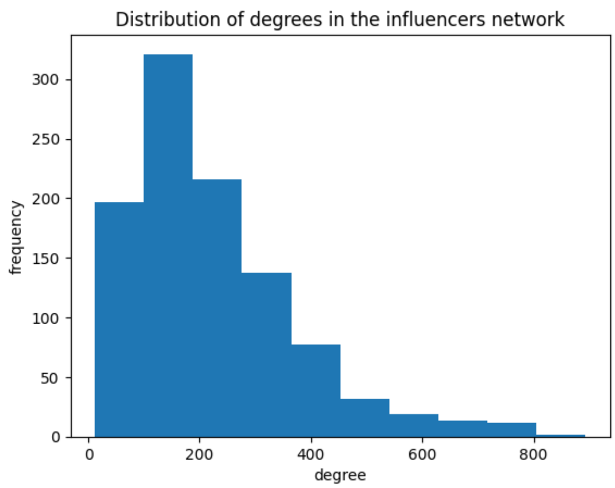


Figure 3. Distribution of node degrees in the influencers network

(dis)assortativity value $\mu_{all} = -0.078$.
However, this result may be due to structural reasons rather than being an intrinsic characteristic of the network, as suggested by the structural cutoff k_s. This parameter is defined as:

$$k_s = (< k > N)^{1/2} \tag{1}$$

with $< k >$ the average degree of the network and N the total number of nodes. It is considered a rough estimate of the threshold at which a random network cannot maintain its neutrality and its structure starts to become disassortative.
In figure 4b the structural cutoff is represented by the black dashed line, and since it is lower than the maximum degree k_{max} of the network it is possible to account the disassortative behaviour of the network to structural reasons. Similar results are obtained when measuring the assortativity of networks with only documents from politicians (A.2) or influencers (A.3), with assortativity values $\mu_{pol} = -0.026$ and $\mu_{inf} = -0.072$, and valus of k_s always lower than k_{max}.

6.2. Comments

In this section, we will report some general statistics on the comments from the TikTok videos. We collected a total amount of 52541 comments of which 38215 ($73\,\%$) were from videos by politicians. Videos by politicians tend to have more comments that are shorter and have a higher emoji density (number of emojis per number of words under a video) than videos by the influencers.

6.2.1 Wordclouds

To better understand, which are the topics discussed in the comments we analyse their content. A first glimpse is provided by figure 5 which shows the most frequent words for videos by politicians and influencers. The most frequent words for politicians include their names as well as ”votare”, ”potere” and ”dovere”. For the influencers, we find ”votare”, ”dovere”, ”potere”, ”volere” as well as names of politicians. This is to be expected since these terms can be easily associated with talk about voting and propositions for the upcoming elections, e.g. ”we will...”, ”we have to...”, ”I think that...” etc. Similarly to the semantic network for the videos, ”Italia” appears as one of the most common words, while ”giovane” is more prominent in the comment sections from the influencers.

6.2.2 BERTopic

To further investigate the topics discussed in the comments, we use BERTopic[14]. The model creates an embedding for each word in the corpus via SBERT, reduces the dimension via UMAP and clusters the results using HDBSCAN. We apply this procedure separately for both, the comments of politicians and the influencers. Figure 6 shows the main topics found. For politicians these include Italy (topic 0), greeting and cheering for the politician in the video (topics 2, 3 and 4), the government (topic 6) and politics (topic 7). For the influencers, we also find Italy (topic 1) the election (topic 2), taxes (topic 3), family (topic 4) and homophobia (topic 5). Further, we also find topic 6 (influencer) which seems to involve people expressing their agreement with the opinions stated in the video. On the other hand topic 5 (politicians) could involve mocking Berlusconi as it includes the term ”bunga”[11] as well as ”ridere”.

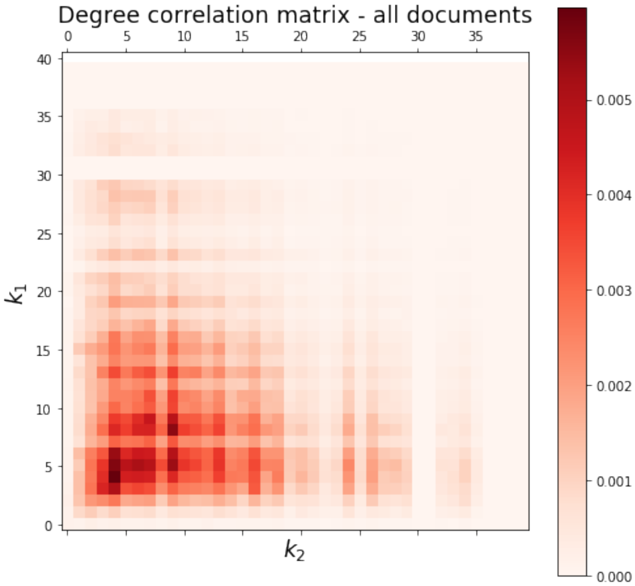
7. Similarity analysis

In this sections we analyse the similarity between different politicians and influencers. This is done by creating an embedding for each person using either their video transcripts or the comments. We apply different embedding technique to test their robustness. The final results, a comparison between politicians and influencers, are reported in table 4

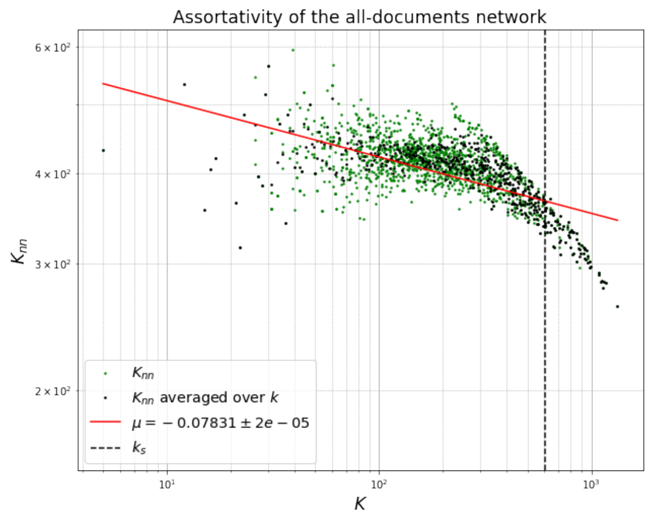
7.1. PageRank

Apart from the purposes explored in the previous section, semantic networks can be used to create an embedding of the data they represent. For this we build a vector with the dimension of our whole vocabulary. Each group of videos belonging to one person V can be *embedded* by the vector containing the centrality measure of each of the words appearing in the network. If a word from the total vocabulary does not appear in V, it will be considered as a node with centrality of 0. This interpretation is extremely useful when

[11] Reference to Silvio Berlusconis’ sexual escapades during his time as prime minister.



(a)



(b)

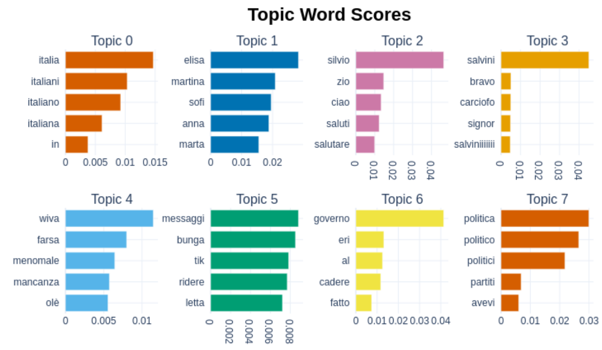
Figure 4.
(a): Degree correlation matrix of the all-documents adjacency network. (b): Assortativity of the all-documents adjacency network, showing the *average neighbor degree* K_{nn} and its average over k. The fitted red line visualizes the assortativity value μ, while the dashed black line represents the estimated degree of the structural cutoff k_s.

Figure 5. Wordcloud showing the most frequent words in the comments for politicians (left) and influencer (right). The size of the words is proportional to their frequency.

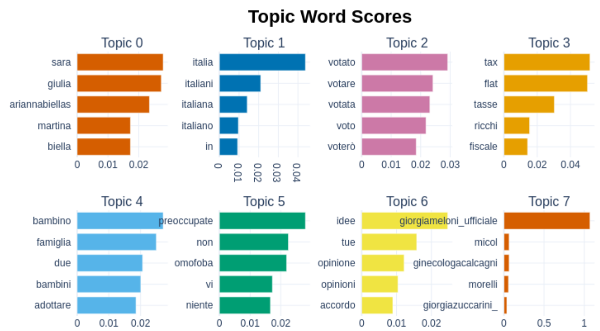
trying to compare speech profiles between certain groups, *i.e.* politicians and influencers. The similarity between two network embeddings X and Y can be measured by computing the *Pearson correlation coefficient*

$$\rho_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \qquad (2)$$

which quantifies the linear correlations between two sets of data. The similarity measure just described has been computed on all the possible combinations of politician and influencer embeddings, and the result is presented visually in the bipartite network in figure 7. The sizes of the node labels depend on their respective degree while the edge weight is the correlation coefficient between the two embeddings. Edges with a weight < 0.1 are removed. This offers an intuitive visual representation of the similarities between politicians and particular influencers, and their degree of correlation. By looking at the specific values one can infer that, while there are no cases of very strong correlations between two nodes (all are below 0.5), this is to be expected since the amount of data for each individual is limited. The highest correlations are achieved between the influencer Torcha



(a) *Politicians*



(b) *Influencer*

Figure 6. Most important topics in the comments from politicians (a) and influencer (b) with their 5 highest scoring words found by BERTopic.

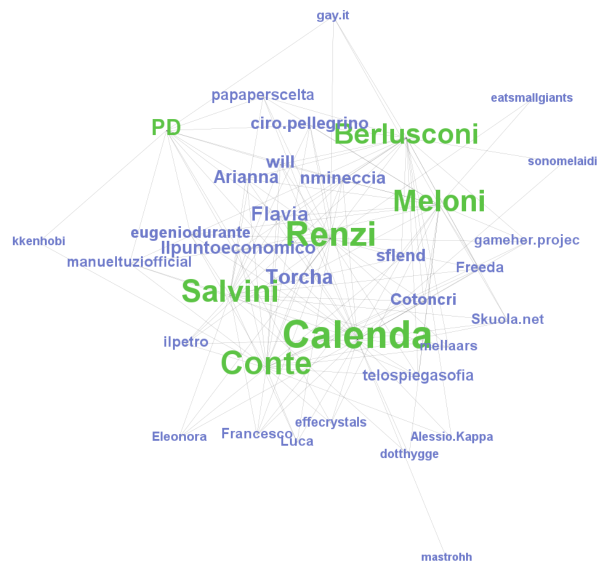


Figure 7.
Bipartite network of the correlations between politicians (green) and influencers(blue) using *PageRank*.

and the politicians Calenda, Renzi and Salvini, with values of ρ_{XY} equal to 0.37, 0.29 and 0.28, respectively.

To have a better estimate of the degree of similarity between the speech behaviour of the politicians using TikTok and the one of the more "natural" demographic in the social network, an embedding of all the documents from influencers is created, and compared with the centrality embeddings of the politicians via *Pearson's correlator*. The results are shown in table 4.
The highest scores are obtained by Calenda and Renzi with values 0.43 and 0.40 respectively, this is a interesting result since the alliance Azione-Italia Viva (informally known as the Terzo Polo), has been able to make inroads to the 10% of the younger electorate, with the better results between the 'big six' parties and a +2% with respect to the general result[15].

7.2. Tf-idf

To evaluate the robustness of the *PageRank* similarity measure, an equivalent analysis on correlations has been done by calculating the **tf-idf** for the same groups of documents. Similarly to the semantic networks, tf-idf is a common procedure that allows to assign weights to the tokens in a vocabulary. The weight of each word depends on two factors:

- **Term frequency** (tf_i): which is the number of times the word i has appeared in the documents, divided by the total size of the documents
- **Inverse document frequency** (idf_i): which is an indicator of the general importance of the term in all the documents, and is defined as the total number of documents divided by the number of documents in which word i appears.

Tf-idf is thus able to quantify the importance of a word, considering not only that more frequent words are generally more important, but also that words appearing in every document are less characterising for that class of documents, *i.e.* their weight must be lowered. Tf-idf has been implemented in Python using the method FEATURE_EXTRACTION.TEXT.TFIDFTRANSFORMER from SKLEARN.

The network obtained with this method is even less differentiated than the one obtained with *PageRank*, and the highest values of correlation obtained are: Torcha with Calenda (0.26), Flavia Carlini with Renzi (0.26) and Torcha with Salvini (0.24). The correlation values between single politicians and the whole collection of influencers are

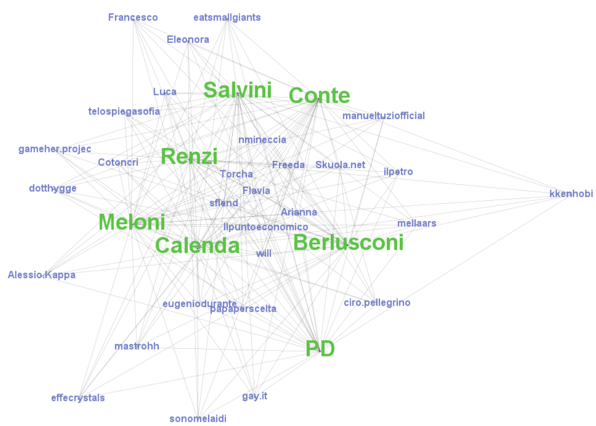


Figure 8.
Bipartite network of the correlations between politicians (green) and influencers(blue) obtained through *tf-idf*.

shown in table 4.

7.3. BERT

Now, we us BERT-BASE-ITALIAN-CASED to embed the video transcripts. It is trained on Italian Wikipedia as well as some other corpora. The input dimension is limited to 512 tokens. Therefore, we need to cut off some videos that are longer than this threshold. For the sentence embedding from BERT we need to add special tokens at the beginning '[CLS]' and end '[SEP]' of the text. The final embedding of the '[CLS]' token depends on the whole document and is used as video embedding. The embedding for the politicians and influencers are again obtained by taking the mean over their respective video embeddings. The dimension of each embedding is 768. Finally, we calculate the influencer embedding as the mean of the embeddings of all the influencers. The similarities achieved are reported in table 4. The values are very high and quite different from our previous results. Except for PD (0.90) all achieve a correlation of 0.98 or higher.

7.4. LIWC

Another way to perform correlation comparison is to use LIWC scaled mean values (explained more in detail in section 9.1) as embeddings, where we cut all the features related to punctuation, and then build the similarity matrix through *Pearson's correlator* method. The Network is shown in figure 9c while the results are reported in table 4. Giorgia Meloni has the highest similarity value with 0.77 while Partito Democratico obtains the lowest one with 0.29. This type of similarity comparison is totally different from the others since LIWC performs a sentiment analysis and not a semantic one. For more details on this type of analysis see section 9.1.

7.5. tf-idf on comments

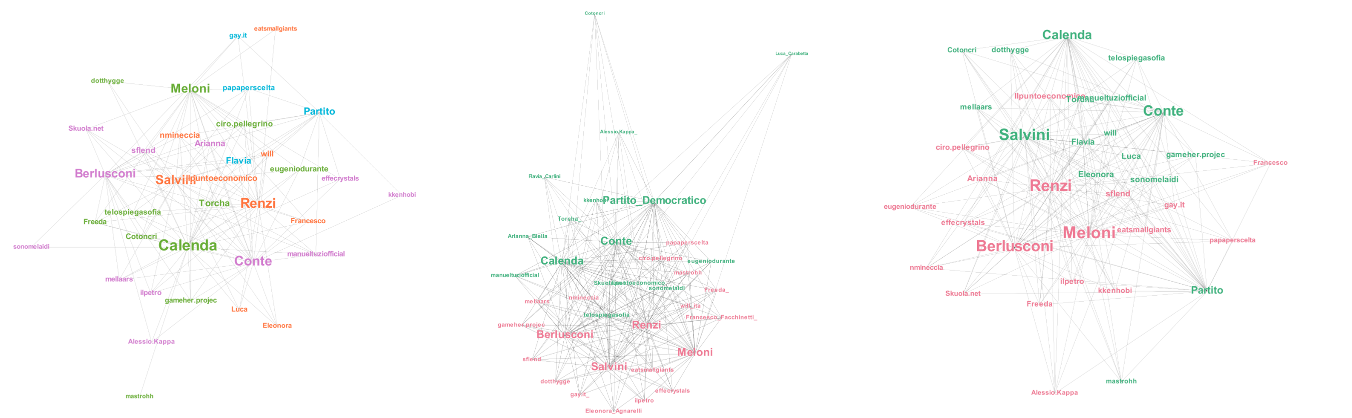
As final embedding method we repeat the tf-idf procedure described in section 7.2. But, this time we apply it on the comments of the videos. As some of the influencers do not have many comments, comparing them directly with the politicians is not feasible. Therefore, we limit the analysis of this part only on the comparison between politicians and the mean influencer embedding. The final embedding has dimension 22665. The results are reported in table 4. Meloni has the highest similarity with 0.9. Berlusconi and PD score lowest with 0.76.

7.6. Final considerations on correlations

This work presents many different ways of measuring similarity between the posts of different politicians and influencers on TikTok. The results obtained are very heterogeneous both in the average correlation magnitude and in the trend they present across the 7 politicians/parties taken into considerations.
Figure 9 shows the same bipartite network, with politicians on one side and influencers on the other, where the edges weights are the correlation between embeddings obtained with the PageRank method (9a), BERT (9b) and LIWC (9c). The node labels are scaled with PageRank and the different colors are communities detected with the modularity algorithm from *Gephi*. Even visually, the networks obtained do not present many similarities and do not seem to show specific patterns. Running modularity clustering detects 4 communities on the PageRank network, while it detects only 2 for the other ones. Each community detected is always represented by one or more politicians accounts, but each time they are grouped in different ways. A recurring pattern in the BERT and LIWC networks is that on one side Calenda, Conte and PD are in the same community, while Berlusconi, Meloni and Renzi are in the opposite one.
Moreover Partito Democratico, which is the only politician account which not represent an individual, occupies a decentered position in all the 3 networks w.r.t. the other politicians. This could be easily explained by the fact that data from the PD account is more scarce, as can be seen from 7.

On the results presented in table 4 the following considerations can be made:

- The mean values of correlation obtained with PageRank and tf-idf are significantly lower than the rest, and their trends are similar, with the highest correlation obtained by Calenda and Renzi and the lowest by PD.



(a) *PageRank* (b) *BERT* (c) *LIWC*

Figure 9.
Similarity networks of embeddings obtained with 3 different methods: PageRank (a), BERT (b) and LIWC (c). Node labels are scaled in size by PageRank anc colored according to modularity.

The result is not surprising since the two methods are based on very similar principles, and their embeddings have the same dimensionality (the length of the vocabulary) and are very sparse.

- BERT correlation parameters are all very high in magnitude. This may be especially due to the fact that its embedding is very dense, with lower dimensionality. In any case, these results are in accordance with the previous ones on the fact that PD obtains the lowest similarity among all the politicians accounts. This is particularly interesting as all other correlation coefficients for this method are nearly identical.

- LIWC results are somewhat in between when it comes to average magnitude, but the trend they show is visibly different, with Calenda being the second lowest correlated by a long margin (only better than PD, which remains last). This result is not very surprising as well since the analysis done with LIWC has a completely different background than PageRank and tf-idf: the first being based on sentiment and the others on semantic content. LIWC embedding is also of a much lower dimensionality than the others. Surely obtaining matching results for these methods was the best case scenario for the research question in exam, but one has to consider that all these methods have important degrees of bias and approximation, and when dealing with semantics and sentiment analysis, obtaining precise results cannot be expected.

- The values obtained by applying tf-idf on the comments can not be compared with the rest since they scan a different set of data, but can be helpful to see if the communities of people following the different politicians present similarity with those following actual influencers, *i.e.* if the TikTok accounts made by politicians can actually hit the target of young people interested in politics. Lastly, the average correlation obtained on the comments is higher than that of tf-idf run on the video texts. This is due to the fact that the comments text comes from a vastly wider range of people, and the content is often not very informative, resulting in a more heterogeneous variety of data. Indeed, the dimensionality of this embedding is larger than the one on videos by a factor of 10.

Correlation coefficient					
Name	PR	tf-idf	BERT	LIWC	tf-idf (comments)
Berlusconi	0.30	0.21	**0.99**	0.71	0.76
Calenda	**0.43**	0.33	0.98	0.50	0.87
Conte	0.36	0.28	0.98	0.61	0.88
Meloni	0.32	0.29	**0.99**	**0.77**	**0.90**
PD	0.27	0.16	0.90	0.29	0.76
Renzi	0.40	**0.35**	**0.99**	0.75	0.88
Salvini	0.37	0.26	**0.99**	0.72	0.89

Table 4. Correlation values between politician account and influencers.

8. Topic analysis

As described in section 4.3, all the videos in the dataset have been labeled choosing from 7 possible topics of interest. This allows to obtain different subset based on the semantic content of the documents that may give some additional insights.

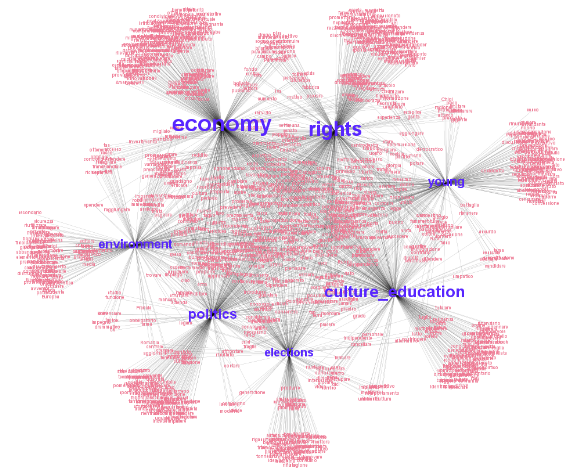
The first networks built are two bipartite ones, with words on one side and topics on the other, for all the documents from politicians (10a) and influencers(10b). An edge is present if a word appears in documents of that specific topic and the weight depends on the number of documents in which a match occurs. In the network, the size of the node label has been set depending on their PageRank value. It is interesting to see how the centrality values of the topic nodes are not always coherent with the number of documents assigned (by hand) to that class, as one would expect. This is shown in figure 11, where it is very clear that, while for some topics the PageRank value is in line with the number of documents assigned, for nodes like *rights* and *elections* the values are surprisingly high in the politicians and influencers bars respectively. This result implies that, while there is surely a certain degree of bias from the definition "by hand" of the topic labels, factors like the length of the documents and the inclination of a person to talk about a certain topic act to mitigate it.

By performing the same analysis done in section 7.1, it is possible to quantify the similarity between politicians and topics and interpret it as a measure of the disposition of an individual to talk about a specific topic. This is represented in the bipartite network in figure 12, where the edge weight between a politician and a topic is the Pearson's ρ between their two PageRank embeddings. To have a more quantitative view, table 5 shows the "preferred" topic by each politician, with the respective value of ρ. The table

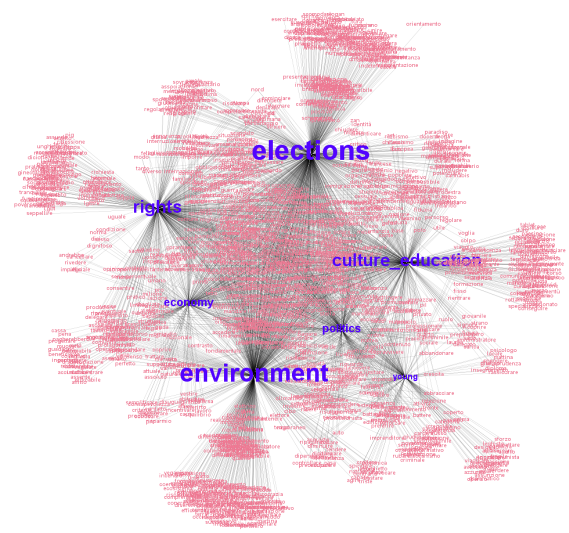
Topic preference			
Name	topic	ρ_{XY}	frac. of documents
Berlusconi	young	0.73	0.50
Calenda	economy	0.53	0.23
Conte	politics	0.53	0.31
Meloni	rights	0.41	0.20
PD	rights	0.35	0.20
Renzi	elections	0.38	0.15
Salvini	economy	0.39	0.23

Table 5. Favourite topic of each politician by the correlation value. The last column shows the fraction of document of the preferred class in the collection of documents by that politician.

provides another confirmation that the bias inserted in the dataset by "artificial" labeling is not extreme: even though



(a) *Politicians*



(b) *Influencers*

Figure 10.
Bipartite networks of words belonging to documents of each topic for videos from politicians (a) and influencers (b).

they present some similarity in their trend, the values of ρ_{XY} shown do not reflect completely the fraction of documents of that topic in the politician document collection.

This results confirm what already discussed in section 6.1.2, for instance Matteo Salvini seems to focus on economy-related topics, where he obtains the maximum Pearson's ρ coefficient.

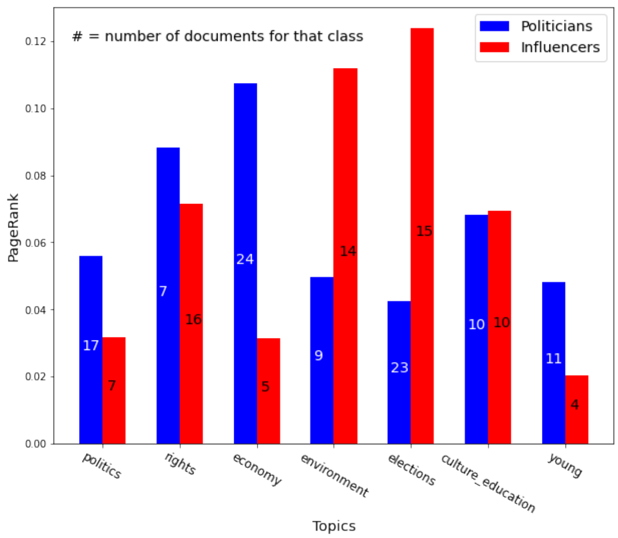


Figure 11.
PageRank values of the topic nodes in the networks of figure 10. Inside each bar the number of document for that class of document and that topic is specify, in order to confront it with the values shown by the bar heights.

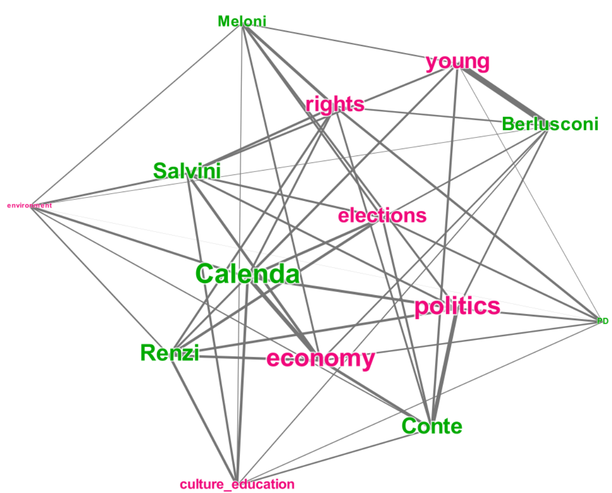


Figure 12.
Bipartite graph showing the similarities between politicians (green) and topics (pink) computed with the PageRank similarity measure. The edge weight is represented by its thickness.

9. Sentiment analysis with LIWC

9.1. Introduction to LIWC: How does it work?

The core logic of Linguistic Inquiry and Word Count (LIWC) comes from decades of scientific research demonstrating that people's language can provide extremely rich insights into their psychological states, including their emotions, thinking styles, and social concerns. For example, if someone is using a lot of words like happy, excited, and so on, they are probably feeling happy, and we can use this information to reliably estimate their current emotional state. But how does LIWC work? LIWC reads a given text and compares each word in the text to the list of dictionary words and calculates the percentage of total words in the text that match each of the dictionary categories. Furthermore, it is important to know that LIWC ignore context, irony, sarcasm, and idioms. For our study we have averaged the values of LIWC features for every politician and influencer, either for texts and comments.
Finally the values are also scaled with the minmax scale method from SCIKIT-LEARN library of PYTHON. This implies that the values shown in the barplots are not the absolute ones, they should only be compared considering the same feature across different politicians/influencers. The value of a feature cannot be compared with that of a different one since the rescaling is based on the min/max values of each feature, averaged on all the people across the dataset.

9.2. Texts Analysis

9.2.1 Summary Features

LIWC-22 contains four summary measures: Analytical Thinking, Clout, Authenticity, and Emotional Tone. Each of the summary measures are algorithms derived from various LIWC variables based on previous empirical research [16] [17].

- **Analytical thinking**: To go viral on tiktok there could be different approaches. Is it better for a Politician to be more formal o informal? Analytic Thinking captures the degree to which people use words that suggest formal, logical, and hierarchical thinking patterns. People low in Analytical Thinking tend to write and think using language that is more intuitive and personal. Language scoring high in Analytic Thinking tends to be rewarded in academic settings and is correlated with things like grades and reasoning skills. Language scoring low in Analytic Thinking tends to be viewed as less cold and rigid, and more friendly and personable.

- **Clout**: it refers to the relative social status, confidence, or leadership that people display through their writing or talking.

- **Authenticity**: What is the honesty level in our politicians' campaigns and how can we detect it? Obviously having a precise measure of the honesty in a text is not as trivial as it seems, but LIWC can help us. When people reveal themselves in an "authentic" or honest way, they tend to speak more spontaneously and do not self-regulate or filter what they are saying. So political leaders with a high measure of authenticity are the ones with little-to-no social inhibitions.

- **Emotional tone**: Although LIWC-22 includes both positive tone and negative tone dimensions, the Tone variable puts the two dimensions into a single summary variable. The algorithm is built so that the higher the number, the more positive the tone. Numbers below 50 suggest a more negative emotional tone.

The results given by LIWC algorithm on the texts from polticians' videos are shown in figure 13.

The lowest Analytical thinking score is obtained by Berlusconi. This is in line from what we expected for many reasons: firstly he is known for his playful and, in some cases, inappropriate behaviour, furthermore he is trying to reach young people which, in general, talk to each other in a more informal way with respect to older generations. On the other hand the highest value for the same feature is reached by Conte, who is known to be more formal and, in addition, he is a university professor graduated in law, which confirms what said above.

The values of the Clout feature for the politicians are very similar with a high mean value. The highest score is reached by Berlusconi and the lowest one, against all expectations, by Salvini. In general the politicians seem to display more confidence and leadership if compared with influencers, as one can see in figure 14.

The Authenticity feature does not lead to reasonable results, this is to be expected since it is very difficult to give a measure to the honesty of someone, especially if the dataset is as biased as ours. In any case, the values for all politicians are very similar and the lowest values are obtained by more moderate politicians as Calenda and Conte.

Last but not least, the Tone feature gives some interesting insights for our analysis. The right wing parties seem to have a more positive tone if compared with the others and Berlusconi has the highest value, probably for the same reasons already discussed. This can be due to the fact that during the elections the right parties had the majority in electoral voting forecasts, so in this case PD, Renzi and Calenda are representing the opposition.

9.2.2 Use of pronouns

The use of pronouns is very important and not so trivial to interpret. Several studies [18] showed a connection with

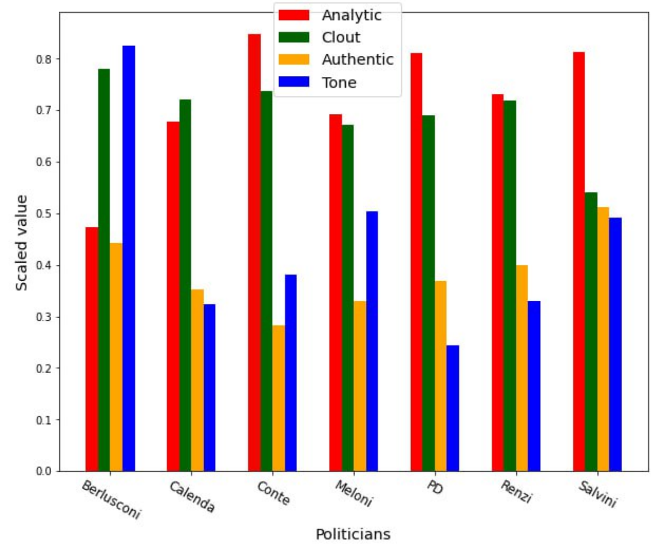


Figure 13. LIWC scaled values for the main summary features in politicians' videos

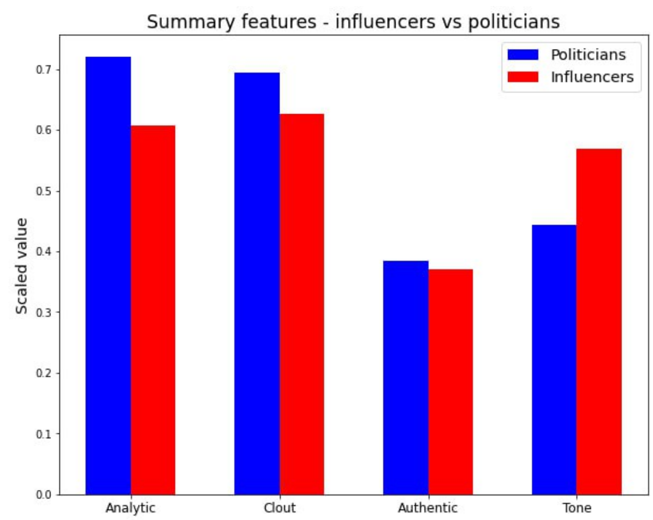


Figure 14. LIWC Scaled values for the main summary features in politicians video text

self-references (e.g "I" and "me") and positive political ads, while other references were more present in negative ones. A particular attention is given to the first plural pronouns "we", "us" and "our" since, in some cases, are signal of group Identity and cohesion, for instance "we" could be used to promote interdependence as in "we can do this". Furthermore the use of greater first-person plural can be correlated with higher rank. Again, the results from LIWC are shown in figure 15.

Berlusconi is in the spotlight for the umpteenth time, he got the higher score for the first singular and second personal pronouns. The self-references are associated to positive po-

litical ads as said before, furthermore it can be associated to his egocentric personality. The second personal pronouns are used, for instance, when he urges young people to vote and assert their rights. Moreover, he gets the highest score even for the first plural pronoun. This can be seen just by selecting randomly one of his videos, he always uses ”we” instead of ”they” when he talks about young people, combined with a conspicuous use of modal verbs as ”we must” or ”we can”, to create cohesion and a connection with his audience.
On the other hand, Conte get the highest value for the third plural pronoun, this can be explained by the fact that in many videos he refers to the right wing parties to explain what they could do if they win the elections or what they have already done in the past and the majority of his videos can be considered negative political ads.
In general, as one can see in figure 16, politicians use more frequently the first plural pronoun ”we” as a signal of identity and group cohesion.

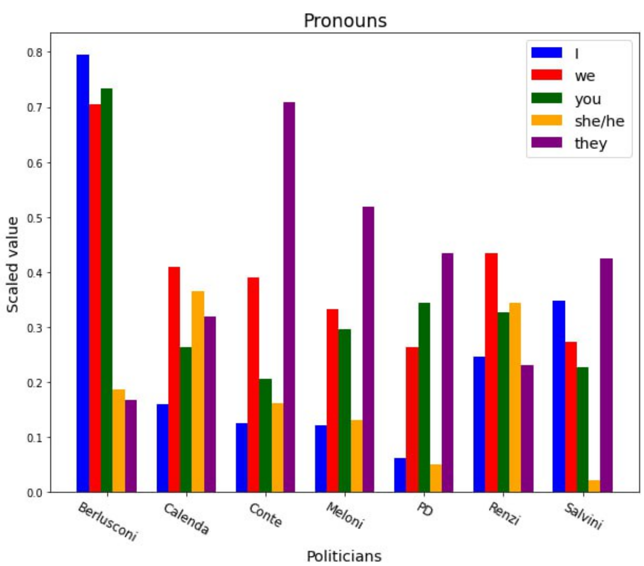


Figure 15. LIWC scaled values for pronouns features in politicians’ videos

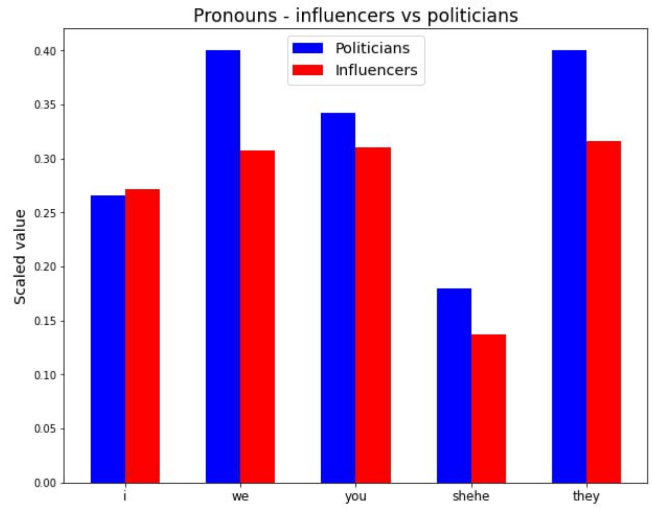


Figure 16. LIWC Scaled values for pronouns features in politicians video text

9.3. Comments Analysis

The same LIWC analysis has been performed for the comments under the TikTok videos of the dataset, trying to understand the reactions and the emotions of each community.

9.3.1 Emotionality: Positive and Negative Emotions

Research suggests that LIWC accurately identifies emotion in language use[16]. For example, positive emotion words (e.g., ”amore”, ”bello”) are used in writing about a positive event, and more negative emotion words are used in writing about a negative event. LIWC ratings of positive and negative emotion words in our case could correspond to the reaction (appreciation or despising) to the politicians’ videos and to their election campaigns.
The LIWC features used for the analysis are: *posemo*, *negemo*, *anx*, *anger* and *sad*. The results are shown in figure 17
As expected, right wing politicians have high values for positive emotions, in particular Giorgia Meloni and Berlusconi, the first one actually won the elction.
Salvini seems to have lower values for the same feature and this can be explained by the fact that he lost over 4.5 millions electors after having joined the Draghi government during the Covid Pandemic, realizing the overtaking of Giorgia Meloni. However, the negative emotions are pretty similar along the politicians’ community.
A fact that immediately sticks out is the anxiety values in Partito Democratico community. The high value of “anxiety” from the comment section of Partito Democratico might be related to the quantity of comments itself, which is low if compared to other profiles we took into study. In addition, from a qualitative and manual analysis, in these comments the anxiety is clearly expressed, keywords like “fear” and “afraid” combined with “future” and “help” are present.

9.4. Further insights on LIWC sentiment analysis

The highest level of communicating without filters, in a natural and spontaneous way, belongs to Salvini. With a further analysis based on the videos, we can say that the medium close-up, combined with the use of a mobile

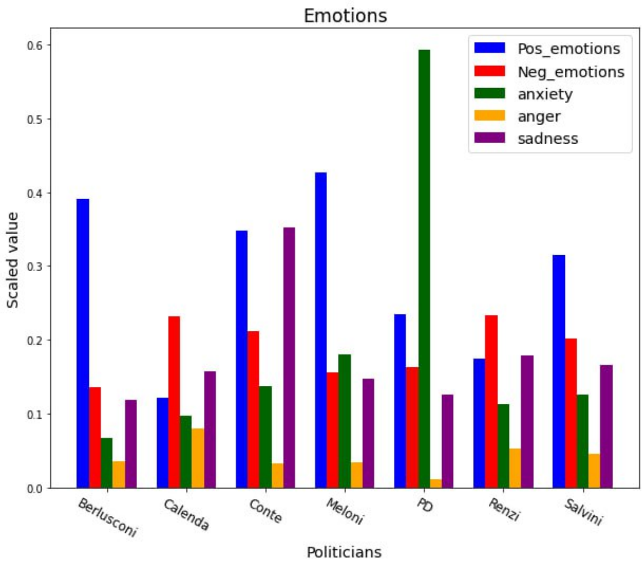


Figure 17. LIWC scaled values for emotions in comments under politicians' videos

camera held in Salvini's hand amplifies the illusion of direct participation. It allows a strong presence of the politician's body, that serves the political (specifically, populist) rhetoric of being in contact with his audience. However, the most important thing is the addressing of individual viewers, and the politician's urgency of addressing them as they appear on the screen (he interrupts himself circa every four minutes to thank them). These interruptions produce a meaning-effect on viewers in that they seem to reproduce a genuine instant interaction with the represented participant's followers ("Thank you, Nicoletta, thank you, David, thank you, Adriana, thank you, Claudia, thank you, Marco"). This is in line with other choices: Salvini always uses vertical angles (both high and low angles) in close shots that allow proximity between the politician and the viewer, increasing social relation and reducing social distance (Van Leeuwen 2008), and make his body more visible[19]. Regarding the Emotional tone, by identifying the positive tone used in the communication, Berlusconi has the highest level of it in his speeches because he frequently used positive words (such as miracle, trust, dream, love, or happiness) and statements like "Forza Italia is the victory of love" or that his party is the "party of love" (Forza Italia, 2013). In this regard it can be said that Berlusconi's strategy has been one of the most successful, both because over the years he has always used simple communication in a coherent way, above all to refer to young people as in this case, but also because he has used his image as a leader to become popular again in a social network that is commonly used by the new generation.

His first video on TikTok has registered over 10 millions views, 700k likes and over 128k shares, moreover his account is the third most popular one between politicians even though he joined TikTok community only in September. The effect of his strategy can be easily seen in figure18, where the number of Berlusconi's google searches has a spike concurrently with his subscription on TikTok.

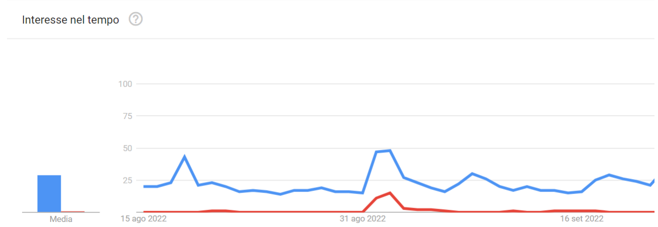


Figure 18. Google trends for Berlusconi(blue) and Berlusconi TikTok(red)

10. Conclusions

It is clear how politicians decided to take advantage of this big platform, TikTok, to approach the young generation as a strategy. The use of social media to communicate with potential voters is a strategy politicians use since the proven efficiency of these portals; relying on concepts like homophily, polarization and echo-chambers which favours the creation of well-defined communities even in cyberspace. These factors have encouraged people to come together into communities on every social media. Since TikTok is mostly young-associated, it has not yet experienced the birth of clear political communities. Therefore, the question comes naturally: did politicians use this platform only to communicate with the younger generations and did it work? Just like in Trump's case in 2016, the social media campaign surely helped in the election of far-right politicians, but based on surveys the young people have not been voting for these extreme parties [12]. In fact, Meloni's case is similar to Trump's one [13](An examination of the 2016 electorate, based on validated voters — Pew Research Center): the majority of their voters are between 35-65 years old, while the under 35 cover the smallest percentage of voters. The results obtained in this work, combined with the outcome of last September elections, were able to confirm that the right party politicians are mostly present on TikTok

[12](Chi sono oggi gli elettori di Giorgia Meloni https://www.rainews.it/photogallery/2022/10/il-profilo-delleletorato-di-fratelli-ditalia-nellanalisi-dellistituto-demopolis-chi-sono-oggi-gli-elettori-di-giorgia-meloni--264299d2-cb42-49ac-a587-ffb1f3ed98c3.html

[13]An examination of the 2016 electorate, based on validated voters https://www.pewresearch.org/politics/2018/08/09/an-examination-of-the-2016-electorate-based-on-validated-voters/

to catch the young generation attention. Indeed, the success of right-wing parties can be seen both in the results of the elections and in their numbers on social media. With a combined social-following bigger than 2.5M followers and 23.1 M likes, the three right-parties Fratelli d'Italia, Lega Nord and Forza Italia hold the hegemony on the social media platform TikTok; unlike the opposition parties who find themselves at a slight disadvantage in terms of engagement.

It is difficult to define whether young adults have oriented their voting preferences around the presence of politicians on TikTok; to do that, it would be necessary to conduct further research using a positional approach. What can be said with certainty is that this was a considered and strategic choice on the part of all the political forces and every party managed it in different ways; there are those who have been more suggested, such as Matteo Salvini and Silvio Berlusconi,(as discussed in sec. 9.4) who appear to be the subjects with the most engagement and whom we find to be the most central nodes in our research, and those who have tried entering the world of TikTok with some difficulty, for example, as shown in tab. 4, Calenda and Renzi seem to be more correlated to influencers but without good results on the platform.

However, as seen from the election results, from the youngest electors (18-24 y.o.) the parties of Matteo Salvini and Silvio Berlusconi received 2.5% and 7.5% [14] (8.8% and 8.1% total electors) of the votes respectively, which is significantly less than Enrico Letta ,13.5,0%(19.1% total electors), who had the lowest presence on TikTok. The only figure who remains consistent in terms of number of followers, engagement and centrality in the networks is Giorgia Meloni, who was confirmed to be the new Italian Prime Minister with 26,0% of the votes, but with an only 15.4 % for the young electors. We can affirm that the presence on social networks boosts their message and allows the politician to be closer to its voters, creating communities, instead of having the usual distant approach. But how would the elections have gone without the contributions from TikTok? Would the percentages of young electors have changed? We do not have the knowledge to answer these questions with the data and methodologies at our disposal. The only consideration which can be done is that the average age of the TikTok users is less than 24, and a large amount of these users are underage (indeed, TikTok was originally created for the minors) which means that the majority of the visualizations is done by people which do not yet have the right to vote. It is possible that the strategy implemented by politicians in their advent on this platform is to aim at future elections (not the 2022 ones), hoping that young people will one day become their supporters. But for know looking at the percentages of young electors, the strategy of exploiting TikTok did not lead to great changes in the final results.

[14] `https://elezioni.repubblica.it/2022/elezioni-politiche`

References

[1] Airi-Alina Allaste. *The future of politics is social media*. University of Helsinki. 2021. URL: https://www.helsinki.fi/en/news/economics/future-politics-social-media.

[2] M. Petrova, A. Sen, and P. Yildirim. “Social Media and Political Contributions: The Impact of New Technology on Political Competition”. In: *Management Science, forthcoming* (2020).

[3] Giulia Pantaleo. “I politici italiani e i social network”. In: *MICS* (2022).

[4] Gaia Pianigiani. “Key Takeaways From Italy’s Landmark Election”. In: *The New York Times* (2022).

[5] Pierri F., Liu G., and Ceri S. “ITA-ELECTION-2022: A multi-platform dataset of social media conversations around the 2022 Italian general election.” In: *arXiv* 2301.05119 (2023).

[6] Yang Y. and Ha L. “Why people use TikTok (Douyin) and how their purchase intentions are affected by social media influencers in China: A uses and gratifications and parasocial relationship perspective.” In: *Journal of Interactive Advertising* 21(3), 297-305 (2021).

[7] Burns-Stanning K. “Identity in communities and networks TikTok social networking site empowering youth civic engagement.” In: *Debating Communities and Networks 11 Conference* Vol. 27, pp. 1-11 (2020).

[8] E Quintelier. “Differences in Political Participation Between Young and Old People.” In: *Differences in Political Participation Between Young and Old People.* 13 (2): 165–180 (2007).

[9] Santo Fortunato. “Community detection in graphs”. In: *Physics reports* 486.3-5 (2010), pp. 75–174.

[10] N. Agarwal H. Bisgin and X. Xu. “”Investigating Homophily in Online Social Networks”. In: ed. by International Conference on Web Intelligence and Intelligent Agent Technology. Vol. 486. 3-5. Toronto, ON, Canada, 2010, pp. 533–536.

[11] *TikTokAPI documentation*. URL: https://davidteather.github.io/TikTok-Api/docs/TikTokApi.html.

[12] *TikAPI documentation*. URL: https://tikapi.io/documentation/.

[13] *SpaCy documentation*. URL: https://spacy.io/.

[14] *BERTopic documentation*. URL: https://maartengr.github.io/BERTopic/index.html.

[15] TGCOM24. *Elezioni 2022, come hanno votato i giovani? Gli under 35 premiano i piccoli partiti.* URL: https://www.tgcom24.mediaset.it/skuola/elezioni-2022-come-hanno-votato-i-giovani-gli-under-35-premiano-i-piccoli-partiti-_55327322-202202k.shtml.

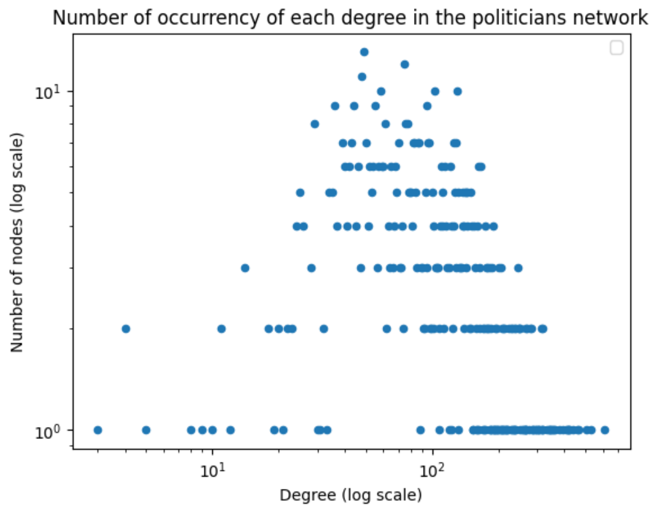
[16] *LIWC-22 software*. URL: https://www.liwc.app/.

[17] Yla R. Tausczik and James W. Pennebaker. “The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods”. In: *Journal of Language and Social Psychology* 29(1):24-54 (2010).

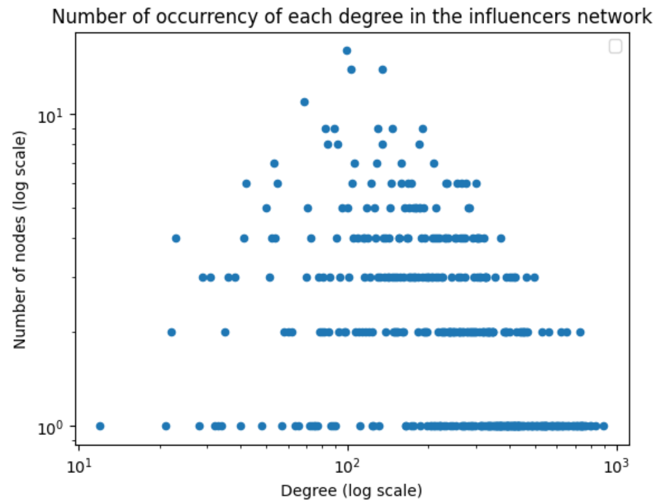
[18] Haynes S. E. Gunsch M. A. Brownlow S. and Mabe Z. “Differential linguistic content of various forms of political advertising”. In: *Journal of Broadcasting & Electronic Media* 44(1), 27–42 (2000).

[19] M. Lya Zummo. “Performing authenticity on a digital political stage, politainment as interactive practice and populist performance”. In: *Iperstoria* 15 (2020), p. 110.

Appendix A. Additional plots

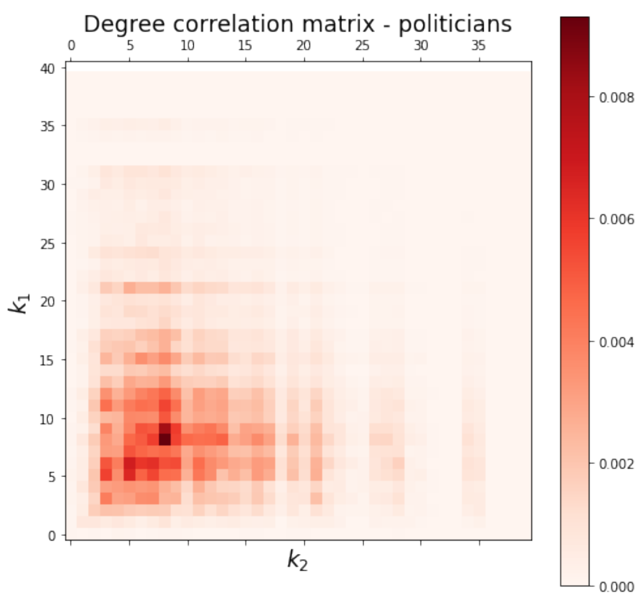


(a) *Politicians*

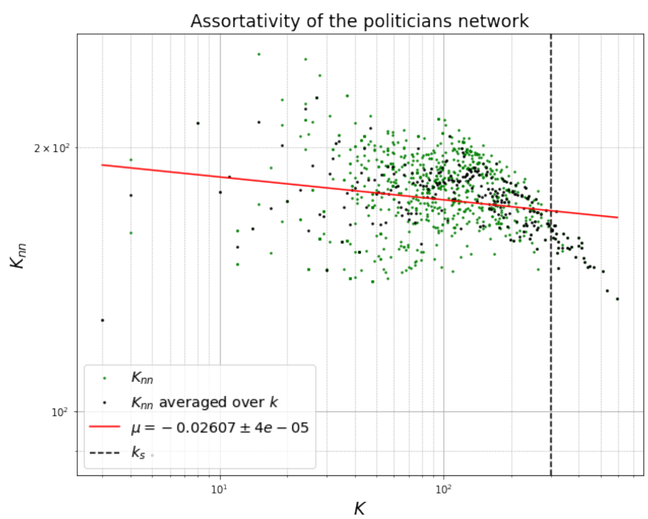


(b) *Influencer*

Figure A.1. Distribution of node degrees using a log-log scale.

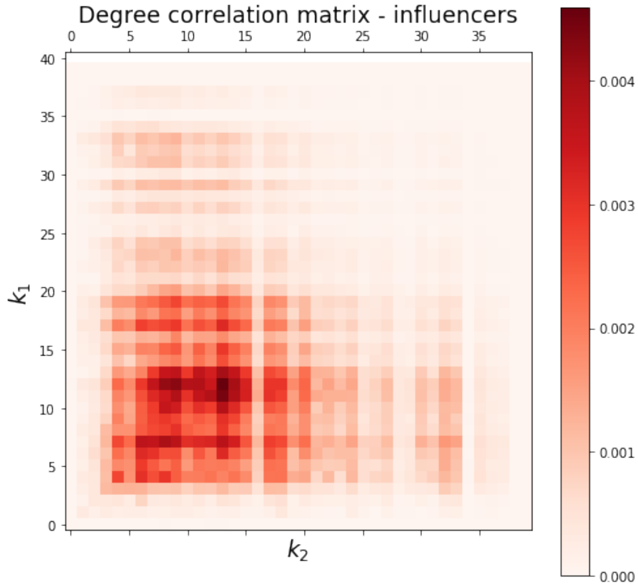


(a)

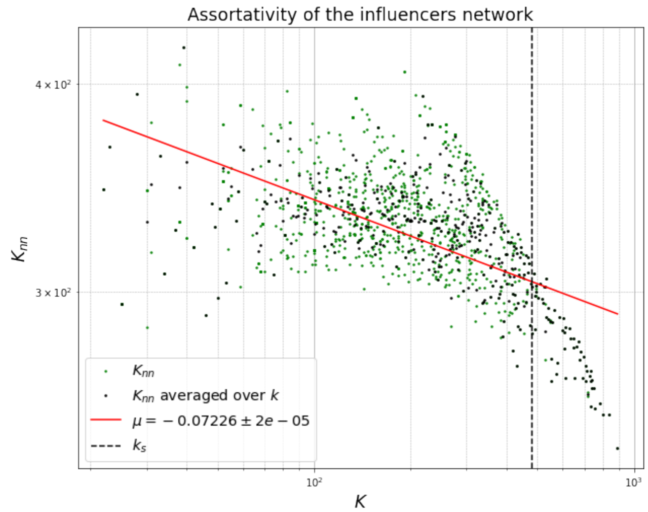


(b)

Figure A.2.
(a): Degree correlation matrix of the politicians adjacency network. (b): Assortativity of the politicians adjacency network, showing the *average neighbor degree* K_{nn} and its average over k. The fitted red line visualizes the assortativity value μ, while the dashed black line represents the estimated degree of the structural cutoff k_s.



(a)



(b)

Figure A.3.
(a): Degree correlation matrix of the influencers adjacency network. (b): Assortativity of the influencers adjacency network, showing the *average neighbor degree* K_{nn} and its average over k. The fitted red line visualizes the assortativity value μ, while the dashed black line represents the estimated degree of the structural cutoff k_s.

Appendix B. Tables

Total Video: 171

Politicians TikTok videos					
Politician	**Political party**	**Username**	**N.of followers**	**N.of likes**	**N.of videos**
Giorgia Meloni	Fratelli d'Italia	@giorgiameloni_ufficiale	1M	6.6 M	10
Matteo Salvini	Lega Nord	@matteosalviniufficiale	841.1 K	11.3 M	26
Silvio Berlusconi	Forza Italia	@silvio.berlusconi	745.6 K	5.2 M	8
Giuseppe Conte	M5S	@giuseppeconte	538 K	6.1 M	16
Matteo Renzi	Italia Viva	@matteorenziufficiale	53.3 K	456.1 K	13
Carlo Calenda	Azione	@carlocalendaofficial	26.4 K	193.4 K	22
Partito Democratico	PD	@partitodemocratico	7.32 K	87 K	5
Total					100

Table 6. : list of politicians and the related information regarding their TikTok profiles, sorted by number of followers

Labeling for the influencers:

- Under 100 K followers = micro-influencer;
- Over 100 K followers = influencer;
- Under 1K followers = personal profile / Non-influencer;

Influencers TikTok videos					
Name	Category	Username	N.of followers	N.of likes	N.of videos
Torcha	Info Page	@torcha	204.4 K	5.1 M	14
Flavia Carlini	Influencer	@flavia.carlini	163.7 K	1.9 M	12
Manuel Tuzi	Political figure/Micro-influencer	@manueltuziofficial	15.3 K	147 K	5
Arianna Biella	Influencer	@ariannabiellas	440.9 K	37.3 M	4
Nicola - Ilpuntoeconomico	Info Page/micro-influencer	@ilpuntoeconomico	18.8 K	256.6 K	4
sflend	Info Page	@sflend	21.1 K	2.2 M	3
Luca Carabetta	Political figure/ Micro-influencer	@luca.carbetta	33.5 K	1 M	3
Sofia Pasotto	Micro-influencer	@telospiegasofia	57.4 K	2.1 M	3
Cristina Coto	Micro-influencer	@cotoncri	38.5 K	570.7 K	2
Will Italia	Info Page	@will_italia	265.6 K	5.4 M	2
Eleonora Agnarelli	Influencer	@eleonora_agnarelli	515 K	18.9 M	1
Mastrohh	Influencer	@mastrohh	156.6 K	7 M	1
Francesco Facchinetti	Celebrity / Influencer	@frafacchinetti	175.5 K	2.1 M	1
Freeda	Info Page	@freeda	244.3 K	6.2 M	1
DottHygge	Micro-influencer	@dotthygge	3.36 K	52.7 K	1
Alessio Krakulli	Micro-influencer	@alessio.kappa	7.86 K	366.1 K	1
Nicola Mineccia	Micro-influencer	@nmineccia	24.2 K	1.1 M	1
Skuola.net	Info Page	@skuolanet	388.8 K	6.5 M	1
Small Giants	Info Page	@eatsmallgiants	3.03 K	24.1 K	1
Domenico Petrolino	Micro-influencer	@ilpetro	41.1 K	656.7	1
Valentina Pano	Influencer	@sonomelaidi	100.7 K	2.6 M	1
Papà per Scelta	Micro-influencer	@papaperscelta	188 K	5.8 M	1
Eugenio Durante	Micro-influencer	@eugeniodurante	13.5 K	134.7 K	1
Ciro Pellegrino	Micro-influencer	@ciro.pellegrino	35.1 K	1.7 M	1
Effe	Influencer	@effecrystals	156.6 K	5.6 M	1
Gay.it	Info Page	@gay.it	20.6 K	197 K	1
GameHer	Micro-influencer	@gameher.project	32.6 K	657.5 K	1
Mella	Personal profile	@mellaars	133	2.08 K	1
Federica	Personal profile	@kkenhobi	912	103.8 K	1
Total					71

Table 7. list of politicians and the related information regarding their TikTok profiles, sorted by number of followers

Appendix C. API tutorials

One of the main issues encountered at the beginning of this work was the extraction of the data from TikTok, and since it is a rather new endeavor, it was difficult to find on internet some examples or tutorials on how these software works. We thus decided to write these two user guides for the APIs, hoping that they can be useful in the future.

C.1. TikTokAPI User Guide

Opening a terminal window and running the two following lines allows to install the two main libraries to use TikTokAPI:

```
pip install TikTokApi
python -m playwright install
```

After the second code line there could be some errors due to the absence of some libraries. To work around this problem look at the output given during the installation, a list of packages is printed. By installing all these libraries one will have all the needed tools.
A different way to install this API is through **anaconda**, where is possible to search which packages one wants to install. In this way, there should not be the same errors because the anaconda browser should install all the libraries.
In the next lines of code is shown how to download a video using this API.

```
from TikTokApi import TikTokApi

with TikTokApi(custom_verify_fp=cos_key) as api:
    video = api.video(id=video_id)
    video_data = video.bytes()
    file_name="Video_TikTok.mp4"
    with open(file_name, "wb") as out_file:
        out_file.write(video_data)
    api.shutdown()
```

To get the custom verify key one has to connect to TikTok using the browser and accept all the cookies, then by clicking to the padlock icon one can visualize all the information characterizing the web site. Opening the cookies list under the folder *tiktok.com* there is a list of different cookies, the needed one is under name *s_v_web_id*.
It is probable that this cookie code is not present, to avoid this issue delete all the TikTok cookies and accept them back.

C.2. TikAPI User Guide

To install this API run from the terminal:

```
pip3 install tikapi
```

Two different keys are needed to use the software.
From the API web page (`https://tikapi.io/`) create an account, then inside the **developer** screen the first key, API KEY, can be found under the **setting** folder. The second one, ACCOUNT KEY, is automatically generated after connecting a TikTok account to the API and can be found in the **users** folder. This software is useful to get different information from TikTok, for example: user information (likes,followers,posts, etc.), posts information (likes,comments, etc.).
In this work it was used to get the comments of the selected posts. An example of how get the comments is:

```
from tikapi import TikAPI,
        ValidationException, ResponseException
api = TikAPI("myAPIKey")
User = api.user(accountKey="AccountKey")

j = 0
# single video request counter
total_n_comments_read = 0
# total comments counter
comments = []
# list to collect comments

try:
    response = User.posts.comments.list(
        media_id = id,
        count=number_of_comment_each_request)
        # (max possible value is 30)

    response_json = response.json()
    j=1
        while(response):

            response_json = response.json()

            if response == None:
                break # no more responses

            # extend output with response comments
            comments.extend(
                    response_json['comments'])
            total_n_comments_read +=
                    len(response_json['comments'])

            # check if we reached maximum
            # n of requests for this video
            if j >= max_requests:
                break #stop

            # get next item
            cursor=response.json().get('cursor')
            response=response.next_items()

            if response == None:
                break # no more responses

            response_json = response.json()

            if response_json['comments']==None:
                print("Continue!")
                response = response.next_items()

                continue #no more comments (?)

            j = j + 1

# handle exceptions
except ValidationException as e:
```

```
    print(e, e.field)

except ResponseException as e:
    print(e, e.response.status_code)

# remove comments from last
# response object to get the meta data
response_json['comments'] = None

# creat dictionary with metadata and comments
out={"meta":response_json,"comments":comments}

with open(file, 'w') as f:
    json.dump(comments, f)
```