UNIVERSITÀ
degli STUDI
di CATANIA

Dipartimento
di Fisica
e Astronomia
"Ettore Majorana"

PHD PROGRAM IN COMPLEX SYSTEMS

ANDREA RUSSO

AI-INFLUENCER TO MITIGATE RADICALISM AND SOCIAL THREATS ON TELEGRAM; AN APPLICATIVE TOOL FOR SECURITY AND COMPLEX SOCIAL DYNAMICS.

PhD THESIS

SUPERVISOR:
PROF. GIOVANNI GIUFFRIDA

CO-SUPERVISOR:
PROF. TAHA YASSERI

CO-SUPERVISOR:
PROF. FRANCESCO MAZZEO RINALDI

ACADEMIC YEAR 2022/2023

July 25, 2023

A human being should be able to change a diaper, plan an invasion, butcher a hog, conn a ship, design a building, write a sonnet, balance accounts, build a wall, set a bone, comfort the dying, take orders, give orders, cooperate, act alone, solve equations, analyse a new problem, pitch manure, program a computer, cook a tasty meal, fight efficiently, die gallantly.
Specialization is for insects.

Robert Heinlein, Time Enough for Love

In ancient times having power meant having access to data. Today having power means knowing what to ignore.

Yuval Noah Harari, Homo Deus: A History of Tomorrow

The day science begins to study non-physical phenomena, it will make more progress in one decade than in all the previous centuries of its existence.

Nikola Tesla

Contents

1 Introduction

> *"If you want to go fast, go alone, if you want to go far, go together"* - African Proverb

What is the main factor that enables humans to go together, and thus to go far? Is it the person's respect, authority or credibility? This thesis tries to see how the credibility and authority of an individual can lead to a phase transition of a social system, forcing the social sample to move one direction or another.

Being able to convince people, to take a common path, with common goals, is the job of a political leader. And to do so, it is necessary to show and be able to communicate, to convince people to follow, or be followed.
Could the main factor that makes people follow the leader be the leader's credibility? Humans do not follow people they do not believe, or who have been enchanted. But also, communication plays a fundamental role. Knowing how to communicate, empower and get people stimulate about the topic under discussion. So we can define as communication the process of sending information by means of messages, where a person is connected to one, or several people, and has the possibility to stimulate personal feeling.
The ability to galvanise people to reach a particular goal is the basis of many studies and theories relating to politics or propaganda.

In the political sphere, the people who are followed are the leaders, but in social networks, where the goal is not only political, but can be also economical, the leaders (or the people who get followed) are called *Influencers*. Therefore, these are the people who can influence other people's values.

However, there are two types of leaders, inherent to the type of topics. There are single-topic leaders (specialised in a particular field, e.g. economics or politics), and generalist-leaders. The transition from single-topic to generalist-leaders occurs when single-topic leaders become so well known in a transversal way that they go beyond the single-topic, and thus become generalists. As a comparison, it can be a river that grows larger as time goes by, and at a certain point it floods, covering different thematic areas.

For a sociological standpoint, an influencer is not someone that has authority to modify the perception or values of others. Instead it is the people who allow and give the influencer the authority to modify the perception or values. [1, 2]
Therefore, an influencer is when, at a given time, influence is exerted.

In 2019 Enke et al. [3] have defined social media influencers as *third-party actors that have established a significant number of relevant relationships with a specific quality for and influence on organizational stakeholders through content production, content distribution, interaction, and personal appearance on the social web.* Subsequently, Influencers define an action plan in the communication sphere in which social media influencers are addressed or perform activities with strategic significance for themselves or the goals of the organization. It was also noted that there is a connection between the average sentiment of the influencer's communication and the effect of the message sent, thus showing that a higher sentiment can have more effect in influencer communication. [4]

The character traits of a leader, that is to say the characteristics of their personality, may be important, but they do not make good leadership. There is no leadership without a group of people (the followers) who accept being influenced by the action of the leader(s).

An influencer, in order to change values or perceptions, needs to obtain authority and social approval from followers. These can be obtained in two ways.

1. Fabricate and raise it over time
2. Acquire it from a rapid (often sudden) event, where competence in handling the sudden event is demonstrated and/or recognised.

The difference between fabricate or acquire depends on the process of creation and the situation. The former indicates a process of construction and/or fabrication starting from a careful construction of one's public image over time. While the second, is something that happens in a short time and is given by the recognition of one or more qualities, from a large majority of people.
Some people want to become leader/influencer and they work to become it, while others were asked. Two concrete examples can be given:

John Cena (John Felix Anthony Cena Jr)
John Cena is a famous American sportsman who became famous as a wrestler in WWE program, he also published rapper songs/albums and then become an actor in several US movies. But in recent years he has become even more famous for a quote, highlighting a *white monkey role* by the Republic of China. During the film's promotional tour in 2021, Cena referred to Taiwan as "a country". He subsequently posted an apology on social media as China considers Taiwan a part of China.
A *White monkey role* is a term used to refer to the phenomenon of white foreigners or immigrants in China being hired for modeling, advertising, teaching English, or promotional jobs on the basis of their race. The phenomenon is based on the perception that association with foreigners, specifically white foreigners, can signify prestige, legitimacy, and international status, who are not expected to be fluent in Chinese in some cases.. White monkey jobs are often related to marketing and

advertising. The "white monkey" may be hired to act as an associate of an individual or pose as an authoritative figure to promote a brand or company, and businesses will occasionally hire these individuals to pose as a founder or executive. [5] In this case, John Cena promotes the CCP (Chinese Communist Party) propaganda. [6, 7]
So, over time, year by year, he has built up an authority where people follow him by being influenced. This power is used by the CCP for its internal and external information and propaganda campaigns.

Vladimir Zelensky (Volodymyr Oleksandrovyč Zelens'kyj)
Volodymyr Zelensky was an Ukrainian actor, director, screenwriter and comedian. Like John Cena he grow in popularity and had the agreement of the Ukrainian population. In early 2019 he used his popularity to became the Ukraian president, and he become president in May 2019.
Despite numerous meetings and attempts at negotiations between Russian government representatives and key European leaders, on 2023/2/24 the Russian army began its invasion of Ukraine, preceded by massive bombardments of major cities. Zelensky proclaimed that martial law had come into force and initiated mass mobilisation, with the voluntary enrollment into the army of citizens who are of age and able to fight. In numerous messages posted on social media, Zelensky stated that he did not want to leave Kiev, besieged by Russian troops, and that he refused a safe-conduct, provided by the United States, which would have allowed him to take refuge in Lviv. [8, 9, 10].
The actions of Volodymyr Zelensky have generated a closeness with the population and institutions. In fact, given his refusal to flee, he was put in optimal conditions by the highest institutional and security organs to continue executing his political role, and through a series of laws that stabilised a precarious political balance (between pro-Russian and anti-Russian) in the government. In this case, Volodymyr Zelensky created an authority over time, built up year after year through his television programs (like John Cena), but he has also been 'demanded' to stay by the population and institutions that trust him. The fact that he refused to flee the war, leaving the country alone, enhanced his political and institutional status.

Authoritative power (meaning that each influencer has authority, i.e. an influencing power) was used by actors in the political sphere.
During the Russian invasion of Ukraine, this kind of power is seen to either improve or worsen conditions in one's own state, in some cases tilting the needle in favour of those who initially seemed doomed.

1.1 Social network, network science and social science

How can an influencer be considered in a scientific way?

The influencer concept comes from marketing and has been used to describe emergence of independent actors who shape attitudes through social media in competition/coexistence with 'professional media' [11]. The term *influence* has been applied to the highly organised leveraging of social media to cultivate neo-liberal individualist 'self-branding' [12] and production of 'Social media influencer capital' through third-party endorsement shaping audience perceptions [13, 14].

Influencers will use their authority to initiate a process of social influencing. Social influence refers to the case when individuals change their behavior under the influence of others. The strength of social influence depends on the relation among individuals, network distances, timing, characteristics of networks and individuals, etc. [15] The main problems of social influence analysis is how to quantify the influence of each user, and how to identify the most influential users in social networks. [15] Some schemes on the evaluation of social influence have been topic-oblivious, but social influence can be measured also by counting how much information related to a topic may be propagated in the networks. Some schemes have been based on pairwise influence, which is defined based on social ties and interactions between users.

An example of social influencing is the Palessi shop event [16, 17, 18], where shoe store in Los Angeles (CA) brand known for budget-friendly shoes, opened a fake pop-up store and invited influencers to the grand opening. The store was stocked with Payless shoes in disguise. About 80 influencers attended over two nights, according to Payless. They shelled out a total of $3,000. One shopper spent $640 on a pair of boots, which represented a 1,800% markup. Payless, however, returned their money and let them keep the shoes. Payless said the influencers were paid a small stipend to attend. Payless said the social experiment was meant to remind shoppers that Payless' affordable shoes are fashionable too. But the prank also points to a reality about the human mind: Consumers are not capable of discerning the quality and value of the things they buy, said Philip Graves, a consumer behavior consultant from Britain.

> Slap a fancy-sounding European label on $30 shoes, and you have an illusion of status that people will pay an exorbitant amount of money for. "The way that we evaluate things is through associations. If you put wine in a nice bottle, people like it more. If you package things up to look more premium, people will like it more; If advertising has high production qualities, people will think it's better." [16]

A study published in 2008 [19] tested the idea that price affects perceptions of quality. Subjects were given inexpensive wine to drink, but those who were told it was more expensive described it as more flavorful and pleasant to drink. This type of marketing, in which real customers are shocked or surprised in order to create a viral buzz, has been tried plenty of times.

But as mentioned, the ability to influence also comes from the network. To understand how an influencer starts to influence other people, it is important to know what kind of network the

communication and passage of information takes place through. Of all the most studied networks, such as the random network [20], the small world network [21] and the complex network [22], the latter network seems to comply more with the procedural dynamic, which identifies when there is the presence of a strong influencer in a network. The main hub(s), having the most connections, given by the preferential attachment, allows for a larger audience of people to be influenced, augmenting what has been said before about the new social media, i.e. since they have given more opportunities to inform and/or misinform, influencers can be used as a means of massive information and misinformation, given their ability to be heard and convinced.

Bakshy et al (2011) [23], they define as individuals who disproportionately impact the spread of information or some related behavior of interest, but ordinary individuals communicating with their friends, for example, may be considered influencers, but so may subject matter experts, journalists, and other semi-public figures, as may highly visible public figures like media representatives, celebrities, and government officials. While exploring a method to quantify influence on twitter, they figure-out that in perspective: individuals who have been influential in the past and who have many followers are indeed more likely to be influential in the future; however, this intuition is correct only on average. They found that by showing how other mid-tier influencers share and repost their content greatly amplifies the distribution of information and thus the possibility of influence. This means that individual-level attributes in particular past local influence and number of followers, can be used to predict average future influence. This also means that individuals who follow many others, or very few others, would be distinct from the average user. Likewise, one might have expected that activity level, quantified by the number of tweets, would also be predictive.

So, individual-level predictions of influence nevertheless remain relatively unreliable therefore strongly suggesting that, rather than attempting to identify exceptional individuals, marketers seeking to exploit word-of mouth influence should instead adopt portfolio-style strategies,which target many potential influencers all at once and therefore rely only on average performance.

In 2015 Lymperopoulos et al. [24] find that there is a similarity between activation neurons and social contagion. They questioning about: How can it be determined whether the output of a social simulation corresponds to the output of the investigated social process? The increasing use of the online social networks and the ample availability of empirical data are instrumental in elucidating the mechanisms of behavioral phenomena appearing in the online ecosystem. To gain more insights into their dynamics, their study adopts a neuroscience perspective and draws inspiration from the structural analogy between the brain and the online social networks. The online social contagion as a dynamical process whereby the users' activity is regulated by the dynamically evolving impact of three sources of positive or negative influence, that is, their self-generated bias, their online interactions, and the external environment stimulus. As in the case of Integrate-and-Fire neurons, the positive influence causes excitation leading to an individual's activation when a threshold is

exceeded, whereas the negative influence causes inhibition, thus deterring the activity propagation. The study of the online social contagion from an open systems perspective whereby the dynamics is not only affected by the internally generated influence, but also by the external environment. This work gave me various ideas on both an ideological and methodological level, so part of my thesis will follow and develop from this work, and in fact, it will encompasses part of the dynamics that will be developed later on. Through the proposed isomorphism, their study aims to introduce an overarching modeling approach, capturing the dynamics of:

1. The stimulation mechanism, through which the influence originating from the individuals' self-dynamics, their interaction and the external environment, diffuses in online social networks.

2. The stimulus processing, whereby the incoming influence accumulates, but also dissipates, thus dynamically changing the level of the individuals' susceptibility to social contagion.

3. The individuals' activation mechanism, by means of which behavioral changes occur.

This and other works by Lymperopoulos et al. will be dealt with in the methodological part as a fundamental part of their methodology is similar to the dynamic which I intend to initiate in the sample of the thesis.

Among the larger audience, most studied networks rely on network and information dynamics, but of course the process of becoming an influencer is a cultural and social definition, that relies on the necessity form society, to obtain different and new information. The ability to be able to serve society as an authoritative person for a given task/goal, is the social basis of what we sociologists define as 'social status', i.e. the position of 'privilege' recognised by a section of society, which gives a power of trust. In fact, in the dynamics of opinion dynamics and information, the possibility of being able to share information by trusting that that information is truthful to the person requesting it, gives additional value not only to the reading and dissemination of the information but also to the credibility of the information itself. The recombination in this process is a core part [25]; there is not just one influencer, but there are as many influencers as topics-news a substantial group need to have;

> Internet is thought to boost self-efficacy because it allows viewers or listeners the opportunity to hear individuals like themselves articulate their political views. [26]

This is how news influencer are born; understanding the need for a group rather than having to assemble an entire solution de novo, influencers combine multiple existing solutions in novel ways. They can combine and re-combine multiple times, while guaranteeing the same authority they already have.

But what is it, or what are the characteristics, that make them influencers and allow them to gain authority from individuals? According to Aristotelian reasoning, message characteristics can fall under either logos (logical arguments) or pathos (emotion, such as humor), and communicator characteristics are referred to as ethos. A large body of work, both qualitative and quantitative, has examined the use of these appeals in political communication. [27] Decades of credibility research tend to agree that the power of a message to persuade is influenced by three categories of factors: the characteristics of the source (ethos), the message (logos and pathos), and the audience.

- Ethos: Source credibility, or ethos, has been one of the most researched topics in communication. One of the main tenets of source credibility is that the communicator's characteristics influence the degree to which the audience finds the message credible and persuasive. Specifically, the degree of expertise (the extent to which the speaker is perceived as being qualified) and trustworthiness (the extent to which the speaker's statements are perceived as valid) are the two main components of source credibility. Highly expert and/ or trustworthy communicators can influence the audience into changing its attitudes and behaviors toward the issue, product, or person being promoted in the message.
- Logos: A common approach to making arguments is using a logical appeal, or logos. By using a logical appeal, an individual provides factual information and arguments to support their position on an issue. This process allows the recipient to evaluate the argument on the basis of that information and decide whether to accept the information as valid. A political example of logical appeals would be a presentation of statistics or a news story based on information in an advertisement attacking an opponent's record. These situations heavily depend on the reliability of the source of the facts, statistics, and so on.
- Pathos: Whereas logical and source-credibility appeals are most often seen in political information, emotional appeals, or pathos, such as humor appeals, are seen in alternative forms of political information. Using emotional appeals adds a different element and/or perspective to information. Humor has been associated with politics in various formats, including late-night talk shows, editorial cartoons, and television dramas. Most of the previous research deals with how shows presenting political information (real or fake) represent politicians. Entertainment media is one of the ways humor is incorporated into political information. In fact, according to a 2005 Pew survey, 48% of adults regularly or sometimes use entertainment media to get political news. Hollander (2005) also found that young voters identified and sought entertainment or humor-based programming as a source of political information.

The English et al. research question is asking whether the message appeal (ethos, logos or pathos) is related to perceived credibility of online videos [27], with their method, they have evidence that the *Ethos* received the highest ratings (M = 3.73, SD = 0.54), followed by the logos appeal (M =

3.15, SD = 0.45), with the pathos or humor appeal receiving the lowest credibility ranking (M = 3.07, SD = 0.46).
Therefore, Ethos, the source credibility, is the main value to obtain credibility and authority in online social networks.

Even so, how is it possible to use this power, hence the authority to mobilise people, and thus the human factor, in geopolitics? Given the authority over people, it is possible to make people do anything within certain limits. For example, encouraging people to perform protective behaviours in the COVID pandemic, such as wearing masks, [28, 29, 30, 31, 32] , getting people to vote for a band during the eurovision competition [33, 34, 35, 36], or voting for one political party over another [37, 38, 39, 40, 41].

This can also influence more sensitive aspects for our security, such as geopolitical, socio-economic [42] or during international meetings. For example, like the Chechen leader and influencer Kadyrov, on the use of nuclear weapon against the Ukrainian army by Russia [43, 44], or to incite new Russian recruits to go to war by showing his sons using weapons before being sent to the Ukrainian front. [45, 46, 47]

1.2 Leadership and information during wars

World War II was followed by the Cold War fought through propaganda, and cultural influence, manipulation of information and spies. Today, things do not seem to have changed; on the contrary, through social media, the situation has been perfected, with algorithms monitoring each of us. Regardless, we are in the midst of an information war that has the conquest of people's minds as its end point. This is the ultimate battleground, with over sixty per cent of the world's population connected to the Internet, which will technically reach one hundred per cent by the end of this decade. The US theorised as early as 1997 about information dominance, exercised through America's power in film, the arts, media, information, publishing, and even science, with universities and Nobel Prize winners.[48]

During the Russian-Ukrainian conflict, we are seeing the disinformation *Communicative strategies* in its full splendour. Putin could win the war on the ground, but he has definitely lost the information and propaganda war. [48] Even Polybius before Christ pointed out: 'Truth is the first casualty of war'. Indeed, one cannot ignore the educational and democratic emergency of this time, which is represented by disinformation, since the manipulation of reality is a weapon inherent in every conflict. Public opinion during the First World War was heavily biased, as historian Marc Bloch also explains. Great Britain organised a campaign of persuasion of American citizens to make them favourable to US intervention in the conflict; the Second World War stemmed from acts of blatant disinformation such as the fake attack by Polish soldiers on a German position. There was also much discussion about the Japanese attack on Pearl Harbor.[48]

With the rise of the internet, peer-to-peer networks, and social media have radicaly created radically new conditions for the circulation of and battle over (mis)information, especially concerning information about war. This is why, during the Russian-Ukrainian war, Russia and foreign organizations and countries have modified the interaction with information platforms, not only was access to social networks limited, but also sites like Wikipedia and many others.

On 2023/02/25 Roskomnadzor restricted the use of Facebook in Russia,[49] and on 2023/03/04 shut down Moscow's Echo radio for its reports on the invasion of Ukraine,[50] also threatening to block access to the Russian Wikipedia in Russia, which reported casualties among Russian military personnel and Ukrainian civilians in its entry on the invasion. [51] Two days later the President of the European Commission, Ursula von der Leyen, announced the blocking of broadcasts of the RT channel (formerly Russia Today) and the Sputnik news agency,[52] and the blocking of RT content was also applied by YouTube, Facebook and other platforms.[53] On 2023/3/05 a law came into force in Russia that punishes those found guilty of spreading false news about Russian military activities with up to 15 years in prison, causing Western media correspondents to withdraw from the country. [54] In particular, it became forbidden in the country to refer to the Russian special operation as an invasion.[55] On 2023/03/11, Meta announced a temporary policy change, allowing Facebook and Instagram users living in and around conflict nations to post content against the Russian military and relaxed censorship of posts inciting the killing of invaders and their leaders; Russia called this decision an "incitement to racial hatred". [56] On 2023/03/12, YouTube also announced the blocking of Russian state-supported media under its policy on violent events.[57, 58]

The limitation from the Russian Federation and the online platform to access social networks (mainly Instagram and Twitter), have provoked a social reaction from many of the Russian influencers, where these platforms were a huge income for them financially, or in the case of some sites where, due to the high online restrictions, pictures of missing family members were shared to ask their colleagues if they had heard from them. [59]

Or again, the tears of some Russian Instagram users, with millions of followers, are for the platform's closure, and paradoxically, not for the dead in Ukraine. And amid tears they invite followers to move to alternative social networks, with Telegram in the lead. According to Statista data [60], Instagram has around 60 million users in Russia, which represents 40 per cent of the population. Roskomnadzor, Russia's media regulator, officially banned Instagram on Friday 2022/03/11, and gave users 48 hours to say goodbye to the platform. The decision came after moves by Meta, which controls Instagram and Facebook, over the war in Ukraine and information against Russia. [61]

The key factor, of the information-war and the propaganda-war in Ukraine, is that Zelensky has succeeded in becoming a *political-influencer* all over the world, when, (assuming a safe time window for visits), the various foreign political leaders wanted to be photographed together with Zelensky, not only to support Ukraine, but also to increase the influence of themselves in their home

countries; and also succeed media-wise in gaining economical, financial and weapons support from much of the world. Especially from the USA and the UK, where by sending in state-of-the-art weapons, the Ukrainians were able to retake important territories lost in mid-September 2022. [62] Obviously, the USA did not only give key weapons to the Ukrainians for the important media or *Soft power* part, but also to neutralise or at least cripple (cease vital state functions of) the Russians, so that they would not be a problem in the short to medium term.

2 Goals

In 2015, in an Italian newspaper [63], an online fake news factory was revealed. The news caused a stir, not so much because of the novelty (since it was still 2015), but because of the subject of the news and the revenue it made from it. The fake news was very much focused on migrants and the crimes they did (mainly rape, brutal murders and torture) obviously all made up. The problem stemmed, however, not from the fact that a completely normal person could create such fake stories, but more from the fact that people believed so much that the (very small) site started to make huge revenues just from sharing the news.

The young man was earning very good money from the site, thanks to Google advertising and the likes. With every click, he was paid a few cents. The important thing that made it all work was not just the stories, but the amount of 'contacts' he could send them to. In other words, the mass propaganda was taken care of by Facebook and other social networks, where he shared and disseminated the fake articles through a whirlwind of fake and fictitious pages, which in turn had an impressive consensus (strengthened most likely by the self re-enforcing effect of echos-chambers):

> One like leads to another, and by dint of being spread, any news story can become viral, and therefore 'true'.

Obviously, in the past, algorithms that decided advertising on the web were not programmed to recognise the categories of good and bad advertising, and only recently have computerised solutions been increasing.

Echo chambers, are metaphorical situations in which information, ideas or beliefs are amplified or reinforced by communication and repetition within a defined system. [64, 65] Often, fake news gains sufficient weight and attention due to the presence of echo chamber, which amplifies both its resistance (in the sense of defending the topic) and its diffusion among people. [66] In fact, fake news does not only gain sufficient traction due to the attractiveness and fascination of the news, but is also greatly increased by the effects of echo chambers. [67, 68] This type of effect, i.e. that a

number of people begin to believe a piece of information (even if it is false) can be the result of an 'echo-chambers' effect.

The problem here is very clear, because it leads to a large number of actors being indoctrinated towards one type of culture. It is one thing, however, if a boy does it, it is another if a whole state does it. [69] During the first pandemic wave of COVID in Italy, between 11 and 23 March 2022, a hashtag #forzaCinaeItalia (long live to China and Italy) and #GrazieCina (thanks China) became very popular, by the fact that a lot of help arrived such as masks, gowns and ventilators against the COVID-19 virus. A large proportion of the tweets showed a positive effect inherent in Chinese aid [69], but even though most users praised and thanked Chinese aid (indeed, Thanks China), the problem emerges under close observation. In fact, almost half of the posts on Twitter for #forzaCinaeItalia, and more then one third for #GrazieCina, are overloaded with bots.
The level of activity, involvement (retweets + likes) and liking (likes) of the Chinese embassy's Twitter account in Rome and of the posts concerning the Chinese government's rescue operation seem to photograph a premeditated operation that has no precedent in Italy. It is no mystery to those familiar with the cybernetic space existence on social networks of bots, accounts created ad hoc to increase, through posts, likes, retweets, and mentions, the reach and effectiveness of a specific message and taking the form of an Echo chamber. The existence of a real bot market, frequently tapped by both private actors and state entities, has long been established. Specifically, the pro-Chinese bots were identified for a number of criteria. First, activity and timezone: 'The selected accounts have an average Twitter post-sharing rate of more than 50 tweets per day, with the most active one reaching 91.72 posts. This activity is to be considered automated. This is also evident from the timezone (time of publication) analysis, which is presumably false due to continuous activity throughout the day, with no breaks between night and day,' the report explains [69] . Second, the amplification rate: the activity of the accounts 'is focused on retweets and mentions. Selected accounts do not produce a high number of organic posts'. Third, follower/following: suspicious accounts are often followed by other 'automated accounts'. Fourth, political affiliation: 'The accounts are believed to refer to the same political affiliation, in favour of Chinese interventions. Interestingly, there are no references to initiatives from other countries (e.g. Russia or the US)'. Fifth, the handle: 'the composition of the nicknames is in fact alphanumeric', which proves that the profiles are 'generated by an algorithm'. Sixth and last, anonymity: 'Some accounts present long periods of absence of communication'.

In Italy, however, laws against the publication or dissemination of false, exaggerated or tendentious news, likely to disturb public order, have been passed and updated since 1997 [70, 71]. In neither case (the Italian one with the migrants, nor the Chinese one), however, did the state infrastructures, such as the judiciary, move. Although the basic law, dated 1981, is indeed old, in fact the offence is

only punishable with an arrest of up to three months or a fine of up to EUR 309.
This aspect places no small focus on the possibility of influencing people, hence the human factor, which certainly has repercussions on society and state apparatuses. Indeed, as Professor Mario Caligiuri, President of the Italian Intelligence Society, explains [72]:

> It might be advisable to start outlining a 'geopolitics of the mind', understood as the battlefield where the struggle for power is taking place, in order to exert ultimate dominance over people and nations, since beyond mind control there can be nothing else.

Over the last twenty years, cyber space has progressively expanded, which is asymmetrical by definition where states as small as territory can represent big powers, such as Israel and South Korea. From cyber space to the sixth domain, for me that of the mind, the step is short, since in 2030 technically all citizens of the world could and possibly will be connected to the Internet. Thus increasing the number of people who can be influenced and conditioned.

Scientific research [72] has shown us that our way of thinking is already to a large extent conditioned by genetics and the environment, i.e. the family we are born into and the social and national context in which we live, which inevitably shape our future, also conveying prejudices about our perception of reality.
During these pandemic years, and now with the Russian-Ukrainian war, is it evident that alongside the real war there is the war of information, which causes devastating distorting effects. To describe them, Marshall McLuhan recalled that 'what fish know absolutely nothing about is water'. The same applies to us who are totally immersed in disinformation and grasp the exact opposite of reality.

> Manipulation has become capillary and uncontainable, being used not so much for political purposes as economic, with states becoming financial entities and multinational corporations heavily conditioning democratic governments.

Says Mario Caligiuri. [72] In fact, even the Italian Ministry of Foreign Affairs, given its recent manipulation and propaganda activities, has become active on this issue. [73]

> Just like a virus, misinformation is also defeated by creating 'antibodies'. A 'method' is needed to fight it on an equal footing.

Said the Italian Foreign Minister Luigi Di Maio in 2022.

There have been some results at the methodological level and perhaps at the level of social 'immune' reaction and defence against fake news [74], as well as at the level of AI and algorithms on social network platforms (such as Facebook, Twitter etc.). [75] But the level of resistance is still not

sufficiently able to reduce the possibilities of control by external agents and thus of threat, although good signals are coming in concerning Trump's hypothetical arrest, seeing as his supporters have not given him the expected support [1]. [76, 77]

In fact, prevention (a vaccine) is much better for defending oneself against a virus (i.e. knowing how to recognise true from false news, and having an information methodology), than when the virus attacks and one must resist. But unfortunately at this time, the virus had already attacked, and part of the population did not have the antibodies to defend themselves.
This has led to cases of extreme action, such as the aforementioned Capitol Hill, and also to extreme action in places where it is not very difficult to hear about shootings and murders against a race or a politician, like what happened in Italy with Luca Traini, and in Japan with Tetsuya Yamagami.

The case of Luca Traini [78]. is closely related to the one described at the beginning of the chapter, i.e. racial radicalisation against an immigrant community. Traini took his gun and to the city of Macerata (a small Italian municipality with forty thousand people) and shot to kill, but only managed to wound several people and hit shops and buildings. One of the points hit was the local headquarters of the Democratic Party. Six people were injured in the attack, all immigrants of sub-Saharan origin aged between 20 and 32. After the shooting he allegedly got out of his car in front of the city's War Memorial, where he gave the Roman salute and shouted 'long live Italy' with a tricolour tied around his neck, before surrendering to the police. The gun attack, which was found to have been specifically aimed at immigrants, was traced back to a racist matrix. [79]

The case of Tetsuya Yamagami, [80] the man who shot Shinzo Abe [81], is different yet similar. The shooter, a 41-year-old ex-serviceman of the Kaijō Jieitai, a Japanese maritime military force, was immediately identified and arrested by the local police. Put under interrogation, the ex-serviceman, who had used a weapon he had made himself to kill Abe, stated that he had a grudge against the Unification Church, which, according to Yamagami, was linked to the former premier. He also stated that he resented the fact that his mother had been brainwashed by the religious group and that the latter brought her down, thus blaming Abe for spreading the religion in Japan.

In both cases, however, one can see a connection with the general radicalisation and the radicalisation of people, one against migrants, and the other against (according to his version) a the person, and that group of people who caused the misery of his family situation. There is therefore an emergency, in some areas of radicalism and radicalisation [82] , which can not only endanger the sovereignty of states, but also endanger society and institutions.

Since we can no longer have antibodies, my thesis is based on trying to deradicalise a radicalised group, more specifically on social networks (Telegram), where a large proportion of radicalised people are already present. [83]

[1] at least for now 2023/03/21

I conducted a search for various Telegram groups that met the characteristics required to achieve my research goals. While I came across several channels, including those focused on QAnon's theories and pro-Russia and pro-Ukraine channels, I ultimately chose no-vax groups due to their history of violent behavior against vaccination centers during the period of my research. I also specifically chose Italian groups since my thesis examined not only data filed but also social and sociological dynamics. Being Italian myself, I felt that I could better understand the unique social and political dynamics and common events from the Italian culture that could modify the dynamics in the telegram channel.

The next chapters will describe the literature on the subject (including the current state of the global security situation), case studies, and in the end, the inherent security needs of both institutions and society, which will take up all the scientific challenges and questions, which I would like to solve and show in my thesis. The challenge of my thesis, therefore, shows a challenge, *the search for information about the future for the safety of people*; there are two research concepts:

- *The methodologies that allow this to be done*, and also,
- *An extremely new field that needs to be analysed and observed in order to understand the complex social dynamics that surround us*.

I can define my research question as:

1. Is it possible to deradicalise a social group within online social networks? and measure the social impact?
2. Can an Open-AI source device be useful to the security domain?
3. Can we obtain a standard method, i.e a common requisite for each individual, to be able to read information? [2]

And and thesis goal as:

1. Observe social dynamics in radicalised group;
2. Construct an Open-AI device helpful for security institution;
3. Obtain sociological evidence of consequential importance of Influencer in everyday-life;
4. Provoke an "Organized Criticality" ($\neq$ Self-Organized Criticality) in Social system.

So basically, I would like to provoke a phase transition of the social group through influencers, not to predict, but to deradicalise people before they become dangerous [3], or to unravel the group before

[2] Popper once said "need a licence to watch television", therefore a prerequisite for every individual to be able to read information shared.

[3] Starting from certain parameters such as sentence sentiment and keywords, which can give an idea of the evolution and status of the people under observation over time. Obviously there is no absolute certainty that I can predict dangerous behaviour, but I think they are good parameters to assess possible dangerous behaviour.

it becomes organized and dangerous.
To define whether people become dangerous before they do harm to society, I will observe and respond to all those messages inherent to violent action, such as violent reactionary ones. Thus, there is no control group to define whether part of the group has a violent and reactionary behavior, because I will reply (if I observe messages instigating hatred or violence) mainly to those individual messages.

A state, in order to defend its population, cannot control each person individually (as in the case of Traini or Yamagami), but focuses more on already organised groups, which can do much more damage than a few gunshots, I am talking about the organisation of popular uprisings against groups or individuals, up to real coups d'état.
However, states already do these things. Or rather, they do not have agents to deradicalise people, but they do have agents to gather information and defuse what may be dangerous actions. So, my thesis is really about deradicalising people and trying to de-structure the radicalised group.

To do this, as mentioned above, there cannot be people applying methodologies all the time and consistently for all radicalised groups, so it is feasible, given the advent of computer technology, to structure an AI that can support the human and exercise the same strategy even when the human, for example, is asleep.

2.1 Why create an AI?

As described by Perrow with his *Normal Accident s' theory* [84]. Perrow argues that if a system has two or more failures/events afflicting its production/protection systems, and that these may follow one another temporally in an arbitrary order, then a specific sequence of events will exist such that all redundancy/security systems will fail, making a catastrophic failure inevitable.[85]
Artificial Intelligence should, in fact, be able to recognise, on the basis of the hundreds of measurements that we collect at any given moment, the correct functioning of each infrastructure as well as the onset of abnormal situations and/or attention.[85] This feature of AI could not only allow me to observe some dynamics that can be observed thanks the collected data, but also to evaluate them "not for the sake of the system itself, but to understand when to attack" and deradicalise.

The AI being able to collect all the real-time data of the various interactions, including also the content of the interactions (message), will give me the possibility to create the network of interactions within the group. When one of these interactions plays an important role (like inciting violence for example) I can both recognise it since I know the interaction and the content of the interaction, and I can act. The AI therefore does not intend to calculate all the hypothetical interactions between the subjects in the sample, but to automate and analyse the interaction concatenation in the sample in real time, so as to highlight the main and re-enacting edges (i.e. interactions) to initiate the deradicalisation process. Hence, a substantial difference between Perrow's theory and my type

of AI, lies in the fact that in Perrow's there is already a network for evaluating the two or three failures/events, whereas in mine, the evaluation of the two/three failures lies in recognising the interactions (thanks to text data) that amplify the negative (danger) potential of an individual, such as incitement to violence, for example.

Perrow's speculations have unfortunately found empirical evidence in the major incidents that have occurred this millennium, starting with the blackout in the United States in 2003, the origin of which, as illustrated in the report jointly drawn up by the US and Canadian governments, is due to an incredible concatenation of events occurring on the electricity grid (grid events), the monitoring and control system (computer events) and the actions taken by operators (human events). It would have been sufficient for just one of these events not to have occurred, i.e. for the time sequence of events to have been even slightly different to prevent what was the largest electricity blackout in the world, with a loss of 61,800 MWatts of electricity suffered by more than 50 million people on the east coast of the United States and Canada, with economic damage estimated at between 4.5 and 8.2 billion dollars.

The AI influencers can do such a thing, catching and locating the trouble spots in the system, to avoid possible social threats. [85] Part of my job, however, is not to convince and deradicalise people with false information, but to share truthful information, removing meta-information such as the author and headline of a newspaper for example, so as not to induce stereotypical reader behaviour. This is because, some newspapers in Italy are not independent towards the political class in general, but are aligned. So if I share information from a newspaper that is openly sided on a subject, I can trigger reactions that do not help create the influencer and thus the process of deradicalisation. So, starting from the principle that I can share any information within the group, I do however have obvious legal limits. Under Italian law, I cannot incite hatred or utter false alarms or exaggerate news to disturb public order. Both on a real and cybernetic level.

But, back to Perrow's book, the inspiration for "Normal Accident's" was the 1979 Three Mile Island accident, where a nuclear accident resulted from an unanticipated interaction of multiple failures in a complex system. The event was an example of a normal accident because it was "unexpected, incomprehensible, uncontrollable and unavoidable". Perrow concluded the failure's analysis at Three Mile Island, was a consequence of the system's immense complexity. Such modern high-risk systems, he realised, were prone to failures however well they were managed. It was inevitable that they would eventually suffer what he termed a 'normal accident'. Therefore, he suggested, we might do better to contemplate a radical redesign, or if that was not possible, to abandon such technology entirely. [86]

In any case, there is no methodology to handle all the various levels of warfare; therefore, it is a cognitive-organisational problem of the people in command. This leads to a higher probability of hitting the nuclear red button [87] , if the leaders in command are not trained and adequate for their

task, like the leaders of a Telegram group in terms of knowing how to defend themselves with their authority and status within the Telegram group.

Another very important factor, or key factor to the question, why "create an AI?", however, which is always related to the human factor, is the person's resistance to analysing deviant content. The continuous analysis of content that does not conform very well to the ethical or value level of the person examining, can cause very high stress, and a withdrawal from the work itself. Indeed, in cases of continuous analysis of child pornography or violent content, for example, analysts can find themselves in a dramatic mental and physical situation.
For this reason, an AI-influencer, manage (even partially) the social dynamic, in the face of shared deviant content, is also a very good reason for the safeguarding of the analyst's psycho-physical health.

Therefore, the human factor is the key part of group dynamics, whether open and social or not. The possibility of a leader being able to lose what he has done, and his authority, will push him to lose everything, thus going 'all in', and this is where the bots controlled by me can make a difference, deceiving the person and driving him to the margins of the group.

2.2 Complexity technique approach

In the last decade, social science thanks to the evolution of computer science, obtained the power to observe social phenomena in a different way. The data gathered from social network, and from everyday life, empower the social method to describe and evaluate recent and past social events [88].
Nowadays, thanks to this rapid progression, social science obtained valuable social information that entrust social method, and several researchers debated over social dynamics, and one of the main outcomes is that *"a natural science model of the research process is suitable for studying the social world but a central issue remaining of whether the social world can, and should be, studied according to the same principles, procedures, and philosophy as the natural sciences"* [89].
Thanks to the observation of complex events on social systems, the recognition and collaboration between social and physical science have started with exceptional results [90], establishing what we have already mentioned, i.e. the *Computational Social Science*.
A certain amount of social complex research, focusing on the interaction between individuals and individuals, or individuals and society, in the political, economical, management, and epidemics sector (and many others) with a computational social science methodology, have obtained importance and regard from the scientific community [91].

Today, public discourse around current events, especially those that occur online, are increasingly influenced by the issues that news media companies choose to concentrate on [92], and from the social media influencers that provide authority on that topic [93]. The authority of the influencers

and certain initial condition of a social system, can generate a Self Organized Critically event (SOC) [94], bringing collective intelligence social action.

Spreading information about some topic or encouraging social manifestation, are common actions and attitude of the online activity. But spreading fake news [95] or encouraging dangerous manifestations, like the Capitol Hill riot, could be a possible threat or corrosive to the democratic institution. Furthermore, on geopolitical affairs, rival governments will always try to take advantage of a moment of weakness, to exploit it to their advantage. Like the United State political election in 2016 [96], and the Chinese aids to Italy during the first wave of the Covid-19 pandemic [69].

People can self organize and become possible *Social Critical Error* (SCE) like a social threat for institution or complex organization. But usually this event arises due to Collective intelligence event or SOC event. A normal event becomes a SOC event due the collapse of many and small and/or micro events on the network, caused by a single (and sometimes unrecognisable or unimportant event) but strong enough to create a cascade collapse of the entire system.
These events are present in physical science (like headquarters) and in social science (like coup d'etat).

Examples of SOC-type manifestations were those in Chile where, because of a small increase in the price of metro tickets, people took to the streets en masse. [97, 98] Those small and/or micro events on the social network are hard to detect and analyze for humans, because there are so many, and often irrelevant for human eyes. For this reason, some natural events are, and will be for a long time, unpredictable, while others, such as earthquakes for example, will be more predictable as technology advances. In table 1, I have listed some examples of natural and social SOC and SCE systems.

Due to this challenge, programming an AI can detect those micro events, and notify possible cascade events on the observed network. In the social network field, we can define a possible micro event, the sum of information and value from society. For example, the sentiment of the population in a specific day/time, vote preference, economic status, shared information, etc, are all valuable input to train the AI on the detection of the social micro-event.
The social sector chosen is the social security field and all related topics (radicalism, terrorism, riot, etc). The social value chosen is easily acquired data due to the high presence on online social networks such as Telegram, like:

1. Information shared on social network (posts, texts and replies)

2. Sentiment status from people that write and share this information

Table 1: SOC & SCE Categories

SOC/SCE	Name event	Force	Dynamics	Outcome
SOC	Send pile	Gravity	Sand moved by gravity has two options: to remain in the same layer or to roll down, potentially carrying other sand with it	Continued option two creates waves
SOC	Earthquake	Gravity and Friction	Excessive pressure from the Earth's crust can cause layer movement and seismic waves until equilibrium is reached.	Layers pushing create earthquakes
SOC/SCE	Manifestation or Riot	Social Need	Similar to a sand pile, a "gravity force" arise when certain limits are exceeded in different domains, such as economics, politics, or society.	Riot, manifestation or Coup d'état
SCE	Coup d'état	Political ideology	Self organize people, institutionally supported individuals in power can enact regime change after a triggering event surpasses the agency's tolerance threshold.	Coup d'état
SCE	Terrorist attack	Ideology	Carried out with the will of an individual or groups that have exceeded their limit of tolerance (inherent to social value and/or social norms). Attacks can cause severe physical and psychological harm to victims and the community.	Terror and violence

Echos chambers, Telegram or Facebook groups are exceptional places where terrorists, rebels, anti-establishment or revolutionary people are active in order to gain popularity or followers (devotees peoples).

The goal of my thesis is to build an AI that helps to detect possible micro social events and tries to deradicalize group of peoples, avoiding SCE. But the social science and the social dynamics studies can help us also to try to mitigate the SCE, thanks to the characteristic of continuous monitoring of the interactions between the various actors in the field (especially the most important or dangerous ones); the type of shared messages that can initiate social dynamics; the social status that these people have within the group (they can influence the masses) and finally to understand when the combination of these factors at a given time (initial conditions of system analysis) can trigger

dangerous events or SCEs.
One method to understand the evolution of SCEs events could possibly be the Kuramoto model. [99, 100] The Kuramoto model is a mathematical model that was originally developed to describe the synchronization behavior of oscillators. However, it has also been applied in social science to study the dynamics of social interactions and the spread of opinions in a group. By analyzing the interactions between individuals and the way their opinions influence one another, the Kuramoto model can help to predict social events such as the emergence of social movements or the spread of new ideas. The Kuramoto model takes into account the social and technical dimensions of the interaction, modeling the complex relationships that exist between individuals in a social network, as well as the way that information propagates through the network. The model can be used to analyze the way that opinions spread through a group, and to predict the emergence of social phenomena such as political polarization or the rise of extremist groups. By using the Kuramoto model and AI technologies, social scientists can gain insights into the way that social networks operate, and develop strategies to influence opinion dynamics and prevent the spread of harmful ideas.

Bur concerning the deradicalisation process and the goals of my thesis, an AI-Influencers can mitigate possible social threats, due the characteristic of gain authority (due to the fact that people give to other people high reputation and control over this micro society) to my accounts giving the helps of my bots, and catalyse their authority on information shared.

In a more specific description, the approach and procedures to obtain an AI capable of understanding social information and sending specific messages to individuals are:

A. Create bots (for data collection and interactions) and let them enter in the various channels/groups

1. Accounts are created to collect data and to interact with other people

B. Collect data to:

1. Analyze the social sample
2. Discover Influencers on the group
3. Create social ideal-types or social models from people in the channels/group [4]

C. Structuring the AI-Influences based on:

1. Type of information assimilated by the participants (articulated or simple)
2. Sentiment and/or emotion detection from topics and individuals
3. Trend topics and argument

D. Structuring data-based communication strategy on:

1. SOC
2. Method to become an influencer (Social simulation)

[4] Based on the characteristics of social status, social role, values and social norms.

3. Authority (of the AI-Influencers) derived from individuals

E. Send messages on selected topics

1. Observe the fluctuations on systems
2. Gain respect for those persons (AI) through the messages (D.3)

F. Those AI account have gained respect and credibility/authority and they cannot be betrayed or treated badly.

G. Begin the anti-propaganda and end the social threat.

People in social networks need to gain respect from people (AI) through messages. The message must get across to people, and that person cannot be betrayed or treated badly, because it elicits emotional and/or rational credibility.
I play on the person's credibility to destabilize the social balance, like the deradicalisation program in prisons, shown in later chapters, and have a phase transition that leads to people no longer following that person (or following another). This, however, is based from successful socio-political programmes, but is not an evidence-based approach from the social network part. In reality, steps have been taken to deradicalise people, but none have been started in online social networks.
My thesis, therefore, is a novelty in this area.

But as online social media becomes a major channels for the diffusion of news and information, it becomes critical to understand how the complex interplay between cognitive, social, and algorithmic biases, triggered by our reliance on online social networks, makes us vulnerable to manipulation and disinformation. This obviously limits my cognitive capacities for judgement and action, as I myself can be manipulated by the information shared and the dynamics created in the sample under scrutiny. But, this is also a positive factor, as not only is there a lot of research to be done on how these two events (manipulation and deformation) occur with current technology, but also whether the tools which I will make available to understand this phenomenon are adequate, limited or outdated. This talk overviews ongoing network analytic, modeling, and machine learning efforts to study the viral spread of misinformation and to develop tools for countering the online manipulation of opinions. Filippo Menczer, in his study called "Hacking Online Virality" [101, 102] have studied the dynamic of online social media, and he shows how to *Hack it*. [103] For example, he showed how 'Unfriending', i.e. being offended by someone sharing that thing, leads to that person no longer being followed, creates chamber echos. Or how this principle can lead to bots finding themselves alone in echos chambers. Or even as 'Follows trains' (i.e. bot-accounts following each other to increase the chances of being followed by others (also fooling part of the system as on Twitter), can lead to the destruction of Echo chambers. Instead, his studies have shown that when there is so much information in a network, the quality of the information and its virality is very low. [103] Or that people who share false information often come into contact with people who share fake information, showing, that it is people who share fake news the most, but it is bots that create it.

Thanks to his information, I have come up with another solution, in addition to deradicalising the group (in case I fail or the results are minimal), which is to 'Hack' the Telegram group itself. That is, being able to greatly reduce the number of people in the group, such as instigating it to go to other groups (controlled by me) because the leaders (hence influencers) of the past group are no longer credible.
In fact, as will be further described and explained in the methodology chapter, the dynamic I intend to apply to the group is very similar to PSYOPS operations. From a research ethics point of view, however, in this case I am not delegitimising the person with the most authority within the sample, precisely because it goes against my AI-Influencer strategy. The reasons, are simple, as he is the person with the most authority within the sample, I initially do not have enough support (from the other members of the group) to go against him, and I am not going to delegitimise him, but the arguments he poses, always giving a coinciding tone to the dynamics of becoming an influencer. You cannot become an influencer if people do not like or respect you (especially if they hate you). As mentioned in the previous chapters, people, in order to call you an influencer and thus have a potential influencer decision towards them, must give you trust and competence for what you do or say.
Obviously, in order to explain how I intend to do this, in both the deradicalisation part and the 'hacking' part, there needs to be a strong IT methodological presence, highlighting the tools useful for one purpose over another.

3 Literature about political influence and security

3.1 Cambridge Analytica and profiling

The Cambridge Analytica case has highlighted how profiling and data acquisition is able to influence certain social dynamics, including political elections and individuals' preferences. In this context, it is useful to take a closer look at this topic, which has implications both on purely commercial and technological issues, but also on important socio-political issues. The Russia gate played an essential role in the political campaign surrounding Donald Trump in the 2016 election. In purely media terms, the Russia gate showed how the correct profiling of users in social networks allowed news and information to be conveyed in order to favour the Republican party during the months leading up to the elections.

The electoral consultation that sanctioned Brexit was not immune to profiling techniques, although according to a report [104] by the Information Commissioner's Office (ICO) on the potential influence on British 'leave' of Cambridge Analytica, its parent company SCL and Global Science Research (who obtained data from Facebook users and their friends via a quiz app) it is reported:

> "it appears that the actors involved may have considered targeting (profiling) British voters, but abandoned the idea".

Furthermore, European Information Commissioner (ICO) Elizabeth Denham [104] writes in this regard:

> "from my review of the materials retrieved from the investigation I have found no further evidence to alter my previous view that SCL/CA were not involved in the EU referendum campaign in the UK - beyond some initial enquiries made by SCL/CA in relation to UKIP data in the early stages of the referendum process".

The role and degree of influence that organisations similar to CA can assume are considerable, especially if they are able to convey the right information to the right people, in order to bring about a political paradigm shift in the short to long term, assuming that there is a need for extensive organisation, funding, and data to properly organise both the message and the people it is intended for.

If the action is carried out by an organisation, or rather a foreign actor, in connection with election campaigns in a given state, with the aim of influencing (more or less) the election result, as was the case with the "IRA" (Internet Research Agency - ed. Russian agency based in St. Petersburg engaged in online propaganda services) during the US presidential elections, this falls within the scope of national security, as there is a real risk that the democratic competition of elections will be distorted, undermining the democratic processes based on the principle of self-determination of

voters' political choices. A subtle, but in certain cases determined, loss of sovereignty could be on the horizon.

The document produced by the US Department of Justice's investigation into possible foreign influence in the 2016 US election identified two different forms of connection between the IRA (Internet Research Agency - Russian company) and members of the Trump campaign. The investigation did not identify (as not mentioned) similar links between the IRA and the Clinton campaign. Firstly, on several occasions, members and associates of the Trump campaign promoted through retweets or similar methods of reposting, pro-Trump or anti-Clinton content from the IRA (the January 2017 US intelligence report described the IRA as a troll factory) [105] . Content was specifically created from IRA-controlled social media accounts, posing as profiles linked to the Tennessee Republican Party, anti-immigration groups, Tea Party activists, Black Lives Matter protesters, or other US social and political activists. Second, in some instances, IRA employees posed as Americans to communicate with members of the Trump campaign in an attempt to seek assistance and coordination in IRA-organised political rallies within the United States. IRA employees monitored the reactions of the Trump campaign and, subsequently, Trump administration officials retweeted their tweets (including some tweets about events organised by the IRA itself). Beginning in June 2016, the IRA contacted several Americans affiliated with the Trump campaign in an attempt to coordinate pro-Trump rallies within the United States. In all cases, the IRA contacted campaign staff claiming to be US political activists working on behalf of a conservative grassroots organisation. The IRA contacts included requests for signs and other materials to be used during the demonstrations, as well as requests to promote the demonstrations and help with coordination logistics. In summary, the investigation established that an entity with a base of operations in Russia (a connection to the Kremlin was not proven in court) interfered in the 2016 presidential election through the social media campaign by means of 'active measures' conducted by the IRA, which is an organisation financed by Evgenij Prigožin [106, 107] (a Russian businessman who came under the crosshairs of the US Treasury Department) and companies controlled by him.

As further explained in the American Justice document [108] on Political Influences (Volume I, Section VA), it is concluded (and the jury affirmed) that:

> "although IRA members had contact with individuals affiliated with the Trump campaign, the prosecution did not show that any Trump campaign official, or any other person in the United States, participated in the conspiracy. This is because the investigation did not identify evidence on the basis that any US person who communicated with the IRA knew they were talking to Russian nationals involved in the criminal conspiracy. The Federal Bureau therefore determined that these individuals did not have the requisite knowledge or criminal purpose to charge them in the conspiracy to defraud the United States (Count One)".

The same document [108] reports,

> "On 16 February 2018, a federal grand jury in the District of Columbia indicted 13 Russian nationals and three Russian entities, including the Internet Research Agency (IRA) and Concord Management and Consulting LLC (Concord), for violating US criminal laws in order to interfere with US elections and political processes. The federal prosecution charges all defendants with conspiracy to defraud the United States (Count One), three defendants with conspiracy to commit bank fraud and wire fraud (Count Two), and five defendants with aggravated identity theft".

Part of the work done by the IRA was to create the content and share it, but this is only part of the activity. A lot of the work lies in the correct profiling of people, i.e. finding through the data on social networks, those people who are most sensitive to certain issues (who do not let the subject matter lapse), and those key people who can most influence the community they belong to. The correct profiling of people plays an essential role in the posting of content on social networks. The possibility of exercising 'manipulative' control over the conveyance of information, is an issue to which the security of every nation must respond [109].

3.1.1 Political influences from social network profiling

A company that knows what its customers want and understands their tastes perfectly well, why should they not rush to create a product/content that satisfies them and perhaps expand their market? Whether we are talking about economics or politics, the converse is similar, just get the right information, then you will be able to trigger a series of social reactions that will optimise the matching of supply and demand in the political/economic market.

Trump's presidential aides effectively implemented a communication strategy to influence the political preferences of the population by sending personalised messages to social network users in the swing states, probably succeeding in 2016 in changing the balance in the disputed states and winning the election. Trump's campaign contributors were not the only determining factor, as the US Department of Justice has shown. Very important in the case of the 2016 election is that the IRA was present in social networking groups antagonistic to their Republican candidate, with the aim of disseminating and proliferating fake news within the opposing political groups. See the cases of Pizzagate and QAnon, far-right conspiracy theories during Trump's election in 2016 and later revived in 2020 [110, 111]. In recent years in conjunction with the emergence of profiling processes, a new political argument called 'populist drift' has arisen. Populism, which in some of its parts (tends to be right-wing) also refers to sovereignism, is often exercised during the Brexit election phases by Nigel Farage. Sovereignists have been active in many national elections, proving that it is not a temporal concept, but a political value that can be referred to, in the sense that it is

not a single cultural reaction in a single state, but being a sentiment present in other nations it can be defined as a global cultural expression caused by the same need/matrix.

But in this argument, profiling played an essential role, as the shared contents (both in Italy and abroad) led to the generation of a sentiment of fear, mistrust and negativity towards European actors. In this context, certain events to which the institutions of the European Union failed to provide a concrete response, such as with immigration, or with Greece in the economic crisis of 2009, caused a negative imaginary towards the European community to be established. But this does not justify that it was only these events that triggered a populist drift in Italy or European countries, as the movement was already in turmoil (Ex-England-Farage, Germany-NPD, Hungary-Orbán etc.). This also stems from the content of the messages conveyed in social networks in recent years and to whom those contents were sent, in fact as agendadigitale.eu describes security [112]:

> "profiling and targeted advertisements are very effective if the message you want to pass is one of fear or anger, which are levers to which people are very sensitive, as for example happens in populist messages. And this explains how Cambridge Analytica in a period of growing populist movements has made its fortune."

In the United States, on the other hand, populism is personified by Trump. Some of his choices, both political and politico-economic, have been questioned by the main assets of the US state, such as the decision to appoint Michael T. Flynn as national security advisor. On 10 November 2016, President Obama warned President-elect Trump against hiring Flynn. During their meeting in the Oval Office two days after the election, Obama expressed 'deep concern' about hiring Flynn in a sensitive and high-level national security post. On 18 November 2016, Flynn accepted Trump's offer for the position of national security adviser. Prior to his appointment, media sources, including the Washington Post and Associated Press, had already criticised his close relations with Russia, and his acceptance of anti-Clinton conspiracy theories, sharing fake news in the 2016 presidential campaign [113, 114].

The hearing before the Senate Select Committee on Intelligence, 115th Cong. 13 (1/11/17) with testimony from Colin Stretch, general counsel of Facebook, on the Influence of Social Media in the 2016 U.S. Election, testified[115]:

> "we estimate that approximately 29 million people received content in their news feeds directly from the 80,000 IRA posts over the course of two years. Posts from these pages were also shared, followed by people on Facebook, and as a result, three times as many people may have been exposed to a story that originated from the Russian operation. Our best estimate is that around 126 million people may have been served content from a page associated with the IRA over the two-year period".

This content has favoured Trump and his campaign, but it has also built the loyalty of a large group of people called Proudboys by defending him in all his actions, even going so far as to storm the US Senate on Capitol Hill against the certification of Biden's election victory on 6 January 2021.

The threat from foreign infiltration both in political elections, but also in simple everyday life, must be managed from a geopolitical point of view, since if other actors have observed that in certain cases there is room for influence in this sphere, others will also begin to do so, considerably reducing the protection of the national interest, increasing the degree of uncertainty and complexity at the international level, considering the possible impacts on national security, but also on international socio-economic and geopolitical strategies between allied and non-allied countries, given the concrete possibility of destabilisation of the attacked country. Cambridge Analitica after the scandal has rebuilt itself into other companies called Emerdata (acquiring CA's data) [116] and Data Propria, which could imply that even after the famous FB-CA scandal there will still be foreign organisations that can with their activities bring repercussions and influence to sovereign states, as the IRA did in the 2016 US elections. The sharing of media content cannot be defined as propaganda, but has to be framed in a different perspective, it has to be defined as inspiring values through shared content, capable of bringing out social behaviour and actions. The purpose of these actions depends on the content of the message, but the medium (i.e. the ability to transmit content) like all technologies is defined by the use made of it.

The illicit collection of personal data by Cambridge Analytica was first reported in December 2015 by Harry Davies, a journalist for The Guardian. The CA scandal will happen in 2018. CA-enforced election campaigns started with data collection. To do this, CA used a Facebook-compatible application where, if a user answered a survey, a certain amount of money was paid out. For the payment of the sum, users had to log in to the site with their Facebook account and approve access to the Facebook-linked app. The APP was developed by Aleksandr Kogan, the Cambridge University academic known for personality profiling research using Facebook likes. For one survey user, the payment process was quick, as after completing the short survey, a payment code was provided by clicking on the payment button. But in those few seconds, two very important things took place. Firstly, the survey-app collected as much data as possible on the user who had just logged into Facebook, through the provision of even personally identifiable information such as real name, location and contact details, something that was not detectable through the survey sites themselves. Basically, *this meant one was able to take inventory of the profile and relate it to a natural person [who is] matchable to the electoral register (Christopher Wylie)*. [117]

Moreover, the app did the same for all the user's friends who installed the app. Suddenly, the hundreds of thousands of people paid a couple of dollars to fill out a survey, whose profiles are not immediately in the clear, become millions of people whose Facebook profiles are an open book. This is where the final transformation comes into play. How do you turn a few hundred

thousand personality profiles into a few million? With a lot of computing power and a huge matrix of possibilities. Even if your sample size is 300,000 people, if a Facebook member likes various posts as one of those 300,000 people, you have a chance of being very similar to him. At some point, psychometric targeting moves into the realm of election campaigns, by means of two strategies [117, 118]:

- The first: images are exploited where only the people most sensitive to a topic could understand the message. For example, images of walls proved really effective in the campaign against immigration. Conscientious people like structure, so for them a solution to immigration should be orderly, and a wall embodies that. You can create messages that don't make sense to some people but make a lot of sense to other people. If you show that image, some people will not understand that it is immigration, and others will understand it immediately.

- Secondly, strategies are used to defame members of the opposition. A series of videos made with a hidden camera without the knowledge of the interlocutors published in March 2018 showed the CEO of Cambridge Analytica bragging about the use of prostitutes, bribery and other methods to discredit opposition politicians on whom he conducts research and analysis for election campaigns. Nix (CA CEO) also claimed that the company "ran the entire campaign on Donald Trump's digital platforms", including possible illegal activities.

This video was one of the most infamous accusations against CA's criminal acts during election campaigns. Cambridge Analytica during the US campaign applied a numerical-election strategy, i.e. it gave a greater focus on the most electorally poised states. Winning by just a few votes in a given state could sanction the victory of those 'big voters' and thus shift the needle between one candidate and the other. The strategy was simple, steps 1 and 2 described above, were channelled into a few key states, so as to achieve victory with the big voters. Through these methods, CA was able to give support for Donald Trump's victory during his first election.

The society distanced itself from what the Cambridge Analityca scandal really was, simply describing the video where it confirmed the use of prostitutes and bribery, as something from old politics that happens from time to time. While in reality, the scandal uncovered a [119, 120]:

> "New kind of weapon"

as Brittany Kaiser (CA's former director of business development) called it during her interrogation to describe the work done by CA, but also to justify the very concept of that AI as a real soft-power weapon.

3.1.2 Profiling

Profiling analysis consists in obtaining (through a process of data collection and processing) knowledge of a group of individuals or an individual, including habits, preferences and information, obtained from 'digital' interactions with reference to political, musical, social issues, including the network of friendships and acquaintances, and much more. This information makes it much easier to convey content, of any kind, as it is possible to understand who is most susceptible and interested in the various topics by influencing their weaknesses, fears and interests. Furthermore, it is possible to know who that content might still propagate it in their network of acquaintances, exponentially increasing the outcome if the person in question is a political leader, social leader or representative. These mechanisms drastically improve the communication chain and the word-of-mouth of content on the web. This can be done through two steps:

1. The collection of data from social networks (Data gathering)
2. The processing of analysing social data (such as Psychometrics)

By performing the search correctly using the appropriate tools, it is possible to obtain an up-to-date list of people and their interactions on social networks. But in order to obtain the preferences of individuals, the data must be processed using social, psychometric and computer parameters. Then we enter the second phase, social data processing, i.e. the process of inspecting, cleaning, transforming and modelling data in order to derive and highlight information. From social network data, it is possible to obtain individuals' preferences from the individual-individual and individual-content interactions in social networks, such as:

1. Number of views of a page or content
2. Number of interactions (likes - comments - sharing) of the content
3. Content of comments and posts (keywords, #hashtags)
4. Content viewing time
5. Number and type of content searched

To these, which are the classical methods, one can add the psychometric parameters that Cambridge Analytica used to profile during its activities, called OCEAN:

1. Openness: which assesses how open a person is to new experiences
2. Conscientiousness: which measures how much a person is a perfectionist or not
3. Extroversion: considers the sociability of the individual being examined
4. Agreeableness: takes into account a person's willingness to cooperate with others
5. Neuroticsm: considers emotionality and the propensity for happiness, as well as sadness

The set of these parameters makes it possible to uncover a set of preferences of the individual, so that we are no longer acting on a mass communication method, but going to address the right content (specifically created) to the right person. [5] The social science used by Cambridge Analytica to evaluate and measure stimuli is Psychometrics, a part of psychology that deals with defining parameters and tools for evaluating and measuring social stimuli in a given context. [121] The effectiveness of psychometrics, however, has been strongly criticised both in its efficacy and in the manner in which the impact has been assessed, i.e. whether it was a causality or a correlational effect. In any case, Psychometrics has become well known to the public due to the Cambridge Analytica (CA) scandal, because by using the collection and analysis of psychometric parameters, was used to influence the US election campaign, and considered to influence other important political events such as the Brexit (although this has not been fully proven, and has been denied by the EU institutions - as per the EU- ICO Commissioner's note). It's important to note that the scientific community has been highly critical of psychometric methods, and questions have been raised regarding their effectiveness. As such, the use of psychometric methods should not be taken to imply a certified correlation between these methods and social dynamics.

3.1.3 Destabilizing Governments with Information

Already the case of Capitol Hill in early 2021, and the Proud-boys showed how the power of social networks, on the part of an important political influencer, can lead to riots or even coups d'état.

> Former US President Donald Trump orchestrated the storming of the Capitol Hill on 6 January 2021 in an 'attempted coup'. This is what emerged from the first meeting of the Congressional Investigation Committee on what happened on Capitol Hill. There is no room for debate at this point, Those who invaded the Capitol and fought law enforcement for hours were motivated by what President Trump had told them: that the election had been stolen and that he was the legitimate president. President Trump summoned the crowd, rallied them, and ignited the flame of this attack."

said Liz Cheney, Republican vice-chair of the Capitol hill committee. [122] The focus on propaganda influence at the political level has not abated since the legal issue of Capitol hill and Donald Trump has not yet ended, but of course, this focus still continues in 2022, when Russia invades ukraine.

The Russian invasion has given a boost to European efforts against propaganda, bots, fake news and deep-fakes. After years of debate and inaction, a real breakthrough in the fight against online disinformation seems to be coming. On Thursday, the European Union will publish a document that within a few months will force Big Tech companies to take stricter and more transparent action on bots, fake news and inauthentic content, in any language. On pain of heavy penalties in case of non-compliance. [123]

[5] The debate on the definition of psychometrics as a science is still ongoing.

A recent proof of interference by foreign states in other countries' national elections occurred recently, where *Yevgeny Prigozhin* (one of Vladimir Putin's loyalists) claimed that Russia 'intervened, interferes and will interfere' in US elections.

> "Carefully, thoroughly, surgically and in our own way, as we know. During our targeted operations, we will simultaneously remove both the kidneys and the liver,'

the businessman said. [124] However, this news, close to the US mid-term elections, on a geopolitical level, also has a characteristic not of trying to change the election as in previous elections, but of establishing a climate of tension and control from Russia towards other states, which can undermine what is the psychic resistance of society, and therefore of the human factor (who goes to vote).

Giuseppe Moles (Mario Draghi government's undersecretary of publishing), in jun 2022, announces a hybrid information war against Italy by Russia. [125]

> There is no doubt that we are facing a hybrid threat, which responds to external influences. There is an infodemic risk that we cannot underestimate. In this vast sea of war news, selecting reliable ones has become a professional's task.

The Italian Parliamentary Committee for the Security of the Republic - Copasir, chaired by Senator Adolfo Urso (FdI), released its annual report on the activities carried out in the period February-August 2022. The document analyses the geopolitical context, focusing on a series of crucial issues: from the conflict in Ukraine to the process of ecological transition, seen as an opportunity for Italy to free itself from foreign dependence; from the recent escalation of tensions in Taiwan to the posture assumed by Russia and China in other sensitive areas of direct interest to Italy, particularly the Balkans and North Africa.

> "We just have to understand that today we are experiencing a phase of permanent disinformation, not limited exclusively to the election campaign, and that it does not only come from outside but is structural to the media and political system," Caligiuri continues. 'In the fluid digital society, it will be increasingly difficult to distinguish the good from the bad and the guards from the thieves'.

says Prof. Mario Caligiuri, president of the Italian Intelligence Society [126].

"This recalls," he concludes, "the real Italian emergency, which is the educational one, since seventy-five per cent of voters do not understand a simple sentence in our language. It is no coincidence that the low level of education may increasingly represent a problem of national security and the reliability of the real nature of democracy in our country'.

In the many points addressed in the report, Copasir draws attention to the hybrid warfare and disinformation strategy implemented by Moscow and Beijing.

> "The editing and dissemination of fake news, social campaigns and the use of trolls are unfortunately establishing themselves as sophisticated and pervasive tools of influence by Russia, but also by other state actors, risking to pollute and distort the public debate in Italy and in Western countries and to convey false and unverified news that does not adhere to the facts and events".

The report identifies as means of hybrid warfare those 'communication strategies articulated in disinformation, misinformation and media manipulation; in data-driven strategies that make use of the 'targeting' systems of large social media companies; in the use of trolling or harassment directed at users' digital profiles; in mass reporting of content and accounts, indirectly exploiting the filtering systems of platforms'.

After reaffirming the primary role of intelligence agencies in identifying and combating hybrid threats, Copasir stresses the need to introduce mechanisms to hold social networks accountable 'as they constitute a venue and channel of information so that by means of automatic mechanisms or algorithms it is possible to prevent the dissemination of messages of hatred and violence or manifestly unfounded news'. The report also suggests 'evaluating the deployment of countermeasures and guidelines within a clear national security strategy that can identify the economic, technological and social vulnerabilities that need to be overcome so that they are not exploited by the adversary'. Lastly, it is recommended that the level of alert be raised by all authorities to 'avert possible forms of conditioning the electoral campaign ahead of the 25 September vote, through hybrid operations, disinformation and cyber attacks'.

3.1.4 Theorising the 'Security Influencer'

States obviously have an organisation that analyses the level of threats to their own state and its allies. These organisations, however, may have become particularly alarmed not since 2021/22 with the events on Capitol Hill or with the advent of the war in Ukraine, but the national security concern involving many state apparatuses begins with the annexation of Crimea.

The annexation of Crimea was initiated through the use of soft-power methodology and human factor. The human factor (which will play a pivotal role in the next chapter) used by the Russian federation, clung to Russian-speaking values in Crimea to exert cultural and social pressure, which resulted in the referendum being won by the Russian supporters without any particular problems. This, however, led to the identification of a hybrid-threat, i.e. according to NATO [127] a methodology capable of adversaries 'of employing conventional and non-conventional means simultaneously, adapting them to the characteristics of their targets'.

If a state, through its values, can exert very strong pressure on a society (or a former colony), inciting the memory of a strong country and a high international position now lost, this raises the security levels of a state, on what can be the various threats (hybrid) a state has to defend itself against.

When I started to take inspiration for my thesis subject, I tried to hypothesise various kinds of sociological methodologies from non-linear interactions in the social world. These insights helped me theorise what is currently my thesis subject, i.e. trying to create an AI-influencer capable of mitigating social threats, especially of the hybrid or human factor type.

Along the lines of this thesis, I found some interesting papers [14], including one in particular who began in 2020 to theorise the authoritarian potential of influencers in social networks, and to lay the groundwork for my idea of the creation of an AI-Influencer.

The Security Influencer was not my original idea, my thesis research will utilise and elevate this concept (as will be explained in later chapters), trying to increase a real impact on a social sample.

The paper based on the "Manchester Arena bombing" during 2017, (where an Islamist extremist suicide bomber detonated a shrapnel-laden homemade bomb as people were leaving the Manchester Arena following a concert by American pop singer Ariana Grande, Twenty-three people were killed, including the attacker, and 1,017 were injured, many of them children. Several hundred more suffered psychological trauma. the paper [14] resulted from this event contributes to understandings of constructivist security through analysing social media outputs to understand who is influential in the security debate and how. In particular, it discusses the creation of spontaneous Influencers that arise when a catastrophic event triggers discussions in social networks by radicalised people ready to accuse one part of society of being 'wrong', as opposed to another 'sane, just and civilised'. Influencers with the highest potential influence identified within in this article are not organised in creating an individual 'security brand' to cultivate influence, instead taking micro or nano 'security' influencer form.

The first type of security influencer emerging within the dataset are 'security broadcasters'. This type of security influencer simply broadcasts a message that becomes influential without further engagement from the influencer, like those paper evidence [23, 24, 128]. This is interesting because critical security studies conceptualise elite generated security discourse in this way. The Copenhagen school sees the elite as in a position of power projecting their security narratives to the audience. [129]

In particular, it is important to note here that none of the influencers have previous security field experience yet influence the debate. The influence is also ephemeral; security influencers and the unlikely individuals becoming important structuring security speak on social media rise from positions of security obscurity and in two cases from positions of having little existing influence on social media.

The contestation of the security speak offers important insights for how debates about terrorism, security and Muslims are structured. Importantly, becoming an influencer does not mean you are agreed with and terms of disagreement are enlightening. The micro-discourse mirrors ongoing social

Figure 1: Content shared by @SpookyAly

and political debate about terrorism, security and integration. Emergence of an anti-Liberalism discourse is interesting although only in one conversation, demonstrating that while illiberal, alt-right, discourse emerges, it is of fairly small magnitude. Far more common is discourse around Islam and Muslims being inherently violent and Islam and liberal society irreconcilability as common themes emerging through all contestation across all actors. An important area of agreement is couched in terms nuancing terrorism and identity with the example of the IRA bombings and how the city's ethnic and religious mix was not disrupted by previous sectarian violence, or were particular communities punished for being co-religionists with terrorists.

The influencers born during the discussion succeeded in changing the community's perception of safety, especially that inherent to the Muslim community as show in Figure 1.

This event shows how it is indeed possible to mitigate possible social actions against one group or pro another, thus also showing the possibility of using these techniques as a means of national defence against those threats from human or hybrid factors.

3.2 Propaganda as facet of national security

As explained in the previous chapter, the human factor plays a key role in both internal politics and geopolitics. The humane factor is what can bring down a government or give birth to one, finance an allied state or establish proximity with an antagonistic state for cultural/economic interests, improve

relations with foreign bodies and states or reduce vicinity to them. The human factor is thus not a simple 'popular will', but that cultural pressure exerted by society, which steers the state or other organisations in one direction or another.

The Ukrainian government would hardly have had all the support it received (economical, mediatic, cultural-social) if it had not established that good reputation with Western countries. The Russians, on the other hand, got a bad reputation by killing defenceless people, civilians, not to mention the fact that they invaded a country (without declaring war), in 2022, i.e. at a time in history when the European culture denies any kind of brutal and unnecessary event against another population/state/society.

It is clear from these words, that if a person has the ability to manipulate the human factor of a state, he has the ability to destabilise (at best) a public balance. In fact, as happened in the case of Crimea in 2014, where the population 'victimised' or succumbed to a cultural-authoritarian power of a neighbouring superpower, both agreed to be annexed by another state and also agreed to be militarily occupied by the same state. This was obviously successful because the social culture in Crimea has never shown any distance from Russian culture (in fact, in Crimea before its annexation, people spoke Russian, even though the state was Ukrainian).

However, it is possible to analyse the dynamics of the human factor? As explained in the various chapters, the balance between society, institutions and the state, is the result of certain social, economic, cultural etc. dynamics that allow a balance to be struck between social interests (living well, economic and social security, etc.) and the interests of the state (independence from other countries, sovereignty of its own resources, facilitating social and political dynamism). However, when one of these necessities fails (either for the state subjects or for society) there it starts to be a discontent, let us take an example, the social/state economic situation is bad.

This does not imply a revolt in and of itself, but it does imply a change of human factor (states are made by people) towards those who caused that negative situation, e.g. Greece towards Europe/Germany during the economic crisis of 2009. To provoke an uprising or riot, however, being the social-political context a complex dynamic (as it is inherent in so many variables and interaction) there needs to be another 'failure' in the system, let's take the cultural one for example, i.e. where civil and social rights have been reduced or cancelled. In this case, there is a much greater chance of an uprising being triggered. As described by Perrow with his 'Normal Accident s' theory [84] . Perrow argues that if a system has two or more failures/events afflicting its production/protection systems, and that these may follow one another temporally in an arbitrary order, then a specific sequence of events will exist such that all redundancy/security systems will fail, making a catastrophic failure inevitable.[85]

A case where the human factor has a key role right now, is what is happening in Iran, where people, due to the economic situation (first negative dynamic) and the repercussions on civil rights

(second failure), cause a social reaction to the adverse events, which also stems from the western passive-human cultural factor (thanks to the internet).

Another case of revolt, starting with the human factor as the key, always analysing the criticality of the system as explained by Perrow, is the case of the revolts in South America as of 2019. Popular uprisings against high prices in Latin America are back in 2019. Violent social outbursts, triggered by a severe economic crisis (first failure) and the increase of public transport fares (second failure), similar to what happened in the 1990s. The first, dramatic one was the Caracazo of 1989, the uprising in the poorest neighbourhoods of Caracas that inflamed Venezuela and spread like wildfire to the south of the continent stunned by the often crude and ruthless horse-riding treatments of the International Monetary Fund (IMF).

At this point, given that it is possible to change and manipulate people's perceptions, thus manipulating the human factor by sharing information in various communication channels (hence propaganda), along with social networks, we have to ask ourselves whether this is actually a risk or a threat.

There is precisely a strong difference between danger, risk and threat [130].

- **Danger** is defined as a radicalised group that until that moment does nothing and is not active, is there even if inactive, but can be a danger.
- **Risk**, on the other hand, is that group or individual that has the possibility of doing something dangerous and if it does, whether it does it involuntarily or not.
- **Threat** instead is where an individual or group, which has an important relevance similar to risk, does something harmful voluntarily.

Therefore, if the action is voluntary, it is a threat to individuals and public or private organisations. The case of Capitoll hill is at issue as evidence of Trump's voluntary action during the event is sought. Although groups such as the prod boys have been defined as a risk, but the capacity for voluntariness and dangerousness was (probably) driven by one person, who therefore voluntarily created a threat.

In fact, given the dangerous nature of the event, Twitter blocked many accounts, both bot and non-bot, that were in such a way linked to both the action of the events on Capitol Hill and the propaganda that triggered this event. [131] In fact, figure 2 show the amount of twitter account banned after the capitol hill riot . [132]

These accounts, sometimes real, sometimes fake, use their authority and authority for their own purposes, some legal, some not. Very interesting is the study and work of Transcrime, where through the FATA project they have tried to strengthen knowledge and public-private cooperation against

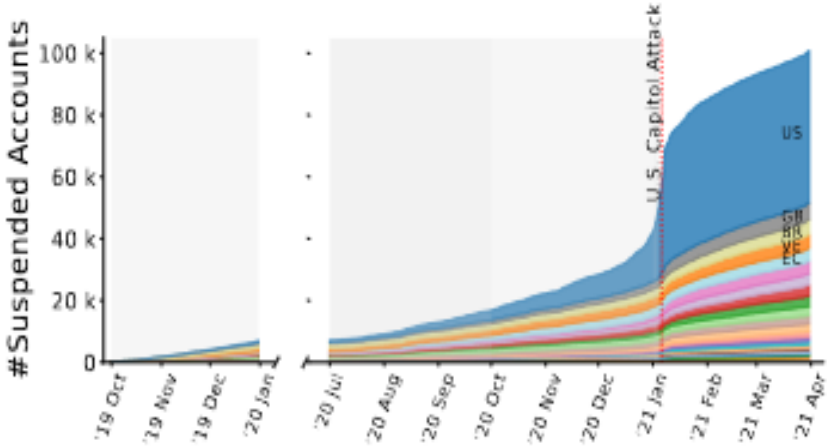

Figure 2: Number of suspended accounts as a function of the date they posted their last tweet, coloured by country

new forms of online counterfeiting. [42] Together with Amazon, they collected data, among those I want to mention [133] of most interest and for relevance to my thesis are two:

The first is in November 2020, where Amazon took legal action in the United States against two influencers who, via their social profiles (Facebook, Instagram and TikTok), sponsored counterfeit products for sale on for sale on various marketplaces including, besides Amazon, also Etsy and DHgate [134, 135].

The scheme was identified by Amazon thanks to intelligence work carried out by the Counterfeit Crimes Unit (CCU), which, in addition to its own marketplace, also proactively monitors other online channels (e.g. websites, social networks). The two influencers and eleven marketplace sellers involved had devised a complex fraudulent scheme fraudulent scheme that exploited social networks to circumvent marketplace controls:

1. the sellers posted advertisements for generic products on the marketplace, without including logos or other distinctive distinctive signs in images and descriptions (thus preventing the marketplace's automatic detection systems from marketplace's automatic detection systems to detect possible violations);
2. the two influencers advertised these advertisements on their social profiles, referring instead to products of well-known brand owners. Photos of the generic products were in fact placed side by side with photos of the equivalent counterfeit product with the text 'order this/get this';
3. the social followers of the two influencers were then redirected to the listing of the generic products on the Amazon marketplace via hidden links;
4. once the order was placed, instead of the generic products they had ordered (order this), the buyers received the corresponding counterfeit product (get this).

The two influencers recently settled the case and paid a penalty, which Amazon donated entirely to the development of brand protection activities. to the development of brand protection activities, including awareness-raising campaigns on the issue [136]. Amazon is now also preparing to take legal action against the sellers involved in the scheme who, although although based in China, had fraudulently indicated residence in the United States [137]. I do not know why the most effective countermeasures have come more quickly when clear commercial interests are at stake, as opposed to political or social ones. But I hypothesise that since trading companies are bound by fewer political, geopolitical, internal legal and bureaucratic constraints, there is a quicker response from economic actors than national ones, such as states.

The second operation, called 'Bologna Luxury' was a very successful experiment in a form of collaboration between intellectual property rights holders and law enforcement agencies. How did the collaboration between law enforcement and the private sector work? The starting point of the investigation was a take-down effectively implemented by a major fashion brand following 'internet brand protection' activities. The legal department of the brand then formally communicated to the "Nucleo Speciale Beni e Servizi of the Guardia di Finanza" (Financial Police) the data relating to the out-of-court enforcement conducted, indicating, in particular, the social accounts involved and the posts removed. Then investigative path tested allowed this information (considered qualified) to be used in subsequent offline investigations, in order to identify the perpetrators and reconstruct the illicit economic flows.

The investigation activities then ascertained how behind one of the reported accounts (called 'follie_of_luxury'), activated first on Facebook and then also on Instagram, was hiding a young influencer who acted as an intermediary between a supplier in China and end customers in Italy. The sales scheme, which generated over EUR 200,000 in a few weeks, was as follows:

1. the potential buyer would choose a certain garment or clothing accessory from the catalogue advertised catalogue;
2. the price negotiation took place directly in a private chat on WhatsApp;
3. the influencer, after receiving payment (by bank transfer or PostePay recharge), would send the money received (retaining a commission) to the suppliers' PayPal accounts in China;
4. the influencer, again via WhatsApp, communicated to the supplier in China the details of the order to be prepared and the addresses of the buyers to whom the products were to be sent;
5. the supplier in China shipped the counterfeit products directly to the customer, with no possibility of return and/or refund. refund.

As part of the investigation, the Financial Police also started an interaction with the social networks involved, through an international rogatory, aimed at acquiring the necessary elements to identify the person hiding behind the accounts such as log files, telephone numbers and payment instruments

payment instruments used. In relation to the last point, once the e-mail address of the person concerned was identified, the financial police also forwarded to Paypal, via the Safety Hub - PayPal Law Enforcement, the request for the sharing of transaction data associated with the account, which confirmed the economic movement economic movements already collected also thanks to the analysis of the transactions made with PostePay.

In both operations, the influencers promotes on a messaging app counterfeit bags made by manufacturers in China that are shipped to the homes of Italian customers.

Can we therefore define that influencers are bad in and of itself, because they wield too much power?
The sharing of media content cannot be defined as propaganda, but must be framed in a different perspective, it must be defined as inspiring values through shared content, capable of bringing out social behaviour and actions (as for Trump as for Zelensky). The purpose of these actions depends on the content of the message, but the medium (i.e. the ability to transmit content like influencers) like all technologies is defined by the use made of it. Thus, the reputation of people towards the individual who wields that power can be considered as a technology (i.e. a means leading to an extension and enhancement of human faculties by which a certain goal can be achieved).

Of course, there may be completely fortuitous cases or emergent behaviour resulting from the interaction of the various actors involved. Some cases concern bizarre events that have caused concern, such as the case of the strange message that appeared on 28 May 2021, on the Twitter profile of the Department that manages the US atomic arsenal [6]. [138]

The message: ' **;l;;gmlxzssaw** ' caused great alarm among the various security departments and many began to wonder about its meaning. Is it a nuclear code? Is it a hacker attack? No, the post was written by a child, the son of a remote working defence employee, who was playing with the keyboard, way similar to a Fat-finger error. [139]

This kind of "error" cannot be predicted, and of course, as described in the Complex system section, even if we know every part of the system, we cannot predict the future of the system itself due to the interaction of all parts present in the system. We can make probability prediction, but not predict it. This work, for physical system and for social system.
These types of threats, both predictable and unpredictable, are examined by the various national intelligence agencies that manage and collect information. Intelligence is therefore a set of people using processes and data elaboration, with the aim of improving the security of citizens, states and authorities.
So, the future cannot be predicted, but intelligence service must [7].

[6] @US_Stratcom

[7] Defined as the government agency responsible for the collection, analysis, and exploitation of information in support of law enforcement, national security, military, public safety, and foreign policy objectives. The most well-known are CIA, MI6 and Mossad.

3.2.1 Intelligence and Complex systems

Intelligence had to predict the future, even though no one can.
This sentence, I think, can describe almost everything about my thesis, there is a challenge, *the search for information about the future for the safety of people*; there are two research concepts, 1) *the methodologies that allow this to be done*, and also, 2) *an extremely new field that needs to be analysed and observed in order to understand the complex social dynamics that surround us*.

Intelligence has as its goal the prevention of threats to the republic and its citizens, so it must foresee the future for the prevention of these potential threats.
Even if we cannot predict the future, it is possible to hypothesise, in a probabilistic manner, how a system may go from what are the immutable variables of the actors in play.
As example of immutable variables, I can cite for example, the Chinese LM-5B rocket (will be mentioned in future chapters about my GSS paper) even though its orbital trajectory is unstable given its re-entry into the atmosphere, hence the unpredictable dynamics of the air currents (density, trajectory) and those of the solar wind (which affect the atmosphere), it is possible, however, to give probabilistic values of where it may land/crash because the rocket's immutable variables (mass, velocity, aerodynamic, centre of gravity) can be exhaustive values for predicting with probabilistic values where/when it will fall [140, 141].
The immutable variables, are of course immutable for a rocket (if we do not count possible re-entry or launch damage) but can be deduced and obtained in a social system, from the collection of initial conditions at a given time on the sample we wish to study.

Many intelligence actions go into preventing what can be various threats, starting with what is intelligence methodology, i.e. the collection of data (through agents, technologies and analysis), the validation of the same, and finally the processing and interpretation of the data then transformed into information, are the main activities of intelligence work for state and republic security; in this paragraph, however, I will focus on those that are the most recent and most in line with my doctoral thesis, i.e. cyber domain, technological domain and information/disinformation sphere.

For example, the Eu members concerning the information from Russia, during the Russia-Ukraine invasion, are facing an extremely sophisticated disinformation/information war that requires government, finance and business people as well as journalists, and others, to know how to read - always better - the information/disinformation warfare that is ultimately fed daily to what we call public opinion.
Country governments are increasingly tempted to influence political life, and the freedom to vote itself, using increasingly sophisticated digital tools of information manipulation. There are no borders to virtual reality, and the goals and tools of disinformation warfare are aimed at altering political, economic and social relations between states. [142]

About the Russian disinformation in Italy, Antonello Giacomelli from AGCOM, said: [143]

> It must always be kept in mind and reiterated that it is not a matter of pursuing the goal of alignment with the single thought, but of countering strategies planned by external autocracies, precisely with the aim of altering the free exchange of opinions that remains the reference value of a democracy. More distinction should be made between opinions and disinformation; Therefore, the rightful work of countering disinformation strategies strictly respects the boundary with the constitutionally guaranteed right to freedom of expression and freedom of the press.

In this sector for example, we can define as Immutable variables and initial condition of the studied system, the literacy level of the society.

An Italian (and also Spanish) problem that ties in with the subject of my thesis is that of literacy. In Italy [144] , only 20.1% of the population (aged 25-64) has a university degree compared to 32.8%in the EU. The shares of university graduates are higher in the North (21.3%) and in the Centre (24.2%) than in the South (16.2%) but still far from European values. There is also a wide distance from other European countries in the share of the population with at least a diploma (62.9% compared with 79.0% in the EU27). Adult participation in training is lower than the European average, with stronger differences for the unemployed population or those with low levels of education. As says Prof. Mario Caligiuri, president of the Italian Intelligence Society, about the de-information disease during COVID and the Ukrainian-Russia war [126].

> "the real Italian emergency, which is the educational one, since seventy-five per cent of voters do not understand a simple sentence in our language. It is no coincidence that the low level of education may increasingly represent a problem of national security and the reliability of the real nature of democracy in our country".

This case denotes how the immutable variables, at the level of national security to assess resilience and the level of the state, are closely related to the most important national assets that distinguish the state in the world. Starting from this personal assumption, it is possible to imagine the assets on which we are stronger than other more prosperous countries in the global sphere. For example, USA and China are fighting for the technological leadership, French and Germany for the economical resilience, Italian (and maybe Russia) work for the international recognition as a superpower state, wile Italy have already the leadership for Food, tourism (with French and Greece) and for the high technological micro-business-ecosystem as opposed to Germany, France, which have large companies, but a small business ecosystem.

This also denotes how intelligence, intense as groups working for the security of their own republic/state, also have, for obvious reasons, facets that are delimited by the culture of the society that

represents them. [145] This is extremely important in the part of negotiation and communication between states.

Variables that are immutable from a social point of view are of course mutable from a temporal point of view, in fact it is necessary (as described at the beginning of the section), that they can be deduced and obtained in a social system, from the collection of initial conditions at a given time on the sample we wish to study.
This methodological part is important, because it is of course possible to change and make these variables change, but this implies that even if these variables change, there may be non-linearities between the change made and the reputation (for better or worse) given to the change. For example, Italian government since some years, have improve a lot the cyber security defence becoming one of Europe's most successful cyber leaders, but the variable mutation wasn't achieved by society, arriving at the paradox where, people think that the State lies and cyber thieves tell the truth. In cybersecurity, criminals are considered nerds, while those who work to secure public administration data are considered incompetent. [146] This figure also gives an indication of the authority and reputation of the state. In fact, as in the case of COVID or the Ukrainian-Russian war, people think that the state lies to its sovereign people, to justify actions such as compulsory vaccination, while those who have become the real heroes are those who misinform people in the various social groups, hence the socio-technical part.
The freedom not to vaccinate was one of the main topics for which people invoked freedom of expression. However, since some social networks are made up of rules for the protection of minorities or the safety of groups of people (especially after the capitol hill cases and Donald Trump's subsequent ban), people have moved away from some classic social networks to other social networks with more freedom (at least promised) of expression. One of these social networks is among others that of former President Donald Trump, called *Truth*. After being kicked off Instagram and Twitter for anti-Semitic comments, Kanye West has decided he wants to become the owner of the American right-wing social. With the former president heading Truth and the Tesla genius buying 100% of Twitter's shares, absolute freedom of speech on social media (which is not necessarily synonymous with democracy) is no longer a mirage. [147]

> "In a world where conservative views are considered controversial, we need to make sure we have the right to express ourselves freely."

Releasing these words to Bloomberg was rapper Kanye West, [147] but if you didn't know better, you might imagine they came out of the mouth of Donald Trump or Elon Musk. Freedom of speech, in the US enshrined in the First Amendment that sanctions inviolable free speech, seems to have become an obsession for all three. So much so that, after being banned from Instagram and Twitter because of some comments of an anti-Semitic nature, Ye - as Kanye now legally calls himself - is moving to buy Parler, the ultra-right's social network. The fact is that in the name of free speech,

the Musk-Trump-West triad could represent an important front to watch, in the free speech, but also for the national security and the social equilibrium between different part of society.

In light of recent events in Ukraine, however, this could be an even bigger problem. A profile of an ordinary Vladimir Putin, for example, could be welcome because the messages he would launch would represent his view of reality, which, although heavily tampered with, would still remain an opinion that in free speech deserves to be considered and, therefore, disseminated.

The example of the Russian president is not accidental, since Musk ended up at the centre of an affair that drew the eyes of the world on him. He took off his businessman's robes and put on those of a geopolitical analyst, suggesting to Volodymyr Zelensky the road to peace. A road that the Ukrainian president refused to take, since he would basically have to please Russia on every point. Even, in an interview with Politico [148], Trump's former Russia advisor Fiona Hill said she was sure Musk was relaying a message from Putin, since he could not have known so many details of the proposed plan - as in the case of the water supply to certain areas that would have to be negotiated, such as Kherson and Zaporizhzhia. Shortly afterwards, he threatened to discontinue the Starlink satellite service that turned out to be a godsend for the Ukrainians, only to retrace his steps.

Musk's peace proposal was followed by another to normalise relations between China and Taiwan. This too, like the Ukrainian one, was rejected by the island of Taipei, which according to the South African entrepreneur's ideas should have returned under Beijing's protective wing. An assist to Xi Jinping, which most experts did not like, but which perhaps finds reason in the great economic interests (that does not make Washington sleep well) [149] that Musk and Tesla have in the land of the East. The result could be to have the owners of major social networks peddling democracy for a more likely free-for-all. Twitter in Musk's hands, Truth in Trump's and Parler in Ye's. A new phase could be opening for social networks and, perhaps 'in retrospect it was inevitable', for giving a voice to people who can't find it. But given the types of comments and content, organisations that have to defend the security of the state and the people, must investigate possible instances of incitement to general violence as Capitol Hill, and/or against races [8]. [148]

Regarding the technological field, the confrontation in the field of technology as AI (Artificial Intelligence), is diriment as far as the strategic competition between the West and China is concerned. Rather than on mere technological capability and the economic, political and security benefits derived from it, this rivalry is based on a value confrontation that finds its focal point in technological competition.
Underlying this competition is a key concept: whoever manages to dominate technologically over the other competitor will be able to set the standards for building and managing the technologies themselves.
Over time, Beijing has deliberately chosen to use technical standards to increase its capacity for

[8] Kanye West made comments on social media where he lashed out at Jews.

mass surveillance (restrictions on Internet use, manipulation of personal data). Through the China standards 2035 program and the Global initiative on data security, for instance, the People's Republic has given great impetus to the promulgation and dissemination of its own standards in the global framework. In addition, Beijing also uses its technological superiority to exert a degree of control over global technology standard-setting bodies, such as the International Telecommunication Union, using its leverage to elect favourable leadership and to push items on the agenda.
One thinks of the Pact on Security and Technology between Australia, Great Britain and the United States. This kind of initiative, however, requires certain political as well as economic sacrifices on both sides of the Atlantic. Significantly increasing transatlantic cooperation in technological terms implies, for example, placing the good of the alliance as a whole above that of each national industry. Other new formats, such as the Trade and Technology Council, seem to bode well for aligning different approaches and creating a more coherent technology policy strategy. [150]

Therefore, the human factor and tecnology (such the AI) has the potential to negatively change society and institutions, whether in the economic, security or obviously political sphere. But is it possible to use the human factor as a positive force, to deradicalise groups of people with a level risk or threat level against the state or society?

3.2.2 Crime and deradicalisation

Deradicalisation is a process in which people reject the ideology they once embraced, and using the authority of influencers within a radicalised group may have the possibility and purpose of deradicalising the group's participants.
Each state of course has its own processes and capabilities to deradicalise people, for example, citizens captured and taken hostage by terrorist groups (such Daesh or similar), forced to read and initialised to the law of the Koran to stay alive. Living in such an environment, for a long time, subject to perpetual indoctrination, either the citizens who have escaped hostage or those who have simply been indoctrinated, participate in deradicalisation or de-indoctrination courses by the authorities.

On 14 March 2022 [151], therefore very recently, the Italian state updated what were the 'Measures for the prevention of radicalisation and violent extremism of jihadist matrix A.C. T.U. 243, A.C. 3357-A', taking up part of the contents of the text approved during the 17th legislature by the Chamber of Deputies (S. 2883), which did not complete its passage through the Senate before the end of the legislature.
The part of the text I would like to focus on concerns:

- favouring deradicalisation, without prejudice to the fundamental guarantees of freedom of religious freedom and in compliance with the principles and values of the Italian constitutional order;

- favouring the recovery in terms of integration (social, cultural, work) of the subjects involved (whether Italian or foreigners resident in Italy);
- creation of the National Centre on Radicalisation (CRAD).

The norm notes that the term radicalisation 'is now used to describe a phenomenon in which people embrace intolerant opinions, views and ideas that are likely to lead to violent extremism violent extremism' and calls on the European Commission to establish an action plan to implement and evaluate the EU strategy to combat radicalisation and recruitment to the ranks of terrorism; The norm provides for two different types of radicalisation, one more focused on the Jihadist matrix, the other on the violent type. But unfortunately, the 'National Plan for the Re-education and Deradicalisation of Prisoners and Inmates' has not yet been formalised and presented (November 2022). [152]

However, there have been observations made about the impact on deradicalisation, especially in Europe. [153] The prison governance styles and the hegemonic forms found in the prisoner community thus define a structural framework with variable geometry, which makes impracticable the hypothesis of prison as a neutral container for radicalisation. In other words, sociological studies on prison give great importance to the variables that affect the processes of internal socialisation. It is only in the comprehension of this variability that serious analytical perspectives and, possibly, effective strategies of prevention can be developed.
Telegram groups, therefore not easy to enter (closed groups), and with a community that knows each other almost on a daily basis, can particularly resemble prisons (echos chambers), where a type of socialisation can help the deradicalisation of people in the group. Of course, a minimal level of generalisation can be identified. If the specific key of prison radicalisation is identified in the violence of the environment and in the crushing of the demands of inmates, implementing strategies to reduce the harms of detention seems to be a useful operation. From this point of view, deradicalisation projects and radicalisation prevention projects do not deviate from the traditional perspectives of rehabilitation treatment [154].
A less oppressive prison, which offers opportunities for schooling, training, recreation (sport, cultural activities), work, contacts with the outside, is destined to be less explicit on the values of containment and stabilisation of the criminalised portion of society, (which is unfortunately, at least on a practical level, difficult to implement in a Telegram group [9]) even from the point of view of those who are imprisoned.
These traditional strategies could therefore reduce the pressures of conflict and the radical instances of social redemption. In an even more specific sense, it is evident that a full (substantial) recognition of religious rights and an organisational effort aimed at guaranteeing access (spaces and times dedicated to the exercise of prayer, respect for food prescriptions), would probably have the effect of

[9] In a Telegram group, people usually enter there precisely because they are looking for the information they want to read.

reducing the perception of hostility and discrimination. It is a matter of obvious, and mere exercises in applying common sense. So, it's all very simple? Not at all, because the afflictive dimension of detention is equally obvious and unavoidable, the objectives of deterrence and neutralisation that the prison systems pursue are equally crucial. But in general terms, however, a prudent and pacified religious practice may be consistent with the above mentioned institutional objectives of the containment of hardship and internal conflict.

There can also be micro, meso and macro models to study the phenomena of deradicalisation.[155] One of the possible factors of deradicalisation, is the micro-level factor, may be the loss of ideological appeal from the individual, but as well as the "Do I want to live my life like this forever? logic". Although the latter option, if initiated by other group members, is reinforced by exposure to alternative viewpoints, e.g. through relevant books and media, it can be a good way to deradicalise group participants, it must also be supported by other group members.
While it may be much more effective, the meso-level element is detachment from the group and its activities, sometimes caused by intra-group conflict and disappointment with the (leaders of the) group, Not only can individuals leave radical groups, such groups may themselves disintegrate and cease to exist.
Wile at the macro level, as explained before, prison can sometimes create a context in which people want to make a new start and be deradicalize.
It is crucial to take into account group membership and the inter-group context that forms the basis of radicalization (thus also creating a strategy to deradicalise). Every radical group is characterised by a strong sense of a (superior) in-group identity, as well as an (inferior and evil) out-group, which is held as responsible for the grievance of the in-group and as such is perceived as a legitimate target of violent attacks in order to achieve societal and political changes. [155] Importantly, people are able to resist the temptations of a radical ideology to the extent that they have a strong shield of resilience. Interestingly, a shield of resilience may also make people — once they belong to a radical group — less susceptible to attempts at deradicalisation.

The **International Centre for the Study of Radicalisation and Political Violence** (ICSR) during his *Prisons and Terrorism Radicalisation and De-radicalisation in 15 Countries* dossier [154], in 2010, evidence how the sample show that authoritative leadership and a conducive environment are essential to making collective processes of disengagement and deradicalisation work. Equally important, however, is governments' ability to recognise these conditions and help facilitate the implementation of programmes.
Governments [154] rarely provide the initial impetus for processes of collective deradicalisation and disengagement, but they can be critical in assisting the process and leading it to a successful conclusion. In doing so, governments need to:

1. Recognise changes in attitudes and beliefs among armed groups' leaderships;

2. Open appropriate channels of communication through which collective deradicalisation and disengagement efforts can be supported;
3. Provide the right mix of sanctions and inducements (the proverbial 'carrots and sticks') that will keep the process on track and move it towards a successful conclusion;
4. Enable the release and re-integration of terrorists and/or insurgents, ensuring that any return to violence remains unlikely.

The dossier describes the dynamics of some countries, I will put a small description for each of them to show how cases of deradicalisation can be initiated through similar strategies. These cases, follow an initial analysis of*Leadership and environment dynamics*, and then the*Repression* and *Inducements* part.

In **Egypt** the Egyptian Islamic Group (IG) is the best known example of prison-based collective disengagement and deradicalisation. Faced with defeat, the loss of popular support and the prospect of an 'Algerian-type' civil war, the IG's leadership decided to embark on a comprehensive re-interpretation of the group's doctrine on the legitimacy of using armed force, which resulted in the publication of 25 volumes. In 2002, the leaders – who had themselves been imprisoned – toured prison facilities across the country, holding discussions (first) with middle-ranking commanders and (then) the group's foot soldiers. By the end, nearly 15,000 militants had agreed to follow the new course. The process led to no significant splits, and no terrorist attacks or armed operations have been attributed to the IG since 1999. The Egyptian government's initial response to the process, however, was marked by a lack of interest, if not suspicion. For four years after the group first declared a ceasefire in 1997, there was little support. Only after the September 11 attacks in 2001 did the government – now under intense pressure from the United States – begin to help the IG's leadership in promoting the new course, mostly by organising the leadership's ten month tour of the country's security prisons. Even so, the government's initial lack of response and even tacit support had not only slowed down the IG's collective process, it also (unnecessarily) prolonged the group's campaign of violence.
The government repression in the 1990s – which involved the mass arrests of Islamist sympathisers and their systematic mistreatment in prison – further radicalised sections of the Egyptian population as well as the prisoners themselves. At the same time, it seems clear that – over time – the government's excesses contributed to the IG leadership's decision to re-think and, eventually, change their mind on the use of violence. Violent jihad, the group's leadership argued, could be justified only if it furthered the movement's aims, yet the group's violent campaign had achieved the opposite: mosques had been closed, preachers banned, the families of activists were suffering, and most of the movement had ended up behind bars. According to one of the group's leaders, if God was on their side, none of this would have happened, which meant that something had to be

'theologically wrong' with the decision to confront the regime.
Once the Egyptian government started embracing the IG's disengagement initiative, it quickly moved from suspicion towards active facilitation. Not only did the repression stop and prisoners were offered relaxed prison regimes (including, for example, better meals, more visits, better opportunities to mix with other inmates, etc.), the government organised the leadership's tour of the country's security prisons and even helped with the publication of its books. Also, for the first time, the group's efforts were covered, and praised, on state television. Critical was the promise of early release for those who had joined the programme. Of the 15,000 IG prisoners in 2002, 13,000 had been released by 2006. In 2007, only a few hundred were left. Furthermore, those who have been released now receive 'pensions', which – though administered by the IG – are widely thought to be funded with the assistance or, indeed, direct support from the government.

The disengagement of the Islamic Salvation Army (AIS) in **Algeria** took place in the second half of the 1990s and concluded with the standing down of the group in 2000. In contrast to Egypt, it was the government which first approached the group in 1993. Three attempts to negotiate a settlement with the imprisoned leadership of the Islamic Salvation Front (FIS; the AIS' political parent organisation) led to nothing, partly because the FIS' leadership still felt confident about its chances of toppling the government. By 1997, the situation had changed dramatically, with Algeria having descended into outright civil war and the various insurgent groups rapidly losing support among the population (see below). This time, the decision to embark on a process of collective disengagement was taken by the AIS' military leaders, who had hidden in the group's mountain strongholds. In the end, the political leadership, though reluctant, had little choice but to accept the outcome of a process which most of their imprisoned comrades (and their families) supported.
The country had slid into civil war, with massacres of civilians being committed by all sides, principally the Armed Islamic Group (GIA) and the government. The massacres often took place in AIS strongholds while most of the group's active members had been arrested and deported to detention centres in the Algerian desert, leaving their families defenceless. The public, on the other hand, often failed to distinguish between different Islamist groups, with the (paradoxical) result that the AIS, whose supporters had been among the principal victims of the GIA and government repression, felt it needed to disengage from violence in order to stop the massacres and prevent the loss of public support. In the leadership's view, the situation had become untenable – the 'jihad', in the words of an AIS leader, 'was about to be buried by its own sons'.
In Algeria, the government went even further. Not only were the AIS' prisoners released back into society, they received promises of employment (which, however, were not always kept). Families of FIS activists who had been victimised by the government were paid compensation, and some former members were given permission to wear guns in order to protect themselves against revenge attacks. Political concessions formed an essential part of the 'package' that had been negotiated

with the Algerian government. It committed to reducing the military's influence over public life, and gave assurances that former members of the AIS would be allowed to establish political parties and participate in political life. Again, not all these promises have been kept, but they illustrate the scale of the government's commitment and the very wide-ranging nature of the process, which – in many respects – resembled a more traditional peace process.

In **Israel**, neither the political nor the military leaderships of the most important Palestinian groups – Fatah and Hamas – have showed any interest in pursuing collective deradicalisation or disengagement outside the context of a negotiated political settlement. Even when they were under military and political pressure, the groups always insisted on being treated as equals, negotiating on behalf of 'their people' and achieving the settlement of the conflict as a whole. As a consequence, there have been no known attempts by the Israeli government to facilitate prison-based collective disengagement or deradicalisation efforts.
Despite the ongoing conflict, the situation never reached a point at which repression alone would have 'broken' any of the Palestinian groups. Palestinian security prisoners in Israeli prisons have not been exposed to the kind of systematic mistreatment that used to be common in Egypt. Unlike Algeria, the situation in the Palestinian territories has never quite reached the point of sliding into civil war, even if the conflict has imposed significant strains on the civilian population for several decades. Moreover, the Israeli government's practice of releasing prisoners in return for kidnapped Israeli soldiers or as 'goodwill gestures' in the context of peace negotiations has not encouraged armed groups to change their behaviour but – rather – allowed them to 'sit tight' and wait until it is their turn to be set free.
Once more, the Israeli case demonstrates why it is impossible to separate disengagement and deradicalisation efforts from the wider political environment. Even if inducements were to be provided to prisoners belonging to certain armed groups, or factions within those groups it is far from clear whether they would convince them to engage in collective disengagement or deradicalisation. From my personal point of view, it is not enough to apply a process of deradicalisation or disengagement on guerrilla actions given the indoctrination and radicalisation that is too great (given the deaths, the still growing tension, and the complex historical dynamics since WWII) between Israel and Palestine. As described, the Israel situation is too complex to apply a PsyOp deradicalisation process on social networks as described in my thesis. Short of a comprehensive settlement, participating prisoners would be released back into hostile communities, with participants likely to be shunned by their families and neighbours. For such programmes to work, all participants would have to receive resettlement aid and move to entirely different locations, which neither the prisoners nor their families would find acceptable.

In the three cases [154] that form part of the sample, government repression has triggered conflicting responses, which – depending on context and environment – have led to further radicalisation or intensified efforts at promoting disengagement and deradicalisation.
Inducements and aftercare are important elements of most disengagement and deradicalisation programmes, because they provide incentives for engaging with the process and then sustain the transition from prison back into mainstream society. In certain cases, collective programmes may go as far as offering political concessions as part of a negotiated 'package' that, more typically, includes promises of early release and cash payments.

Processes of collective disengagement and deradicalisation have proved powerful and effective in facilitating insurgents' and terrorists' transition from violence to non-violent activism. However, it should be remembered that the circumstances in which they are likely to succeed are quite limited:

- They seem to work best when the group's military leadership and most of their followers are behind bars.
- They also presuppose that groups have strong and authoritative leaderships, which are able to exercise a degree of hierarchical control.
- Most importantly perhaps, collective disengagement and deradicalisation tends to happen when the feasibility and/or viability of the armed campaign has come into question.

The process, therefore, requires a conducive environment, that is, a situation in which the leadership perceives that the armed campaign has faltered or is at risk of inflicting unacceptable losses to the group or its wider constituency. Even when those conditions are met, the local and historical context may dictate different kinds of approaches. For instance, the Israeli case shows that collective disengagement or deradicalisation are unlikely unless they are accompanied by a full settlement of the conflict. In fact, it may at times be difficult to distinguish the two, especially when – as in the Algerian case – political concessions form part of the 'package' that is offered to the terrorists and/or insurgents. When success is judged to be likely, governments should help facilitate the process:

- They can support the process of reform and dialogue, for example by allowing reforming leaderships to meet with their followers; relax prison regimes for those who have signed up to the programme; and disseminate their writings and publications.
- Repression is a double-edged sword. While it may be useful in 'breaking' an armed group's will to carry on and shift their perception from 'possible victory' to 'certain defeat' and 'unacceptable losses', it may also have radicalised many prisoners in the first place.
- Whatever role repression plays prior to the beginning of a process of collective disengagement and deradicalisation, the cases that have been looked at seem to show that – once a

process has begun – the easing of repressive measures provides an immediate reward for those who have signed up to the programme.

- Inducements and aftercare can make an important contribution to sustaining the success of collective disengagement and deradicalisation. Early release is by far the most important inducement for those still in prison, while financial support and/or employment appear to be critical in ensuring that armed groups and their members remain peaceful.

The case of Algeria showed that political reform and reconciliation may form part of a 'deal' that brings about collective deradicalisation and disengagement. This raises important questions about the nature and identity of the process: if a full 'package' is negotiated that includes the standing down of the armed groups in return for political concessions, collective disengagement and deradicalisation may turn into a fully-fledged peace process and benefit from drawing on the full set of instruments provided by conflict resolution and peace-making.

Obviously, the cases mentioned above are part of a real social and political ecosystem from a physical point of view, thus with completely different dynamics from online social groups. So, how to approach a deradicalisation process in a Telegram group? A more detailed description will be given in the next chapter, but it is possible to identify, starting from the complex social dynamics, and those of dynamics in networks, that:

1. By exploiting what is the very structure of the network, i.e. an echos-chamber, it is possible to send messages with information about articles, where both the standard of the newspaper and any other reference to classic newspapers is removed so as not to stigmatise the information shared;
2. The socialization process (like the many cases exposed) activated by the information shared in the network, can start a social dynamics that can:
    a) Reduce and mitigate the authority and the power of the leader;
    b) It makes the participants lose interest in the common vision and goal of the group.

But as explained in the various chapters on complexity, and as also described by the title of this thesis, social dynamics are difficult to manage, as there is a huge (not infinite) set of possible variables and interactions that can make the direction of the dynamics change from one side to the other constantly and perpetually. To overcome this constant change, one brain alone has an enormous difficulty, which is why my goal is to create an AI that is able to manage and understand, at least in part, all those possible dynamics and variables.
Obviously, an AI is not a holy grail, it obviously has both cognitive and programming limitations. AI has shown great and important features and improvements for society and research, but there are still important limitations. For instance, on the level of information management as for organisations, AI

has made a significant contribution in terms of efficiency and computing power even for sensitive organisations such as the CIA (Central Intelligence Agency) [156], thus giving it a greater capacity to process both sensitive and non-sensitive data, making the various analysis and security operations more efficient. [156] The use of AI systems in social media presents simultaneous limitations and advantages from both social and technical dimensions. Opinion dynamics and information propagation are crucial factors that influence how information spreads on social media platforms. AI systems have the potential to improve the efficiency and accuracy of information propagation by identifying and targeting specific user groups. However, the influence of these systems on opinion dynamics is still unclear and requires further research. Additionally, the interaction between the technical limitations of AI systems and the social dynamics of social media can create unintended consequences, such as algorithmic bias or the spread of misinformation. Thus, a comprehensive understanding of the interplay between social and technical dimensions, opinion dynamics, and information propagation is essential for developing effective and responsible AI systems in social media. On the other hand, however, AI still shows various imperfections both from the point of view of human-machine interaction, as well as on the level of compatibility between human needs and its management (the danger that certain technologies may end up in the wrong hands) [157], and finally on the ethical level. There are not a few cases where AI takes on radical or racist behaviour towards communities or organisations, but this also depends very much on the quality and diversity of the data fed to AI [158].

So I do not intend to say or show that an AI can manage and do it all; AI can make mistakes, just like humans, but certainly the support of an AI (capable of analysing so much information, and processing those small dynamics that escape human cognition) is a fundamental aid to the goals of my thesis.

4 Social & Complex systems literature review

4.1 Sociology of complex dynamics

The general science development trend in the 20th century, which is also passed in the new century, is the gradual penetration of ideas and methods of physics in natural as well as traditional humanities. Since the 1970s, the methods of mathematical and then physical modeling have been increasingly used in such sciences as demography, sociology, economics, history, and political science. In all these sciences, the desire for an objective and, preferably, a quantitative description of various social and economic phenomena is increasing. The development of quantitative models in sociology, political science, theory of transport, and other areas of society investigations is gradually moving relevant tasks from the humanities and engineering sciences to interdisciplinary applications of mathematics and physics. [159]

Ettore Majorana, famous Italian physicist, in one of his last publications, he explained how sociology, and other social sciences and humanities, could begin to be studied as a hard science, given the collection of social information and analysis with statistical laws. [160]

> "But the introduction into physics of a new type of statistical, or rather simply probabilistic law, which hides, in place of supposed determinism, beneath ordinary statistical laws, forces us to revise the basis of the analogy we established above with social statistical laws."

One must first convince oneself that the formal analogy could not be stricter. When stating, e.g, the statistical law: 'In a modern society of the European type, the annual coefficient of nuptiality is close to 8 per 1,000 inhabitants', it is quite clear that the system on which we must carry out our observations is defined only on the basis of certain global characteristics, deliberately renouncing the investigation of all those additional data (such as, for example, the biography of all the individuals that make up the society under examination) whose knowledge would undoubtedly be useful in predicting the phenomenon with greater precision and certainty than the generic statistical law allows; not otherwise when one defines the state of a gas simply by its pressure and volume, one deliberately renounces investigating the initial conditions of all the individual molecules. A substantial difference can be seen in the mathematically defined character of the statistical laws of physics, which is matched by the clearly empirical character of social statistical laws; but it is plausible to attribute the empiricism of social statistics (we mean precisely the inconstancy of their results beyond the part pertaining to chance) to the complexity of the phenomena they consider, for which it is not possible to define the conditions or the content of the law exactly. [160]

The significance of statistical laws according to classical physics can be summarised as follows:

1. natural phenomena obey an absolute determinism;
2. ordinary observation does not allow us to recognise exactly the internal state of a body, but only to establish an innumerable complex of indistinguishable possibilities;
3. having established plausible hypotheses on the probability of the various possibilities and assumed the laws of mechanics to be valid, the calculation of probabilities allows the more or less certain prediction of future phenomena.

It is possible now to examine the relationship that passes between the laws established by classical mechanics and those frankly empirical regularities that are known by the same name particularly in the social sciences. [160]

But Social systems are exceptionally complex, non-linear and highly sensitive to the initial conditions, thereby rendering the **accurate prediction** of the outcome of a social process an elusive goal [161]. They are also non-autonomous systems driven by time-dependent endogenous and exogenous stimuli giving rise to non-stationary activity patterns. [128]

As described above, data collection allows us to obtain information after analysing, even if we obtain a large amount of data, it is not certain that we can obtain 'acceptable' predictions, this difficulty stems both from the fact that a social system, as described, is itself complex (we shall see in the next chapter), and thus easily influenced by the various interactions of individual/agents/members, but also from the fact that the data. In wartime, we have to work with limited data, a constantly changing landscape and many assumptions. The work often has to be tactical, and what was produced the day before often has to be completely revised the next day because new information has arrived. [162] And then when more data arrives: At the same time, the challenges faced during infectious disease threats pose the questions and problems for the rigorous and fundamental research that allows the field to progress after the emergency has passed. [162] In order to be considered optimal, must comply with various characteristics, these characteristics are called the 5 Vs of big-data.
With the passage of time, these 5 Vs increased to 6, then 7 etc.. At present, here I describe those characteristics necessary for data to be considered reliable for social research in complex systems.

1. Volume: Volume therefore refers to this huge mass of information, which cannot be collected with traditional technologies.
2. Velocity: The challenge facing organisations is the need not only to collect this data but also to analyse it in real time, in order to be able to make important decisions as quickly as possible.
3. Variety: *"More isn't just more. More is different." - Chris Anderson*. Variety refers precisely to the different types of data available today from an increasing number of heterogeneous sources.
4. Veracity: *"Bad data is worse than no data"*. The data must be reliable, tell the truth.

5. Variability: The changeability of their meaning is something to be taken into account when interpreting data.

The combination of these characteristics in a data set (a set of related information that is composed of separate elements but can be manipulated as a unit by a computer), gives the possibility of being able to work and obtain/extract meaningful information from the event or set of events we want to study.

But information alone, in a social system, does not do much. We must not only know the initial conditions of the system, but also we need to *understand* it, in order to hypothesise the evolution of the system. To understand society, we must also understand the mechanisms that make it up and identify it from others. Furthermore, knowing how these mechanisms work (hence the dynamics) is essential to understanding the present, and possibly future, activity of the system.

Various sociologists such as Ulrich Beck, Vilfredo Pareto, Bruno Latour, Edgar Morin and many others have studied complexity and emerging phenomena related to society.

Bruno Latour with Michel Callon, for example, based on John Law' works, have present the actor-network-theory [163], which bases its development on a theory of the 'fabrication' of scientific fact. For these theorists, the world should not be thought of in terms of social groups, but in terms of networks (in a precise sense that in no way corresponds to that of the Internet). What makes the social is the 'association', the formation of 'collectives' and the set of relations and mediations that hold them together. In this way, the social is understood as an effect caused by the successive interactions of heterogeneous actors, i.e. the actor-network. Every actor is a network and vice versa. The action of an entity in the network leads to the modification of the latter; any action involving the whole network has an impact on the components of the network. Therefore, action does not have a precise source, it always involves a series of entities and mobilises the collective force that they represent. The relative stability of an actor-network can collapse at any time if certain actors are removed from the network. For example, if the telephone, the banks or the president are removed from the network in question, the network will become destabilised. Similarly, the actor-network is both local and global, depending on its size or length; it can be both a micro and a macro-actor since its size varies over time and depends on the translation operations of its spokespersons.

Vilfredo Pareto was an economist, sociologist and engineer. He started out as an economist, but I would like to put him in my thesis in complex social systems, not only because of the famous pareto curve, which often indicates the presence of complexity (hence the complexity signature [164]), but also because of his interdisciplinary studies from economics to sociology and political sociology. Interdisciplinarity is not only a scientific model with which complex systems are studied, but also a comprehensive approach to understanding those complex dynamics, which if individual disciplines alone have difficulty with.

Pareto's move from economics to sociology was based on economic studies, but starting from the constants of human nature and the rationality of the agent. The statistical study of the distribution of incomes had provided him with initial evidence of the stability of human nature despite varying historical-geographical situations. On the other hand, the observation of not only economic but more generally social behaviour led him to observe how the social individual only rarely acts according to an instrumental rationality of means appropriate to the end. The starting point of this new sociology, which, according to him, neither Comte nor Spencer had been able to conceive, is that in most cases, the social individual behaves in a non-logical manner, i.e. without an apparent purpose and, in any case, without a clear consciousness of the purpose pursued. [165]

The study of political dynamics opens up another important area of study inherent to the social and complexity sciences, is the geopolitical dynamics aspect, of one or more countries. Geopolitics, defined as the relating to politics, especially international relations (as influenced by geographical factors), is itself complex because of the difference of the actors in the field, but especially because of the international relations (and thus the dynamics) between the various states.
An emerging behaviour of the contemporary geopolitical situation (2022/10/1), is definitely the BRICS. The BRICS is a non-official collection of international states [166], which due to the Russian invasion of Ukraine, and defensive takeover by NATO states and passive affiliates, has resulted in a non-linear polarisation of the global geopolitical network, of which the BRICS is at the opposite end of the spectrum from NATO. This has arisen, not only for economic, military or strategic reasons, but for international reasons. Inasmuch as the BRICS nations (Russia, China, India, Brazil and South Africa) culturally distinguish themselves from the Western 'white-European' cultural hegemony (present in NATO countries). The BRICS, in the coming years, could also expand to other international states such as Argentina and Iran, which are also in antithesis against Western (especially US) culture.

This obviously leads to a greater focus on the possible dynamics of risk and danger between states. But also on this aspect, another important sociologist has gone on to study the sociology of complex dynamics, Ulrich Beck.

A very interesting concept from Ulrich Beck, in sociology related to complex dynamics, is the "reflexive modernisation". Beck identifies with a transition from a first to a second modernity and considers the role of sociology within this transition. Postmodern sociological theories and reflexive modernisation, in fact, differ from those of simple modernity (such as functionalism and Marxism). [167] "Reflexive Modernisation" is a text written by Giddens, Beck and Lash, The concept is treated differently by the two scholars in that Beck assumes the concept of reflexive modernity for what concerns economic and political processes, while Giddens focuses more on the sphere of traditions. Reflexive modernity is a kind of reflection on modernity, it is a process through which all the institutions we have built (the economy, science, family, state, welfare), and which we

thought produced order within society, begin to lose their functionality. In this process, the more modernisation takes root, the more its (often unforeseen) consequences threaten the fundamental institutions of modernity itself. An example often cited by Beck, a great Polish sociologist and colleague/collaborator of Giddens, is that of climate change: if we globalise the ways and principles of industrial state capitalism, we produce unforeseen effects that threaten not only the institutions of modernity but, indeed, all of humanity and nature (the climate change).

According to Beck, the risk in old societies is that of the distribution of wealth (from which wealth and security are derived). Currently, however, risk is understood as a systematic way of dealing with the insecurities and randomness induced and introduced by modernity itself, such as the global worming. A difference between Morin and Beck, is the availability of information. For Beck the advantage of this transformation lies in the greater 'democratic' nature of science and technology: in the class society, science and technology are isolated 'on an ivory tower', and therefore unreachable to the layman, whereas in the current transformed society they have become more 'open' and self-critical. [168]

But as mentioned earlier, beck defines various levels related to the criticality of threats and risk. Beck define, at least, three axes of conflict in world risk society that must be distinguished in the communicative logic of global risks:

1. Environmental risk conflicts, which spontaneously generate a global dynamic;
2. Global financial risks, which are at first individualised and nationalised;
3. Threat posed by terrorist networks, which are both empowered and disempowered by the states.

In the third case, of environmental risks that pose physical threats, there is on the one side affluence-induced environmental destruction, as in the case of the hole in the ozone layer and the greenhouse effect, which may justifiably be laid primarily at the door of the Western industrial world, though their impact is, of course, global. [169]

Building on the most widely accepted definitions and distinctions in the sociological literature, Battistelli & Galatino [130], following Beck's work, propose integrating into the conceptualization an aspect which has been only partially explored in theories of risk and uncertainty: the intentionality of social actors in the production of risks. This integration, while retaining the distinction proposed by Luhmann (1993) between danger and risk, also allows to distinguish between risk and threat. The difference being that risk is attributable to positive human intentionality, so that harms constitute ' "bads" that are tied up with the production of goods' (Beck, 2002: 44), while threat is attributable to negative intentionality, the harm being deliberately produced by ill-intentioned actors. Their proposal moves in this direction by introducing the notion of intentionality (positively/ negatively

oriented) and inserting along with danger (non-intentional) and risk (intentional and positively oriented), the concept of threat (intentional and negatively oriented). [130]

While particularly useful for the purpose of defining what risk is and what it is not, in their opinion these three specifications of the concept do not address the tangled question concerning the intentionality of the agent. Faced with the difficulty of capturing a phenomenon like terrorism within his conceptualization of risk, Beck (2002, 2009a) [170] takes some tentative steps in this direction acknowledging that :

> ... an essential difference between environmental and economic dangers, on the one hand, and the terrorist threat, on the other, resides in the fact that in the latter case purpose takes the place of chance. Environmental crisis and economic threats due to global financial flows, in spite of all their differences, exhibit a commonality: they must be understood in terms of the dialectic of goods and bads, and hence as contingent side-effects of decisions in the process of modernization. This is not true of the new terrorism, which annuls the rational principles of former risk calculation because purpose replaces chance and maliciousness replaces good will.

First, risk has to do with uncertainty as it is related to unknown future outcomes. Second, those outcomes are not necessarily negative as they may range from benefits to harm and be different for different actors. Third, risk has an agency element as it is associated with human decision.

Battistelli & Galatino [130], continuing their work with a sophisticated way of thinking about the dynamics of risk and threat. The typology their are suggesting here is inductively constructed on the basis of field research that the authors have carried out over fifteen years in different areas, such as international crises, crime, urban conflicts, health risks and migration. [130]

Regardless of whether we were investigating institutional actors, the media or lay people, their results confirmed that potentially or actually risky events are differently perceived by people and institutions and managed accordingly. At the same time, the technical notions of probability of occurrence and severity of outcomes revealed their limitations in explaining such differences in subjective and societal reactions, leaving many questions open.
Why are people more worried by international terrorism than by earthquakes? Why do they fear (relatively rare) small crimes more than (more frequent) car accidents? Why are institutions and society as a whole more prone to mobilizing material and symbolic resources to address urban insecurity than pollution? The fact that in the sociological debate events and problems that differ in nature are merged together, ignoring the distinctions between them, does not help in the search for an answer to such questions and leads to underestimate the psychological, social and political impact of harms whose author is not nature but human beings.
In order to address these issues, Battistelli & Galatino [130] typology is constructed on the basis

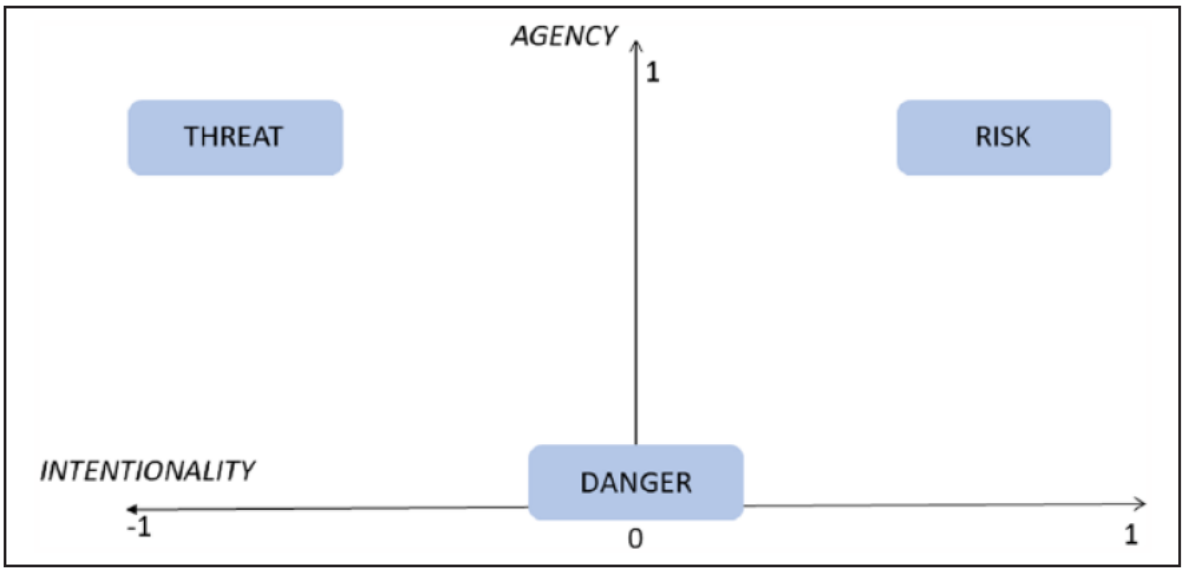

Figure 3: A typology of danger, risk and threat

of two main variables, agency and intentionality, which in the previous section we identified as essential properties for delimiting the conceptual borders of the notions of danger, risk and threat and for specifying their distinctive connotation.

Their model, as shown in the third figure of my thesis, demonstrates a very interesting concept inherent to threat dynamics.

In their model, agency and intentionality are not considered as discrete variables (presence/absence; positive/negative) but as continuous ones, since they may assume an infinite number of states within a given interval. Accordingly, agency, represented on the y-axis, can vary from absence of agency (conventionally 0) to presence of agency (1); intentionality, represented on the x-axis, can vary from negative intentionality (–1) to positive intentionality (1), passing through the intersection of the axes, where intentionality is absent (0). The axes' origin is the location of danger, a concept characterised by the absence of both agency and intentionality. Typical examples of danger may be infectious diseases, earthquakes, weather-related phenomena, and so on. Naturally, the absence of intention in the genesis of danger does not in itself mean the absence of human responsibility in the prevention/handling of its effects, as shown by the examples of the tsunami that devastated the Indian Ocean in 2004 or Hurricane Katrina in 2005. Both these cataclysms brought to light varying degrees of inefficiency on the part of the authorities in the face of the disaster, both ex ante (prevention) and ex post (management of the emergency), but it would not be possible to attribute their genesis directly to human agency and intention.

On this basis they suggested a tripartite typology which does not only distinguish between danger and risk in terms of agency but also between risks and threats in terms of positive/negative intentionality of the agents. At analytical level, a typology of danger/risk/threat may be helpful as a means of

addressing a persistent theoretical gap and limiting the recurrent conceptual stretching of risk. It can contribute to advance the understanding of differences in subjective and societal perceptions and reactions to potentially harmful events.

Although the concept may seem tenuous, knowing whether a negative event stems from the intentionality or not of the person who caused it, at the level of complex dynamics, is very important to understand whether it was the cause of an emergent behaviour (polarisation, homophily) or a simple rational event not directly influenced by external factors. For example, the key dynamic is to understand whether that individual was prompted by others to do that action (hence he got drag), or whether it was a conscious personal initiative.

Thus, depending on the type of voluntariness of the action, one distinguishes risk from threat. This transition is very feeble and labile, as a person with authority or a common influencer, can exert his authority with a certain type of message, to galvanise and induce people to move from involuntary, to voluntary action.

4.1.1 Information theory and dynamics

Spreading the right information to the right audience has always been the underlying goal of proper communication. However, it is hard to estimate whether the intended addressees of the communication have understood the information correctly, in particular when the communication is irregular, disseminated among different channels and/or complex in nature.

Today, public discourses around current events, especially those that occur online, are increasingly influenced by those issues that news media companies choose to concentrate on [171]. Not only do news media have the power to create public awareness around social issues [172, 173] but they can also influence how public perceives the importance of those issues [174, 175].

However, today's news site are mostly designed for a *one-to-many mono-directional* style of communication. Even though, on some sites, readers can comment, share, and rate articles, most of the public reactions about specific issues do not take place on the newspaper itself. As a matter of facts, readers who want to express opinions on particular articles tend to transfer that discussion on social media (Twitter, Facebook, etc.) connecting to the originating source through the usage of "hashtags".

Nowadays, social media are by large the most relevant online communication mean designed for networking [176] and interactive communication. When used effectively, their applications can promote dialogue [177], facilitate information transfer and understanding [178], engage stakeholders [179], and improve communication and collaboration in online environments [180]. In contrast to online news sites, social media are designed for a *many-to-many* and *bi-directional* communication among participants.

In some cases a message may create social media hype on certain topics as part of a well designed communication strategy. This is particularly true in political scenarios when having people (just) talking about something becomes important, even though they do not fully understand the real meaning of the original message. In this case, the communication to be diffused is properly designed to trigger social media activity.

I this case, it is better to focus on those public messages that need to be properly and quickly grasped by the readers. In such cases, it is important for the source to make sure the audience understands what was intended by the communication. Thus, for the source it becomes important to be alerted, as quickly as possible, if this is not the case. And, if necessary, the source may produce additional communication to clarify (e.g., through the usage of Frequently Asked Questions, FAQ.)

From a methodological standpoint, users on social networks often use *hashtags* embedded in their discussion. Thus, we use hashtag related data combined with an unusual usage of the question mark symbols (denoting questions asked) in the tweets to estimate the need for additional explanation about a particular issue [181]. Such analysis are performed within a specific time frame as the readers' interest decay degree over time is also important.

When a message is enticing or alluring, people in social networks may start a deep discussion on the topic. If, however, the alluring or alluring information has been shared to spur people on and make them react (typically in the political sphere), it may mean that some of this information assimilated by followers can be used for other purposes.
The case of Donald Trump on Capitol Hill for example, was an important event, where it was seen how the potential of an influencer with political and social authority, exploited by sending out information, to incite people to criminal acts, thus interfering with public safety, and in this case also with that of the state.

4.1.2 Information as interference and misinformation

Disinformation and interference by Moscow [182] where the editing and dissemination of fake news, social campaigns and the use of trolls are unfortunately establishing themselves as sophisticated and pervasive tools of influence by Russia, but also by other state actors, such as China, risking to pollute and distort the public debate in Italy and in Western countries, and to convey false and unverified news that does not adhere to facts and events. In this scenario, in Italy there is 'a substantial weakness in the interventions to counter disinformation and the various forms of interference. A clear deficit and delay of our country compared to commitments, tools, strategies and measures that have already been operational for some time in the international context'.

It is possible to define many ways of communicative strategy to interfering in other in the affairs of other countries. The document published from the COPASIR (Italian Parliamentary Committee for

the Security of the Republic) [182], evidence some methods used by foreigners actor during the Italian election campaign in 2022.

1. *Disinformation* all those practices of voluntarily creating and disseminating untrue or misleading information, with the aim of deceiving the recipient of the message
2. *Misinformation* is defines as all those practices of disseminating untrue information, without the knowledge that it is false and, therefore, in the absence of the voluntariness of deceiving the recipient of the message, in media manipulation
3. *Targeting* method by Data-driven systems of large social media companies,
4. *Trolling* defined as the act of creating and disseminating online messages and comments of a violent or defamatory nature, prompting the recipient to an emotional response
5. Harassment directed at users' digital profiles
6. Mass reporting of content and accounts, indirectly exploiting the filtering systems of platforms

These characteristics of information, also give insight into how the dynamics of voluntary or involuntary information sharing works. As described by Beck, or in the work of Battistelli & Galatino, voluntariness or the knowledge that that information is false, can change the dynamics of information sharing within the reference network.

Assuming, therefore, a social system, where one part of the people knows that information is false, while another does not, entails a dynamic that goes beyond the individual parts of the system itself, as communication and interaction between people, entails a dynamic, that dynamics produce events of polarisation, homophily or others effect, that will limit or enhance the dissemination of information in the same and other networks.

4.2 Complex systems

The most famous quote about Complex Systems comes from Aristotle, who said that

> "The whole is more than the sum of its parts"

Complex systems are systems where the collective behavior of their parts entails emergence of properties that can hardly, if not at all, be inferred from properties of the parts. Examples of complex systems include ant-hills, ants themselves, human economies, climate, nervous systems, cells and living things, including human beings, as well as modern energy or telecommunication infrastructures. [183]

Various scientists have discussed and talked about complex systems, and hypothesised models for understanding those dynamics that usually escape the human eye and brain, or, dynamics for predicting the future.

Simon de Laplace for example, who wrote in 1776:

> The present state of the system of nature evidently follows from what it was at the previous instant, and if we were to imagine an intelligence that at a given instant understood all the relations between the entities of this universe, it could know the respective positions, the motions and the general dispositions of all those entities at any instant in the past and future... But ignorance of the various causes that concur in the formation of events as well as their complexity, together with the imperfection of analysis, prevent us from attaining the same certainty with respect to the great majority of phenomena. There are therefore things that are uncertain for us, things that are more or less probable, and we try to make up for the impossibility of knowing them by determining their varying degrees of verisimilitude. It thus happens that to the weakness of the human mind we owe one of the finest and most ingenious of mathematical theories, the science of chance or probability.

According to Laplace, knowing the positions and velocities of all the particles in the universe, and the laws governing their relationships, it would be possible to predict their evolution for eternity. Laplace's conception configures probability in the description of physical processes as accidental, linked to our ignorance, but in principle elusive. But in 1903 the great mathematician, physicist and philosopher of science Henri Poincarè wrote:

> A very small cause that escapes our attention causes a considerable effect, which we cannot fail to see, and we then say that the effect is due to chance. If we knew exactly the laws of nature and the situation of the universe at the initial instant, we could predict exactly the situation of the same universe at a later instant. But even if it were the case that the laws of nature no longer held any secrets for us, even then we could only know the initial situation approximately. If this allowed us to predict the subsequent situation with the same approximation, we would need no more and we would have to say that the phenomenon has been predicted, that it is governed by laws. But this is not always the case: it can happen that small differences in the initial conditions produce very large ones in the final phenomena. A small error in the former produces a huge error in the latter. Prediction becomes impossible, and you have a fortuitous phenomenon.

However, I would like to make it clear that the idea of Laplace's demon is madness from an entropic, dynamic point of view with respect to the non-linearity of the interaction between the individual

components, and also on a thermodynamic level. I wanted to include Laplace in my thesis because as described above, as a matter of logic, this 'demon' can exist, and in fact deceives the reader's mind. But starting from this basis, I would like to highlight the efforts of many other scientists to show the arduous paradigm shift.

The extreme sensitivity to initial conditions described by Poincarè paves the way for the modern concepts of 'deterministic chaos' and 'complexity'. In its most general sense, the concept of complexity challenges the idea that the study of complex systems can in any case be traced back to the study of their constituents.

Thus, a small initial uncertainty causes the system to become unpredictable after a short time, but this results (deterministic chaos) in the loss of knowledge of the initial conditions.Since we do not control these small uncertainties, it is impossible to predict their future evolution.

Therefore, the goal of complex systems, is to study the result of nonlinear dynamics, called *emergent behavior*, therefore, that something that lies beyond the sum of its individual parts.

On complexity science, there are complex dynamics, which are more or less omnipresent in all dynamic systems.
Some already mentioned such as, polarisation and homophily are common examples for social dynamics, but emergent behaviours of Collective Intelligence (CI), Collective behavior (CB), Self-Organised Criticality (SOC) or Stochastic Resonance (SC) are common in many natural and social events.

Self-organized criticality (SOC) is a property of dynamical systems that have a critical point as an attractor. The concept was put forward by Per Bak, Chao Tang and Kurt Wiesenfeld ("BTW") in a paper [184] published in 1987 in Physical Review Letters, and is considered to be one of the mechanisms by which complexity [185] arises in nature. Its concepts have been applied across fields as diverse as geophysics,[186] physical cosmology, evolutionary biology and ecology, bio-inspired computing and optimization (mathematics), economics, quantum gravity, sociology, solar physics, plasma physics, neurobiology [187, 188, 189, 190] and others.
SOC is typically observed in slowly driven non-equilibrium systems with many degrees of freedom and strongly nonlinear dynamics. Many individual examples have been identified since BTW's original paper, but to date there is no known set of general characteristics that guarantee a system will display SOC.

The macroscopic behaviour of systems with self-generated criticality exhibits spatial and temporal scale invariance typical of a phase transition, e.g. 1/f noise. Unlike a common critical point which is reached by an external adjustment of the order parameter, systems exhibiting self-regenerated criticality spontaneously stay close to the critical point, hence the name.

Self-organised critical phenomena can be observed in non-equilibrium systems with many extended degrees of freedom and with some degree of non-linearity. Many examples validating this concept have been identified since the original paper, but there is still no agreement on the necessary and sufficient conditions for a system to exhibit self-organised criticality.

Collective behavior (CB) refer to social processes and events which do not reflect existing social structure (laws, conventions, and institutions), but which emerge in a "spontaneous" way. Use of the term has been expanded to include reference to cells, social animals like birds and fish, and insects including ants.[191] Collective behavior takes many forms but generally violates societal norms [192, 193]. A Mexican wave is effectively a CB in that the behaviour is simulated by many others after a certain stimulus. But the initiating movement, i.e. the single point that stimulates interaction with others is a gesture (the *hola*) that is hardly used when there is no goal (in football). Collective behavior can be tremendously destructive, as with riots or mob violence, silly, as with fads, or anywhere in between. Collective behavior is always driven by group dynamics, encouraging people to engage in acts they might consider unthinkable under typical social circumstances.[192] Hence, collective behaviour is a more or less spontaneous social behaviour that numerous individuals manifest at the same time, in the presence of the same stimulus or related situations, whether they are gathered in one place, as is typically the case with crowds, or physically separated and dispersed, as is the case with social movements or, in a more restricted sphere, with fashion. It is possible to distinguish collective behaviour from one another is to stick to the manifest form it takes, since it usually corresponds to a particular mode of unfolding, but given the complexity of the event itself, and the dynamics it brings, it is difficult to be sure that the result of collective behaviour always derives from the same initial social condition.

I will take a lot of inspiration from theories and past events of collective behaviour, as it is an integral part of my thesis to find a method to avoid collective behaviour that can be a threat to society, organisations and institutions.

Social scientists have developed theories to explain crowd behavior.

- Contagion theory – the Contagion Theory was formulated by Gustave Le Bon. According to Le Bon crowds exert a hypnotic influence over their members. Shielded by their anonymity, large numbers of people abandon personal responsibility and surrender to the contagious emotions of the crowd. A crowd thus assumes a life of its own, stirring up emotions and driving people toward irrational, even violent action.[194] Le Bon's Theory, although one of the earliest explanations of crowd behavior, is still accepted by many people outside of sociology [195], but the bulk of media studies has strongly rejected his theory. However, critics argue that the "collective mind" has not been documented by systematic studies. Furthermore, although collective behavior may involve strong emotions, such feelings are not necessarily irrational. Turner and Killian argue convincingly that the "contagion" never

actually occurs and participants in collective behavior do not lose their ability to think rationally.[196]

- Convergence theory – whereas the Contagion Theory states that crowds cause people to act in a certain way, Convergence theory states that people who want to act in a certain way come together to form crowds. Developed by Floyd Allport [197] and later expanded upon by Neil Miller and John Dollard as "Learning Theory", [198] the central argument of all convergence theories is that collective behavior reveals the otherwise hidden tendencies of the individuals who take part in the episode. It asserts that people with similar attributes find other like-minded persons with whom they can release these underlying tendencies. People sometimes do things in a crowd that they would not have the courage to do alone because crowds can diffuse responsibility but the behavior itself is claimed to originate within the individuals. Crowds, in addition, can intensify a sentiment simply by creating a critical mass of like-minded people.
- Emergent-norm theory – according to Ralph Turner and Lewis Killian,[196] crowds begin as collectivities composed of people with mixed interests and motives. Especially in the case of less stable crowds—expressive, acting and protest crowds—norms may be vague and changing, as when one person decides to break the glass windows of a store and others join in and begin looting merchandise. When people find themselves in a situation that is vague, ambiguous, or confusing new norms "emerge" on the spot and people follow those emergent norms, which may be at odds with normal social behavior. Turner and Killian further argue that there are several different categories of participants, all of whom follow different patterns of behavior due to their differing motivations.
- Value-added theory – Neil Smelser argues that collective behavior is actually a sort of release valve for built-up tension ("strain") within the social system, community, or group.[199] If the proper determinants are present then collective behavior becomes inevitable. Conversely, if any of the key determinants are not present no collective behavior will occur unless and until the missing determinants fall into place. These are primarily social, although physical factors such as location and weather may also contribute to or hinder the development of collective behavior.
- Complex Adaptive Systems theory – Dutch scholar Jaap van Ginneken claims that contagion, convergence and emergent norms are just instances of the synergy, emergence and autopoiesis or self-creation of patterns and new entities typical for the newly discovered meta-category of complex adaptive systems.[200] This also helps explain the key role of salient details and path-dependence in rapid shifts.

But, a special place among collective behaviour is occupied by hearsay or rumor. Hearsay - a succession of uncontrolled rumours, partly true and partly false, but without the possibility of

distinguishing between the false and the true - may constitute collective behaviour on its own, and run its course without being accompanied by other forms of manifest behaviour. Other times, however, rumour constitutes the main information system between individuals involved in collective behaviour, as is characteristically the case in panic episodes of economic, war or catastrophic origin.

Collective or group intelligence (CI) is manifested by the fact that a team of cooperating agents can solve problems more efficiently than when these agents work in isolation.[201] The concept of collective intelligence has been used to address very diverse groups of agents: insects living in colonies, teams of humans, collaborative robots, although in the latter case it would be more appropriate to speak of distributed intelligence.

For Pierre Lévy, it is a question of [202]

> "intelligence that is distributed everywhere, constantly enhanced, coordinated in real time, and which results in the effective mobilisation of skills"

It results, among other things, from the quality of the interactions between its members (or agents). While the knowledge of the members of a community is limited, as is their perception of the common environment, and while they are not aware of all the elements relevant to the goals, agents can accomplish complex tasks or find innovative solutions thanks to different mechanisms, methods..., such as *stigmergy* [10] [203]. Forms of collective intelligence are very diverse depending on the type of community and the members it brings together. Human groups, in particular, do not obey such mechanical rules as other collectives, for example social animals such as insect colonies or associations of cooperative robots.

The concept of collective intelligence can be studied as a particular example of the manifestation of emergent behaviour that takes place in non-linear dynamical systems (such as flocks of birds or fractal systems). In systems of this kind, the atomic parts that represent the primitive, constituent elements of the whole, taken in their own right, possess properties and functionalities that distinguish them in a univocal, linear manner. But as soon as a large number of these primitive elements aggregate in such a way as to form a system and reach a critical threshold, as a result of the relationships established between them, properties and behaviours begin to manifest themselves in the overall aggregate, often of a non-linear type, of which there was no trace in the atomic elements and which therefore denote so-called emergent behaviour.

The *Emergence* occurs when an entity is observed to have properties its parts do not have on their own, properties or behaviors that emerge only when the parts interact in a wider whole. The Emergent behaviour, then, occurs whenever a high-level pattern or configuration originates from the thousands of simple interactions that occur between local agents, therefore the interactions within

[10] Stigmergy is a communication method used in decentralised systems whereby individuals in the system communicate with each other by modifying their surroundings. Stigmergy was first observed in nature; for example, ants communicate with each other by leaving a pheromone trail, so an ant colony is an example of a stigmergic system.

the system define whether the system has emergent properties. Such systems (a complex systems) – and the self-organization and emergent phenomena they manifest – give the amount of action and dynamics, lie at the heart of many challenges of my thesis, my discipline and global importance for the future of the worldwide knowledge society.

Steven Johnson speaks of emergent systems considering bottom-up, i.e. bottom-up, self-organising mechanisms, focusing on connections.[204] Taken individually, an ant or a neuron is not particularly intelligent. However, if a large enough number of such simple elements interact and self-organise, a unitary, complex and intelligent collective behaviour, also referred to as swarm intelligence, can be activated. If this behaviour also has an adaptive value, we are dealing with an 'emergent' phenomenon like an ant colony or our brain.

Collective intelligence can be interpreted, in the light of these reflections, as precisely a systematic aggregate of individual intelligences, whose reciprocal relations and collaboration produce massive cultural, sociological, political and anthropological effects of an emergent kind that are difficult to study with the criteria applied to the individuals who are part of it. In his 1995 book Out of Control, [205] Kevin Kelly argues that artificial machines and social systems are reaching such a level of complexity that they will soon no longer be distinguishable from biological apparatuses. In this regard, Kelly speaks of a kind of global mind emerging from a techno-cultural network integration. However, I believe this assumption made by the author is overly simplistic. Even if one possesses knowledge of all possible variables in a simulation, be it biological, chemical, or social, it remains exceptionally challenging, if not impossible, to fully grasp and comprehend all the intricacies and dynamics within the simulation. In fact, I contend that it is far more articulate than we can imagine to discover emergent behaviors in simulations, even with consideration of all variables, that exhibit true freedom and independence from the simulation itself. This is precisely what occurs in the real world, as it is influenced by various ecosystems, environments, and unpredictable factors that we still struggle to fathom and predict with accuracy.

The new media are often associated with the promotion and enhancement of collective intelligence. Their ability to easily store and retrieve information, mainly through databases and the Internet, allows them to be shared without difficulty. Thus, through interaction with the new media, knowledge is easily accessed by passing from one source to another and giving rise to forms of collective intelligence. Take for example the case of GameStop, incited by the activity of a nerd-community on reddit or social actions such as rapid flesh mobs etc.

In the theory of nonlinear dynamical systems (particularly in chaos theory) and in the theory of stochastic processes, **stochastic resonance** (SR) is a mathematical mechanism whereby a nonlinear system, immersed in a certain stochastic background noise, becomes sensitive to external perturbations, which are too weak to affect it in the absence of such noise, which contains a wide spectrum of frequencies. The frequencies in the white noise corresponding to the original signal's

frequencies will resonate with each other, amplifying the original signal while not amplifying the rest of the white noise – thereby increasing the signal-to-noise ratio, which makes the original signal more prominent. Further, the added white noise can be enough to be detectable by the sensor, which can then filter it out to effectively detect the original, previously undetectable signal. [206] [207] In order for such a mechanism to take place, nonlinearity is crucial.[208] Originally proposed in the context of climate dynamics,[209, 210] with the passage of time it has assumed great importance in numerous fields,[211] notably in information theory [212], neuroscience [208] and social science [211]. Stochastic resonance was first proposed by the Italian physicists Roberto Benzi, Alfonso Sutera and Angelo Vulpiani in 1981, and the first application they proposed (together with Giorgio Parisi) was in the context of climate dynamics. [206] Although this is a very interesting topic for complex social dynamics, it is unfortunately not an integral part of my thesis (only in marginal situations). One of the earliest usages of the term "complexity", in the social and behavioral sciences, to refer specifically to a complex system is found in the study of modern organizations and management studies. However, particularly in management studies, the term often has been used in a metaphorical rather than in a qualitative or quantitative theoretical manner. By the mid-1990s, the "complexity turn" in social sciences begins as some of the same tools generally used in complexity science are incorporated into the social sciences.

A case of stochastic resonance in the social sphere may have been the event of the Bologna massacre. Unfortunately, it is not possible to prove this, but I hypothesise (with all the possibility of being wrong) that the explosion of the bomb, turned a very unstable political and social situation into a more stable and balanced one. The Bologna massacre was an attack committed on Saturday 2 August 1980 at 10.25 a.m. at "Bologna Centrale" railway station in Bologna, Italy. In the attack 85 people were killed and over 200 injured. It was the most serious terrorist act that occurred in the country after World War II and has been referred to by many as one of the last acts of the Tension Strategy. The strategy of tension is precisely that political and social instability I described earlier. The strategy of tension in Italy is a political theory that generally refers to a very tormented period in the history of the Italian Republic, particularly in the 1970s, known as the 'anni di piombo' and which, through a subversive design, tended towards the destabilisation or disintegration of pre-established balances. The strategy is based on a prearranged series of terrorist acts aimed at spreading a state of insecurity and fear in the population, such as to make people justify, demand or hope for authoritarian-style political turns; it can also be implemented in the form of military tactics consisting of committing bomb attacks and attributing the blame to others. The noise of the event, brought a situation of balance in the Italian political and social system. In fact, in that period, various types of polarising events (the Reggio Calabria uprising or the national political system) led to tension in various Italian hotspots, including, reggio calabria, rome, emilia romagna and milan.

4.2.1 Social complexity

All these different dynamics are part of what is called social complexity. Social complexity is a conceptual framework used in the analysis of society. Contemporary definitions of complexity in the sciences are found in relation to systems theory, in which a phenomenon under study has many parts and many possible arrangements of relationships between the parts. At the same time, the complex and the simple are relative and can change over time. [213] Current usage of the term "complexity" in the field of sociology typically refers specifically to theories of society as a complex adaptive system. However, social complexity and its emergent properties are central recurring themes throughout the historical development of social thought and the study of social change. [214]

This emphasis on interconnectivity in social relationships and the emergence of new properties within society is found in theoretical thinking in multiple areas of sociology.[215] As a theoretical tool, social complexity theory serves as a basis for the connection of micro- and macro-level social phenomena, providing a meso-level or middle-range theoretical platform for hypothesis formation. [216] Methodologically, social complexity is theory-neutral, which means that it is adaptable to both local and global approaches to sociological research, also applied in networks and simulations.

In the first decade of the 21st century, the variety of application areas has increased [217] with the development of more sophisticated methods. Social complexity theory has been applied to studies of social cooperation and public goods; [218] altruism; [219] education; [220] global civil society; [221] collective action and social movements; [222, 223] social inequality; [224] work and unemployment; [225, 226] policy analysis; [227, 228] health systems; [229] innovation and social change, [230, 231] to name a few. An ongoing international scientific research project, Seshat: Global History Databank, is explicitly designed to analyse changes in social complexity from the Neolithic Revolution to the Industrial Revolution.

As an intermediate theoretical platform, social complexity can be applied to any research in which social interactions or the effects of those interactions can be observed, but particularly when they can be measured and expressed as continuous or discrete data points. A frequently cited criticism of the usefulness of complexity science in sociology is the difficulty of obtaining sufficient data [232]. However, the application of the concept of social complexity and the analysis of this complexity began and continues to be an ongoing area of research in sociology. From childhood friendships and teenage pregnancy [214] to criminology [233] and counterterrorism [234], theories of social complexity are applied to almost every area of sociological research.

In communications and computer science research, the concept of self-organizing systems emerged in the mid-1990s in scientific communications research[235]. Scientific and bibliometrics are

research areas where discrete data exist, as are many other areas of social communications research, such as sociolinguistics [214]. Social complexity is also a concept used in semiotics [236].

In an historical and practical terms, if the world had found a balance before the Second World War, between more or less equal forces, things changed with the Manhattan Project and the atomic bomb. The atomic bomb, changed the geopolitical relations between states, for example, between those that had or could create themselves an atomic bomb, and those that for political (Italy and Germany, the losing states of the IIWW) or economic (third world countries) cases could not get it.
As explained in the previous chapter, with the example of the BRICS, relations (hence interactions) between states can change from the actions taken. The Russia-Ukraine case is an example. But in the case of the atomic bomb, a special factor was added.
The atomic bomb not only shows an authoritarian position of who has it, and who can use it, but can drastically change global state-to-state relations precisely because of its destructive potential.
In fact, the complex dynamic that I would like to highlight, lies in the fact that the atomic bomb, has brought an additional layer not on the scientific field, but on the political and geopolitical field. This (it is not the first one, there have been similar cases, such as with smallpox and vaccination), places on the field how a highly and purely scientific element, goes out of its sphere, and also influences the social, economic and political one. Indeed, the entire field of scientific and technological studies suggests that this is now not the exception, but the norm.
Oppenheimer was one of the four in the Manhattan project, he pointed out after the Hiroshima event that science (physics in this case) has gone far beyond its own field, and has succeeded in bending even sectors such as politics and economics.

> 'Now I am become Death, the destroyer of worlds' - J. Robert Oppenheimer

Since this event, the world has become a little more complex. Oppenheimer basically added a variable in the social world, and in fact, he refused to work on the hydrogen bomb given the results of the atomic bomb.

Another example, could be in geopolitical level, the Ukrainian victories of mid-September (2022). This wars-victories were also strategically conceived in view of the Samarkand meeting in 2022 between Putin and Xi Jinping, where Putin's image and authority was marginally reduced given the recent defeats on the ground. Indeed, the various leaders, including the most important ones (Xi Jinping and Narendra Modi) showed a cooling of relations and an uncommon vision in the near-medium future [237, 238]. The Ukrainians, however, by succeeding in making Russia lose territory, and thus showing Putin as not a strong leader, have embarrassed Xi Jinping (since they are allied with the Russians against the West) by doing the US a favour in electing Xi Jinping to the next Communist Party convention in late October 2022. [239]

Leaving the political and geopolitical world, however, and relying mainly on complexity science, we can found many social complex dynamics, almost present in many non-liner dynamics:

Homophily (i.e., "the love of similarity/equality") is the tendency of individuals to associate and form bonds with others who are considered similar because of some peculiar characteristics. The presence of homophily has been described in a wide range of studies which have observed homophily in various contexts and modalities and established a similarity connection [240] in different domains including age, gender identity, social class, and organizational role.[241]

Individuals in homophilic relationships share common characteristics (beliefs, values, education, etc.) that make communication and relationship formation easier. Homophily often leads to homogamy-marriage between people with similar characteristics. [242] In the context of social network analysis, the term used is closest to the original in sociology: "homophily" here corresponds to the phenomenon caused by the similarity between individuals in one or more relevant characteristics, within the social network in which they are interconnected with each other. [243] Robert K. Merton and Paul Lazarsfeld distinguish between "status homophily" and "value homophily": the former describes individuals with similarities in social status who, given their characteristics in common, are more likely to associate with one another; the latter, on the other hand, refers to the tendency to associate with other like-minded people, regardless of status differences.[243]

The opposite of homophily, here is heterophily or intermingling, witch means from a sociological perspective includes the various forms of weak or absent ties interactions between individuals that go against a particular society's cultural norms.

Another important dynamic that can be found in social groups (especially online groups), is polarization. Polarization is the result of a dynamic that distances one group of people, relative to another starting from the fact that the values or various characteristics that each group possesses, are different, dividing the extended group into two or more groups. Polarization is the opposite effect of homophily. The polarization of people into groups, affects various personal spheres, such as political, economic and social.

Digital media, and particularly social media, could potentially play a role in encouraging social polarization.[244] This is because social media sites like Facebook can help cluster friends and acquaintances into homophilous circles, and social news sites like Digg can facilitate a consumption of news that is biased by its user's choices. In the extreme, a lack of "a common public sphere" could lead to isolated, polarised groups which could even be hostile towards one another.[245] For example, during the Arab Spring uprisings, it was observed that social media furthered the social stratification already present in several Arab states. Polarization observed in a particular social media site need not necessarily be a result of events and discussions that happen on that platform. For example, observed trends of polarization in online social media may therefore emerge from activities of users in other online platforms or offline activities.

Polarisation also depends very much on the quality of the information, hence on the credibility of the message shared. As an instance from a 2019 study, messages propagating anti-climate change beliefs on Twitter were collectively found to carry no credibility.[246] Hence it is highly unlikely that such messages which are not credible can increase polarization of climate change opinions on Twitter, this also highlights from my point of view how, the credibility of the message has an impact on how close some people are to a piece of information. Indeed, the aim of my thesis is that the role of influencers within the social sample, gives greater credibility (given the authority of the influencer) to the news, thus influencing the possibility of reducing or increasing polarisation.

Beyond social media, however, polarisation is also 'naturally' found in the non-cybernetic realm, such as the field of politics. Most discussions of polarization in political science consider polarization in the context of political parties and democratic systems of government. In two-party systems, political polarization usually embodies the tension of its binary political ideologies and partisan identities.[247, 248, 249, 250, 251] However, some political scientists assert that contemporary polarization depends less on policy differences on a left and right scale, but increasingly on other divisions such as: religious against secular; nationalist against globalist; traditional against modern; or rural against urban.[252] Polarization is associated with the process of politicization, [253] which can also result from a non-sharing of ideals, values and sometimes social norms, thus creating a condition of socio-cultural polarisation.

Political scientists have shown politicians have an incentive to advance and support polarised positions. [254] These argue that during the early 1990s, the Republican Party used polarizing tactics to become the majority party in the United States House of Representatives—which political scientists Thomas E. Mann and Norman Ornstein refer to as Newt Gingrich's "guerrilla war." [255] What political scientists have found is that moderates are less likely to run than are candidates who are in line with party doctrine, otherwise known as "party fit." [256] Other theories state politicians who cater to more extreme groups within their party tend to be more successful, helping them stay in office while simultaneously pulling their constituency toward a polar extreme. [257]A study by Nicholson (2012) found voters are more polarised by contentious statements from leaders of the opposing party than from the leaders of their own party. As a result, political leaders may be more likely to take polarised stances. [258]

Another interesting dynamic that can be found in social networks and social sciences are Echo chambers. The Echo chamber is a metaphorical description of a situation in which information, ideas, or beliefs are amplified or reinforced by communication and repetition within a defined system. Within a figurative Echo chamber, official sources are often no longer questioned and different or competing views are censored, disallowed, or otherwise underrepresented. It originates from the physical phenomenon of echo, where sounds reverberate in a hollow enclosure. The phenomenon is particularly evident in the case of social media and the use of it by politicians, institutions and

other organizations for the purpose of circulating their own messages to the detriment of others, including hoaxes of various kinds. The mechanism is especially amplified when a person's circle of friends and acquaintances, as is often the case, share similar ideas and thoughts. In this way, news, articles, and comments will appear on the social page that will contribute more and more to amplify a one-sided and uncritical view on that topic.

In fact, Echo chambers may increase social and political polarization and extremism. [259] On social media, it is thought that Echo chambers limit exposure to diverse perspectives, and favour and reinforce presupposed narratives and ideologies.[64, 65]

John Scruggs, a lobbyist for the cigarette manufacturer Philip Morris, described in 1998 two "Echo chamber" mechanisms [66]. The first one consists in the repetition of the same message by different sources. The second mechanism consists in the diffusion of similar but complementary messages by a single source. Scruggs describes the Echo chamber as a strategy for increasing the credibility of certain information in the eyes of a target audience.

Often, individuals isolated within media Echo chambers are drawn into an intellectual and informational confinement, constructed according to their worldview, which reinforces their beliefs and leads to an inability to critically and constructively consider the opinions they are presented with. Thus, these people are less able to judge the quality of the information to which they are exposed.

Filter bubble and Echo chamber, as elements of the online social structure, have interactive effects that are relevant for the understanding of online radicalization [68], and Echo chamber, as described, can lead to the radicalisation of people [67]. This passage is important in the following chapters, where most of the experimental work will take place in radicalised Echo chamber.

4.3 Computational social science

One of the most widely used methods to study complex systems, is the use of computers, yielding data collection. This computational methodological use for the social sciences is called Computational Social Science (CSS) [260], that is, the exploitation of the potential of computers for the study of social events (complex or not), but also for calculations, analyses, simulations, etc.; Thus, to explain, hypothesise and analyse social phenomena, such as economic dynamics, social change, crime and deviance, urban planning, political science, communication, information, etc.

Since the big-data are large in themselves, a part of them can be "sensible". Later on we will see the case of Cambridge Analytica and what it entailed, but for now it is important to dwell on the organisational necessity that the massive amount of data obliges us to respect.

In the above-mentioned event of the 'Manhattan Project', a higher organisational necessity appeared for the first time than in the various projects of the past, because physical, engineering, political,

national security (secret services - remember that the Second World War was still in progress), economic, information (newspapers and TV), transport infrastructure and military bodies were brought together for the first time with a high degree of dimensionality. All this gave rise to what was called 'Big Science', i.e. not only as the English dictionary *Collins* defines it: 'scientific research that requires a large investment of capital', but also because of the political-social-economic-security interweaving.

Already in the 2010s, but especially in 2022, in the midst of the approach and application of big-data, amidst pandemics, physical and information wars, political and economic crises, the computational approach to social sciences is beginning to take on what is known as 'Big Social Science' [261], i.e. similarly to what happened with the Manhattan Project, political-social-economic-security interactions are closely intertwined, and having an organisation and a security perimeter to work with is necessary not only to optimise work, but also to be able to avoid external and internal influences that can undermine the national security perimeter.

The use of collection and processing from social data is widely used in various fields, such as politics for elections, in business to better predict and analyse the economic spectrum of competitors and partners, for one's own products/services, but also in general for research or to improve citizen services. In social networks such as Twitter, Facebook, Telegram and others, it is possible not only to collect data, but if you have a large number of accounts, you can also try to influence, for better or worse, the evaluation of hashtags or users. This method originated in politics during the US national elections with Obama, where through massive use of information soft-power could be used to gather consensus. The same methodology has since been used by other international leaders, such as Trump, Matteo Salvini [262, 263], Matteo Renzi [264], but we can say that all the most relevant political parties in international field, they more or less all use a system of information control, although, some do it to improve the image of themselves [265], others to worsen that of others. Of course, it is possible to do both at the same time. [266]

It is not only the political sector that has moved, but also the private sector [267], in fact, there are not a few cases in which CEOs of companies arm themselves with small armies to influence communities in social networks. One of the latest cases is that of Paris Saint-Germain Football Club [268], where the company has made it known that it already has a small army of accounts for the company's objectives. From a strategic point of view, this is very similar to the strategy of nuclear deterrence, i.e. to deter any kind of aggression. The company justified this news, to defend its players Mbappé and Neymar from homophobic attacks that may reduce the players' performance (given the psychosomatic pressure). The same kind of justification can also be found in the political sphere, i.e. the various leaders can take advantage of the Big Social Science complex, to attack and project themselves, but above all not to die in the social world, which even if it is true that it is only

a fraction of society and does not reflect the real country, it is still an additional level to be taken into account for social, and above all political, dynamics.

In fact, the various organisations that deal with credibility, authority and, above all, the image of the person, when taking data from the various social networks, must take into account what has been described in the previous chapters, i.e. take into account the social dynamic aspect, the structure and topology of the network, the actors in the field and the possible unforeseeable spin-offs given the previous action taken.

All this reasoning closely resembles the principles of intelligence, i.e. a similar observation referring to social intuitions and data-based intuitions, with the aim of understanding the various factors, dynamics and structure of the field in order to defend oneself and avoid possible threats. In fact, a security perimeter has recently been initiated in companies as well, not only because they can be targets without intelligence, but above all because if attacked (and they show considerable damage), their credibility can be attacked. And credibility, or the 'brand-factor' is often what keeps the company itself going, i.e. based on that relationship of trust between the producer of the good/service and the final consumer of the good/service.

To analyse the value of textual information, and thus a possible impact on people, is used a test analysis technology called Natural language processing (NLP), the NLP is a subfield of linguistics, computer science, and artificial intelligence. The goal is that a computer could be capable of "understanding" the contents of some text, including the contextual nuances of the language within them.

4.3.1 Exploring the application of computational tools in my research

In this chapter, which precedes the methodological part of my thesis, I show my research and publications done through the use of used Computational Social Science method. Some of these same tools, I have used in my thesis as the fundamental basis of both the methodological and also the conceptual part, like the paper on geopolitics, mentioned for explain the difference between immutable and mutable variables (Intelligence and complex systems chapter).

One of the first jobs where I used NLP tools, was related to the dynamics between the requesting of information, and reading newspaper articles. In this paper [269], we have shown that it is possible to calculate and, in some cases, predict, how many articles a newspaper has to write, to satisfy the readers' demand for information, based on social network data. In particular, by using Shannon's entropy, it was possible to observe how the behavior of certain hashtags can give insight into the information dynamics between who sent a piece of information and how much of this information was understood by the recipients. This is significant because it is possible to anticipate from the entropic level of the topic, the significance and importance of the topic, well in advance. We have

Table 2: Information Entropy and average tweets/hours.

Type	Topic	Tweets	Hours	Entropy
HEE	DPCM (Apr 26th)	894.500	78	**5.6113**
HEE	BLM	8.4 M	127	**5,5760**
HEE	Beirut	3.2 M	69	**5,1962**
HEE	Debates2020	3.7 M	47	**4,8745**
HEE	Megxit	52.100	63	**4,3127**
HEE	DPCM (Oct 25th)	618.150	42	**4.1972**
LEE	Int.WomensDay	1.5M	21	**4,1457**
LEE	ValentinesDay	812.900	34	**4,0572**
LEE	Tokyo2020	3.1 M	24	**3,6664**
LEE	SuperLega	45.400	27	**4.0541**
LEE	G.dellibro	11.600	14	**3.8232**
LEE	JuveInter	48.500	16	**3.6126**
LEE	MilanNapoli	16.800	10	**3.5672**
LEE	DPCM (Oct 13th)	26.700	12	**3.1569**
LEE	Dupasquier	<10.000	7	**3.0201**
LEE	GazaU.Attack	623.800	8	**2.9740**
LEE	ObiWan	20.300	7	**2,6907**
LEE	SuperLeague	616.700	48	**2.5417**

shown that it is initially possible to distinguish between two different types of entropy events, HEE (high entropy events) and LEE (low entropy events). Table 2 shows the difference in entropy levels between various topics in 2020. After that, we collected data on the display of newspaper articles, and showed how after a certain period of activity on social networks, if the topics are HEE, a shift of activity from social networks, to newspapers, is shown in figure 4.

In another paper [270] , however, I obtained data on space mission hashtags, in order to assess whether it is possible using an equation and various other data, to create a 'Geopolitical space score', i.e. a geopolitical state score, relating to the successes and thus, the prestige of the success or failure of space missions. This work, however, also showed another factor, that even if a mission is successful, it is possible to have negative side-effects. In fact, the data showed us in table 3, that the sentiment of some missions, even if concluded 'positively', have negative sentiment. Such as the case of the uncontrolled re-entry of part of a Chinese rocket, which endangered vast areas, as the same rocket could have fallen on many important cities between the southern and northern hemispheres. Figure 5 shows the Geopolitical Space Score result by years, given by the equation:

$$GSS = \frac{S}{G * B} * (F + Q) \tag{1}$$

S is the Sentiment value from the Twitter event; B is the amount of money invested (budget); G is a difficulty rate associated with the country that launched the space mission, "not because they are easy, but because they are hard" that imply Geopolitical effects; F is related to the success or Failure of the mission; and Q stands for the statistical Quality and difficulty of the country in spaceflight launch organization. The S and Q parameters are evaluated by data scientists through statistical methods. The G score is the only parameter that needs a subjective value because it depends on personal evaluation and hypotheses to evaluate the difficulty rate of reaching that goal.

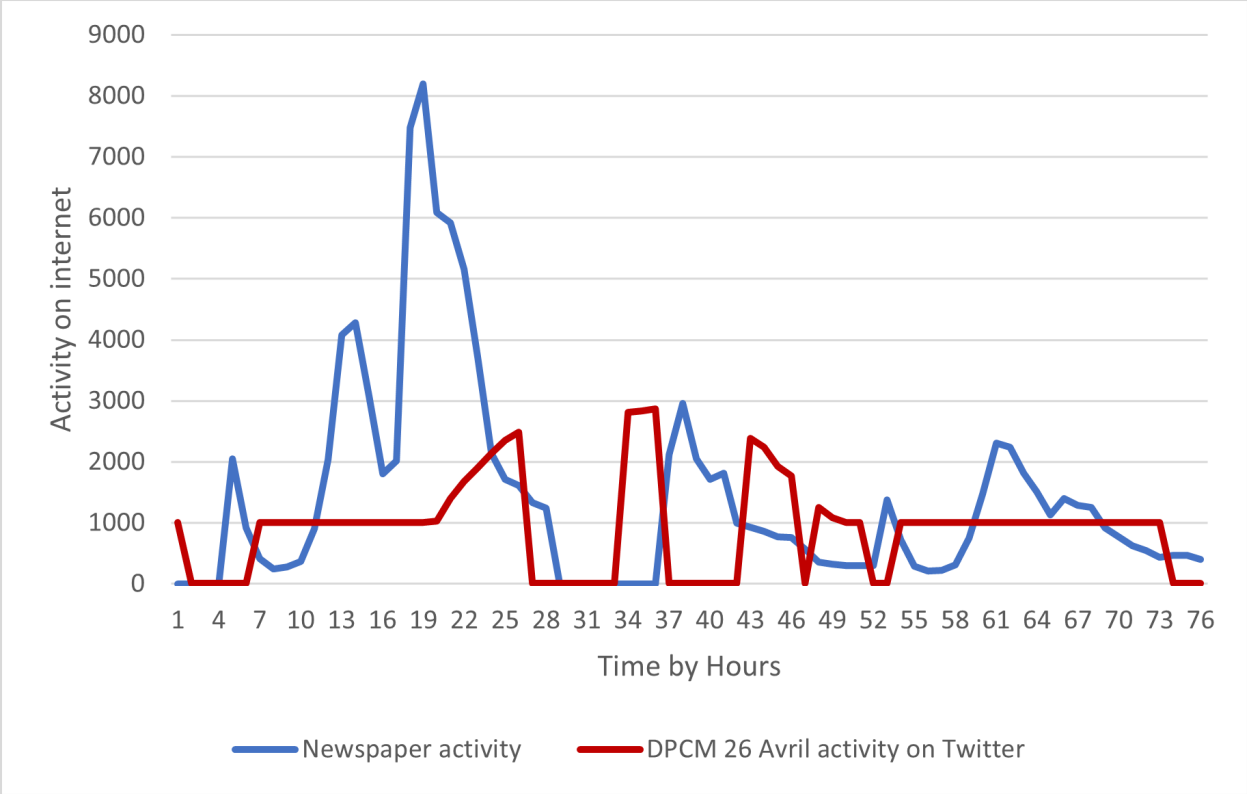

Figure 4: Newspaper activity and SN - 26th April

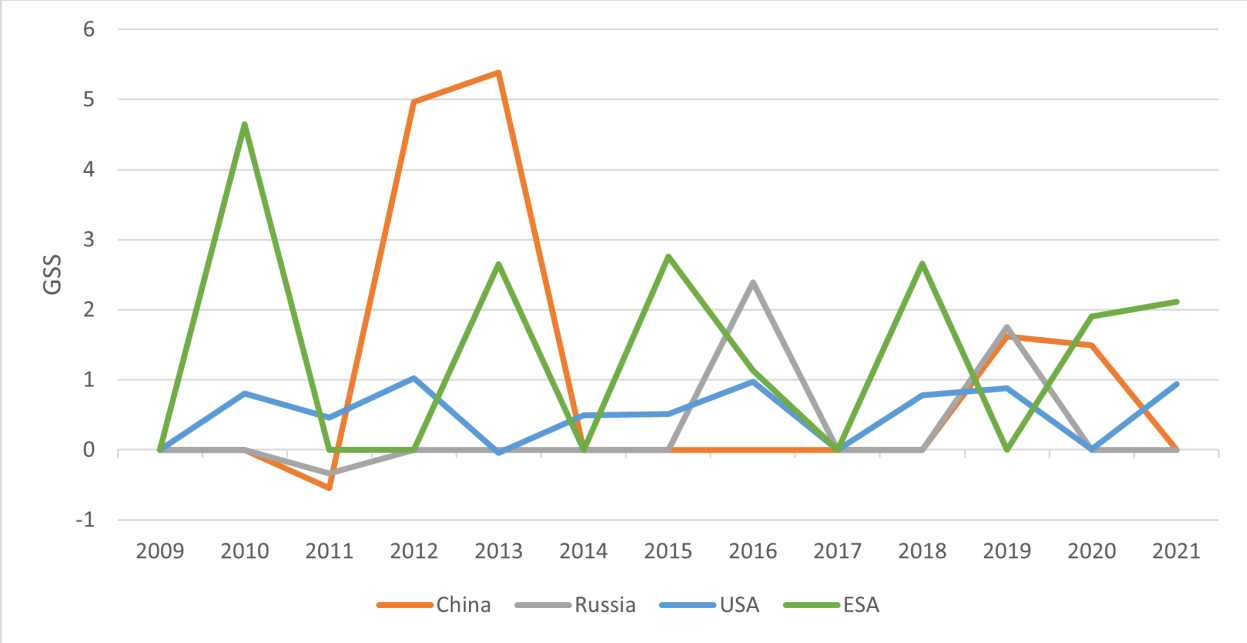

Figure 5: Geopolitical Space Score

Table 3: Geopolitical Space Score

Mission	Geopolitical Space Score
Tianwen-1	9,547438889
ChineseRock	-0,208492857
ASAT	-0,659007937
Perseverance	2,198416733
Mars 2020	2,502628333
JWST	2,469102533
Rosetta	7,588575333

An example of the use of sentiment analysis was my paper on firestorm issues towards companies [271].

Companies that are interested in saving their image and reputation tend to produce quality products or services, both from a B2B and B2C perspective. It happens, however, that every now and then something goes wrong, like *Social Failure* [272, 273] [274] , *Communication Failure* [275] or *Product or service failure* [276], and as consequence the quality level of the product/service falls. [277] [278]

At this time, it can happen that a number of people can speak out in social networks, criticising the brand. If a critical number of people do this, however, the phenomenon can turn into a Firestorm. [274, 275, 276]

Nowadays, a firestorm is a cluster of consumers' digital word of mouth [279, 274, 280] that highlights some communication error, or some terrible mistake made by a company [280]. The definition of Firestorm [281] depends my the context, but in social media is defined as *"a sudden, and sometimes violent reaction"*.

I have been working on a paper in this area [271], because I have noticed that when a company is under pressure from one of the events mentioned above, it is possible for a cyber attacker to attack the cyber part of the company or public organization.

Due to the large number of posts that bots can create within seconds, they can be used to amplify any idea on social networks, influencing political affairs [282] and business company value [283]. In fact, in my thesis, I will use various accounts to increase the authority potential of both individual accounts (what will be called influencers) and shared news, with the aim of being able to impact the social dynamics within the sample. Using bots or secondary accounts, greatly increases the possibility of impact on the group, as the volume of accounts and the impact that a sufficient number of them can have when a critical mass is reached, has the possibility of greatly impacting the sample mainly for two reasons:

- Based on Le bon's studies, a mass of people can change social norms and thus the dynamics;
- The influence on the type of news has a considerable impact based on the entropy of the news (defined as the unexpectedness of the news).

Regarding the first point, however, it must be said that Le bon's studies have been refuted by the bulk of media studies. In fact, the difference between social dynamics in the 'physical/real' sphere is very different from the cybernetic one. However, this does not detract from the fact that if a majority of people follow a certain pattern, there are cases of collective intelligence or copying behaviour by other actors in society. The so-called *Bandwagon effect*. defined as the tendency for people to adopt certain behaviors, styles, or attitudes simply because others are doing so.[284] As also demonstrated by the *Asch conformity experiments*. But returning instead to the impact of bots for firestorm in social networks, they can be used to amplify any idea on social networks also influencing the political and economic sectors. For example, due to a cyber-attack on a Twitter newspaper profile, such newspaper shared a fake news about President Obama being injured by a bomb in the White House, causing a flash-crash in Wall Street and the stop all of economic transactions for some minutes. This led to a loss of about 121 billion dollars for S&P 500 and its related companies [285].

However, during the writing of the paper, an event happened, which we can define as a case study. On December 10, 2020, CD PROJEKT RED released a long awaited game called Cyberpunk 2077. This game was very popular even before its release and it generated continuous social hype from the video game community throughout its development, also winning the "Best Game Awaited" from Golden Joystick Awards for two consecutive years. As shown on Figure 6 and Figure 7, hype for the game substantially increased during the 10 days before the release of the game, reaching its apex on December 10, when the hashtag `#Cyberpunk2077` was tweeted 193,900 times on Twitter, from users of 53 different nationalities. During this timespan, many other hashtags regarding the game were very popular, for example `#Cyberpunk2077Hype` was retweeted 10,000 times.

However, a few days after the release, the Cyberpunk 2077 topic arose again, this time associated with queries related to patches and refunds. In fact, the game was released too early and many bugs were present: due to this, several people had asked a refund to CD PROJEKT RED, often also writing a bad review for the game on online stores like Steam. This created a mood-disease within the company, were in this case, CD PROJEKT RED's employees became stressed and felt pressure related to the quality of Cyberpunk 2077, in which they had invested more than two years of hard work.

In early February 2021, only 60 days after the game's release, CD PROJECT RED was hit by a ransomware attack and attackers were able to exfiltrate the source code of several games, including administrative files [286]. The attackers then threatened the company of leaking or selling the stolen code and files, unless the firm paid a large amount of money to the cybercriminals. In the end, CD PROJECT RED refused to negotiate with the attackers, stating on a press release that they would "not give in to demands or negotiate with the actor", also confirming that no personal information was obtained in the attack and that they were working with law enforcement to track down the

attackers [287, 288]. Later on, security analysts found the stolen source code while being auctioned on the dark web for a minimum price of 1 million USD. The auction was closed after the attackers stated they had received an offer that satisfied them. Within a week of these auctions, the code was shared online via social media, and CD PROJECT RED began using DMCA take down notices to remove posts containing their code.

The social hype that CD PROJEKT RED generated for Cyberpunk 2077 was used by hackers to threaten the company in order to prevent the release of future patches, while damaging the company's reputation to undermine the sales of other long awaited games.

In Table 4 we show the results of the sentiment analysis, obtained from tweets and comments for the hashtag `#CDprojectRED`. Data collected from Twitter respects the timeline of Cyberpunk 2077's release and its developers; data shown in the table can be organised in three categories: before release (October and November), during release (December and January) and after the release of Cyberpunk 2077 (February).

Table 4: Vader Sentiment on #Cyberpunk2077 from Twitter

Months	Negative	Neutral	Positive	Compound
October	0,085	0,757	0,150	0,163
November	0,079	0,766	0,149	0,163
December	0,087	0,750	0,161	0,153
January	0,143	0,758	0,093	-0,111
February	0,104	0,745	0,145	0,120

It is possible to observe that in October and November the sentiment remained neutral-positive with a few oscillations. In December, when the game was released, we can observe a small increase in the negative sentiment due to the high number of bugs present in the game, however, this increment is quite negligible. In January, when a greater number of players were playing the game, the negative sentiment became stronger than the positive one, causing not only a negative compound (-0.111), but also a neutral-negative sentiment for the game and for the developers. Finally, on February the sentiment returned neutral overall, however, the presence of negative sentiment is still stronger compered to the one in October and November.

These data show how much pressure the CD PROJEKT RED company had to experience during the release of the game. Additionally, in Figure 8, we show the financial value of the company during the whole game release timeline, also marking the two critical events that occurred: the yellow line indicates the release of the game, while the red line indicates the ransomware attack. We can see that, after the release of the game, the financial value of the company suffered a sudden drop, that was likely conditioned by customers losing trust in the company due to the presence of many bugs in the game, bad reviews and critics. We can see that the company regains more than half the value lost during the next two months, however, the ransomware attack causes another drop in the

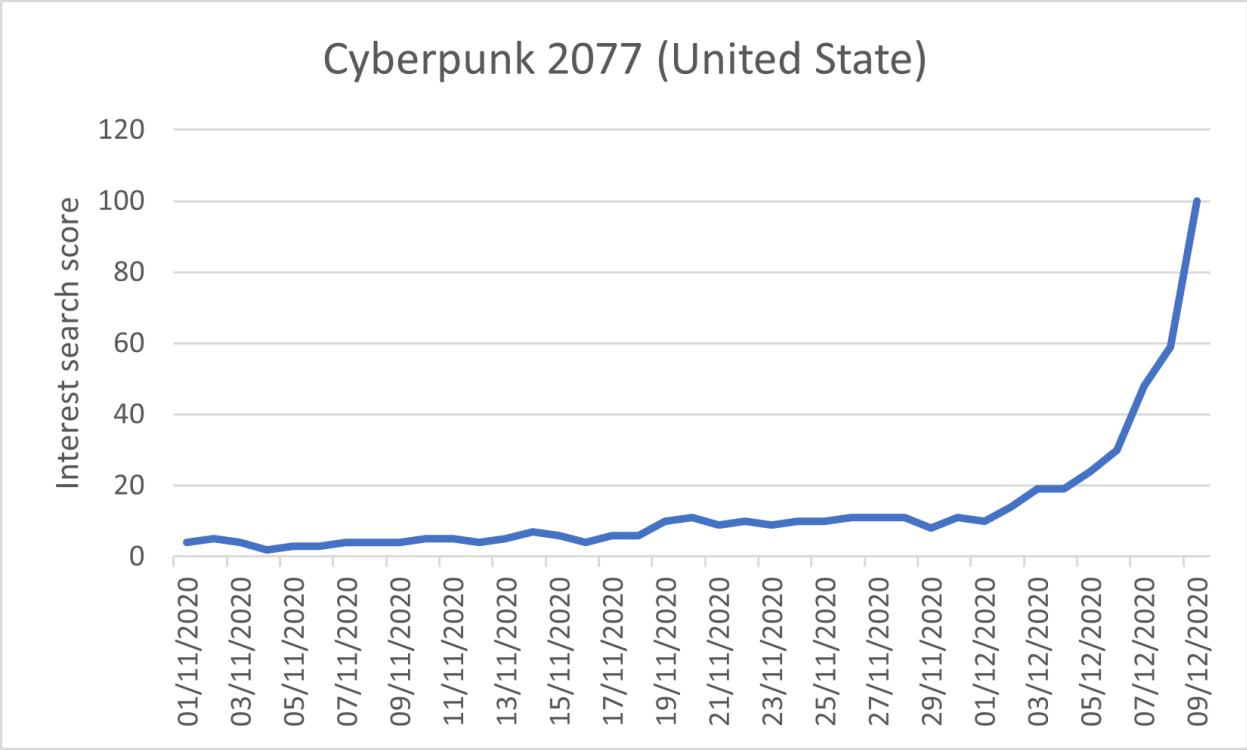

Figure 6: Interest Score showing social hype for the release of Cyberpunk 2077

financial value of the company due to customers losing trust in the company again, this time from a security perspective.

These results show how an organisation (public or private) can be more vulnerable to cybernetic attacks when it is under both social (image and reputation) and managerial pressure.

One last piece of work useful for my thesis, but this time not in the NLP field, is a simulation (currently working in progress) . The aim of the simulation is to understand "how an individual can become an influencer, and starting from the how (i.e. the initial conditions, which the parameters) it is possible to trace back to when he/she will become one".
The simulation features agents and events. The agents are individuals, and each of them has its own characteristics. While events are the messages exchanged in the simulation. All agents can read a message (event) at any time, while only active agents can send one. Influencers are defined dynamically, in relation to the number of followers they acquire. The characteristics of the agents are:

1. The personality of the agent (need to seek news, need to share news, or neutral);
2. Active (can send messages) or Passive (cannot send messages);
3. Ability to embellish (i.e. to create highly appealing messages);

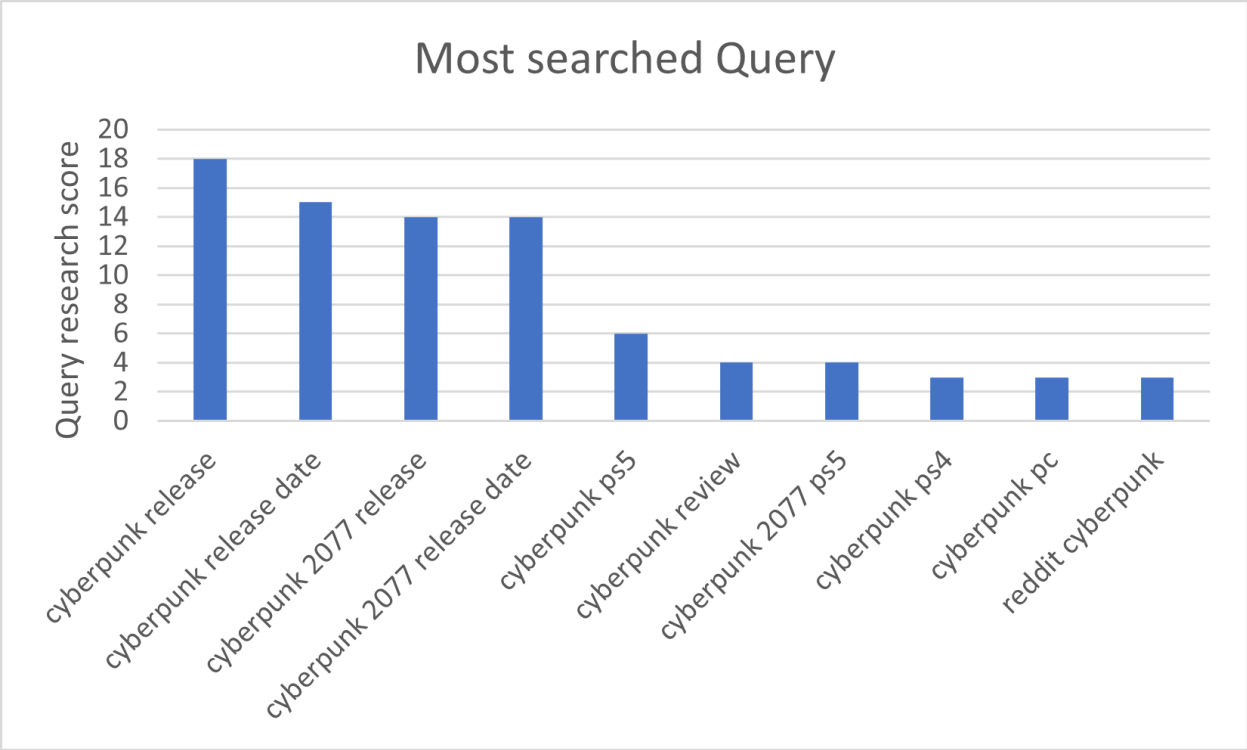

Figure 7: Queries showing social hype for the release of Cyberpunk 2077

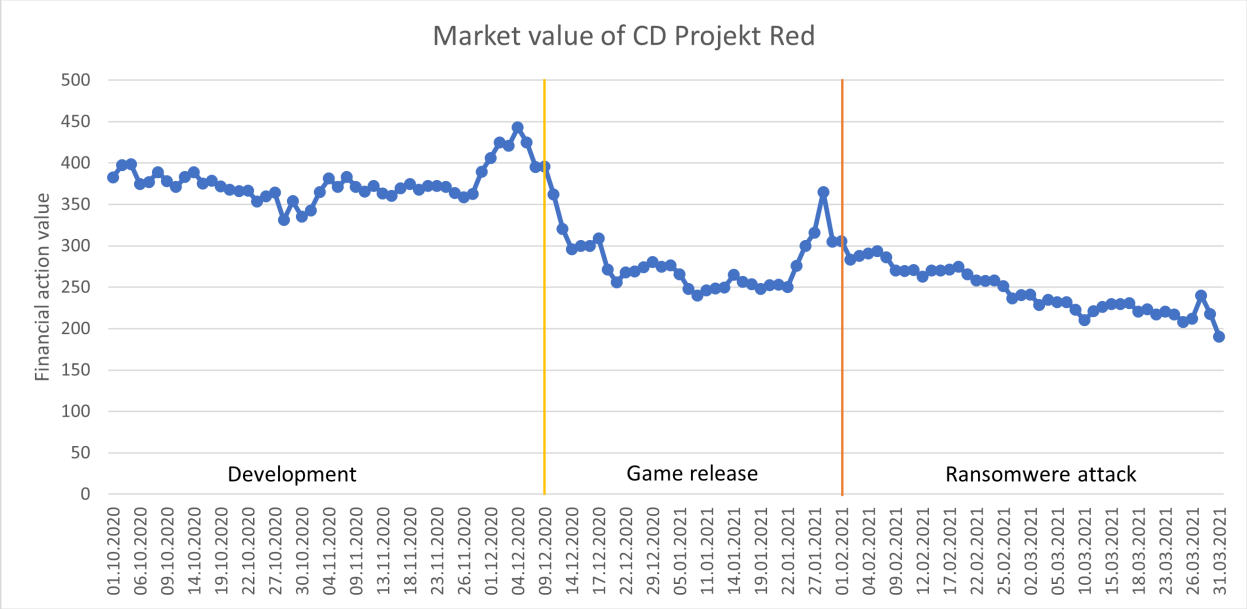

Figure 8: Financial value of CD PROJEKT RED and critical events

4. Interest in each of the N topics;
5. Competence for each of the N topics.

Depending on the dynamics, it will be possible to see who has had more messages and how much will be shared. Sharing, in fact, will allow for the creation of a network, and being an influencer, a person capable of attracting consensus and thus connections to him/her, the individual with the highest message characteristics, and who will also have the hub dimension in the network, is the influencer of the simulation. This simulation, is highly necessary to understand both to justify a social dynamic on how one becomes an influencer, but also to understand more deeply, the dynamic between information and social relations, which allow one to become an influencer.
This part will be necessary during the interaction dynamic, the main subject of my thesis.

These tools used for my work, meant that through the acquisition of information from the people taking part in a particular discussion, I was able to obtain important information about their thinking. On the level of privacy, it is very important that I do not know the sensitive data of the people from whom the data can be taken, even though that is public data and open to all. it is an ethical factor, not a technical one. In fact, in the past, cases such as Cambridge Analytica have caused more attention to be paid to these issues, both on an ethical and national security level.

5 Methodology

In this chapter, we will detail the methodology employed in my research, highlighting the steps taken to achieve my research objectives and the reasons behind my choices.

At the methodological level, it is essential to consider several aspects that play a critical role in understanding social dynamics and the effectiveness of deradicalisation processes. One of the key factors is the availability and collection of accurate data on initial conditions, which forms the foundation for the analysis of the social environment. In addition to this, the integration of innovative technologies such as Natural Language Processing, Artificial Intelligence[11], is becoming increasingly important in both the interaction and analysis phases of the research. The use of Artificial Intelligence in the analysis of social media data can provide valuable insights into user behavior and communication patterns, contributing to a more comprehensive understanding of the phenomena under investigation. Therefore, a rigorous methodological approach that integrates these various aspects is necessary to ensure the accuracy and reliability of the research findings.

There are many publications on the dynamics of deradicalisation, some of them (though very rare, but increasing), have ventured into deradicalisation on social networks [289], and even less on the Telegram platform. Being a new topic, most of these publications, especially in anthropology (given the dynamics of deradicalisation), have limited themselves to giving the basic methodological foundations for good results [290], while also preserving a good ethical compromise (in relation to data and privacy). [291]
For instance, the report *Developing a social media response to radicalization*, [289] demonstrated how, through the use of narratives on platforms such as YouTube, Twitter and Facebook, already radicalised people did not pay attention to the narratives of deradicalisation. But this method, they suggest, *might best be used for prevention purposes for a well-defined target group. Perhaps for those individuals who can be identified to show some curiosity or sympathy for extremist groups, but are not yet active supporters.*
But, in the specific case where there is a possible danger, therefore a threat to society, institutions, organisations and governments immediately shut down the various communication channels (Telegram and Facebook mainly), closing the groups where people discuss and organise.
For example, in Italy, during the pandemic, Telegram groups were often used for no-vax and no-mask discussions.
Some of these channels, however, organised themselves and orchestrated a whole system of bribery and false vaccinations, in order to obtain fake Covid-passes [12]. [292, 293, 294, 295]

[11] specifically GPT-3.5 model

[12] Covid-passes have been used to gain access to public places such as hospitals and administration, but also to private places such as restaurants etc. So if you didn't have a covid-pass, in other words vaccinated, you wouldn't have been able to go to a restaurant.

This strategy (which has also been implemented in other countries [296]), has limited most of the case studies, inherent to the no-vax dynamics on Telegram. In fact, if I put *deradicalisation no-vax Telegram* as keywords on google scholar [13], only six articles are found. Of these some are related to the cases of Capitol hill and Qanon. Whereas if I search only for 'telegram deradicalization' in scientific articles only four pages come up, of which only four articles mention the word Telegram in the text (and none in the title).

But, returning to the Italian case, the closure of groups took place as soon as a possible threat was detected, i.e. that of an increase in Covid-19 cases given the efforts of the various governments to reduce the Covid cases, so a decision was subsequently made to close the Telegram groups. In Italy, after the closure of various no-vax channels due to the fake Covid-passes, there are only very few groups left. I have chosen to study no-vax groups in this thesis, as no-vax organisations are similar to other groups that do damage to public and private bodies. As was, for example, the case with *ISIS/Daesh*. [14]

5.1 Case Study

Although the objectives of the thesis may appear to be psychological operations (PSYOPS), in reality they really are. The term PSYOPS refers to those actions with the main purpose of influencing the opinions, emotions, attitudes and behaviour of hostile groups in such a way as to further national goals, [297] so it is possible to make the deradicalisation process part of the national security interest.
PSYOPS actions can also be part of those actions where the human factor is a fundamental key to the achievement of both objectives and national and international balances, and therefore there are different methodologies for different objectives.
I mainly relied on two documents [298, 299] to observe PSYOPS operations, especially in the internal, not external, sectors application, precisely to improve the methodology of my thesis. These data-driven papers showed how PSYOPS operations can be compared to marketing actions, but with due care and modification, of course.

But before we talk about data and methodologies, we need an important evaluation of the sample studied. Are we sure that the Telegram sample is actually radicalised and/or dangerous?

Figure 9 shows an interesting behaviour, for the non-Italian speakers, here's the translation [15]:

> *Let's fill the Sapienza University (Rome) with shit.*
> *Share it.*

[13](2023/03/31)
[14]Islamic State of Iraq / Islamic State of Iraq and the Levant
[15]Thanks to Deepl - *https://www.deepl.com/translator*

The first undergraduate course on Gender Studies to train assholes who will go out and make propaganda for those feminist scumbags.
Let's not stand by while those feminist cancers are destroying our society.
Enough is enough! We must counter-attack!
Let us not be destroyed by a small minority.

The message goes on to give suggestions on how to do damage at the Sapienza University in Rome, attaching e-mails, telephone contacts, addresses and other public and easy to obtain information.
This message can be defined as incitement to violence, and so, can be defined as a threat due the voluntary of the action.
There were no physical actions dangerous to people's safety, but actions such as intimidation and shitstorm [16] were reported. Days after this event, some Italian newspapers [300, 301] wrote about this event, reporting and confirming the intimidation to public employee. After these events, the "Infosapienza Centre" (of the same university) developed in-house 'blacklists' made up of thousands of malicious IPs, constantly updated in a fully automated manner, resulting from observations, analysis and continuous study of traffic 'anomalies' coming from the Internet, complementing the existing Threat Intelligence systems. [302]

Moreover, Italian no-vax groups (such as the one in the study) have already carried out vandalistic and dangerous acts, such as setting fire to a vaccination hub, fortunately without injuries because it was done at night. [303, 304]

In Italy there are mainly two 'radicalised' groups on Telegram, the most extremist one is called *" !Chat stop dictatorship! "* but has few members (about 100) while the second and more voluminous one is called *"Put down the covid-mask"*, and has about 450 members.
I found these groups by searching both on Telegram, with keywords such as Covid, no-vax, no-mask and others (but found no structured groups), and online, where I mainly went looking through newspaper articles to find out who the various no-vax groups in Italy were. The newspapers gave me an exhaustive list of the largest and most relevant groups (see the Sapienza University case), many of which have been closed due to the problem of the spread of fake or real covid-passes.
The only truly structured and organised ones with daily active members are these two. Most of the information circulating on these two groups is similar, but the group *"!Chat stop dictatorship!"* is more extremist than the other, in the sense that there is no community discussion, but there are often angry and resentful people insulting each other, perhaps to show that the one who shouts and insults the most is stronger and more vigorous than the other.
While in the second group, despite often having similar content, the community manages to 'not insult each other' and actively discuss the topic itself, expressing different ideas and opinions, even

[16] A situation in which a lot of people are disagreeing and arguing with each other online

if radical.
In the first group, it is very difficult to achieve a deradicalising process, as there is no possibility of discussion. However, this is not due to the people themselves, but to the fact that they are in an echos chamber. If people often see insults, at some point over time, it makes that the norm. One can get out of this situation by inserting different information into the Echo chamber to amalgamate the group over time. Therefore, given the objectives and conditions of the social samples obtained, I decided to work mainly on the second group, i.e. *"Put down the covid-mask"* as one of the most interesting, considering the aim of this work.
Despite its relatively small size of 450 members, which pales in comparison to the vast galaxy of no-vax and no-mask communities in Europe, this particular group stands out as the most compelling and vibrant choice. Its engaging and diverse discussions make it a standout among other groups, many of which have diminished in activity due to ongoing closures and restrictions.

Given, however, that data and algorithms cannot confirm whether the group is actually radicalised (it is only possible to compare it with others), are there any methodologies to assess whether that group is actually radicalised from a sociological point of view?

The evaluation sciences within the social sciences present various types of critical thinking to evaluate the desired subject. One of these, the 'realist' one, has as its methodology the Context-Mechanism-Outcome Pattern, developed by Pawson and Tilly. [305, 306] Realist evaluation researchers orient their thinking to context-mechanism-outcome (CMO) pattern configurations. A CMO configuration is a proposition stating what it is about an information systems initiative which works for whom in what circumstances. A refined CMO configuration is the finding of information systems evaluation research.

In Figure 10, is shown the Context mechanism output model. The CMO's works as described: within a Context (where norms and roles are agreed upon), a Stimulus (Mechanism) such as a piece of information is inserted, and finally a response/behaviour (Output) is obtained, which allows me to evaluate and confirm the initial hypothesis.

This method allows us to assess whether the social group actually has radical characteristics. Some studies such as Gustave le bon's 'The Crowd' [307], for example, have already spoken of group dynamics and their characteristics, but the substantial difference lies both in the obvious fact that one thing is a physical group moving in a delineated physical space (therefore also with physical limits, which can modify its behaviour), and also in the fact that the dynamics of information that mobilises the group is completely different. Le Bon's dynamic, for obvious temporal reasons, does not include the internet and instant messaging platforms, something that today's crowd groups have and do. The ability to send information at the right time, and of a certain weight, can become the key that initiates a phase transition.

Figure 9: Radical and dangerous groups

These dynamics, however, can be easily simulated by computers thanks to new technologies. So, in principle I could use an ad-hoc simulation as a methodology for my thesis and model it until I have a certain similarity with some already existing social events. Starting from this assumption, however, I make an important assumption from a methodological data point of view.
Why download data when you can do simulations? Lymperopoulos et al (2015) [24] described that in social systems the dynamics might be defined at an abstract level, thus making the traditional engineering approaches not always pertinent to the validation of social simulations. So The question remains: "How can it be determined whether the output of a social simulation corresponds to the output of the investigated social process?" The answer to this question is closely related to the purpose served by a social model. In their case, the purpose is to provide qualitative evidential proof of the proposed isomorphism between the dynamics of networks of Integrate-and-Fire neurons and the dynamics of interacting online social network users. In this case the simulation results are considered valid if some of the empirical observations are reproduced by the model. This outcome can be interpreted as evidence of the credibility of the fundamental qualitative assumptions

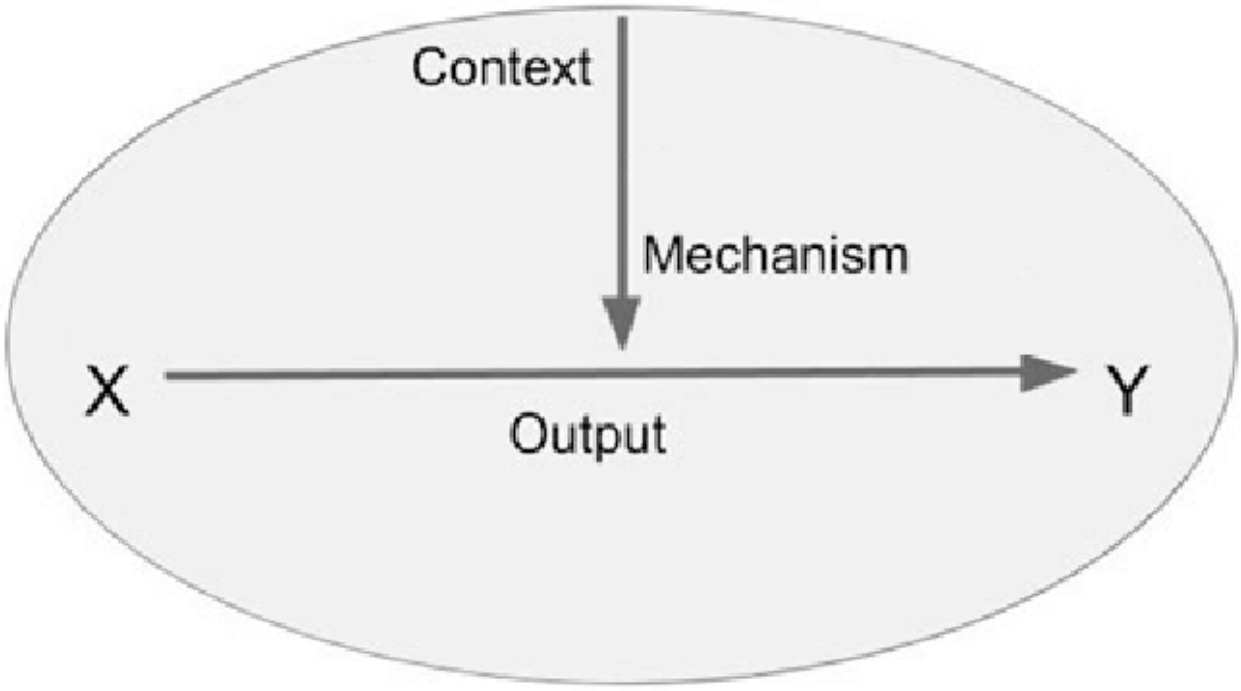

Figure 10: Context-mechanism-outcome - Pawson and Tilly

accounting for the underlying causal mechanisms leading to the phenomenology of the analysed social system. Fulfilling this purpose can take us a step further in the development of a more accurate online social contagion model based on neural network dynamics. [24]
This can show how: by going closer to what the scientific community defines as 'deradicalisation', it is possible to approach the result of one's simulation if it is close to what the scientific community refers to as 'deradicalisation'. Starting from this assumption, I can make a social simulation, using software such as Netlogo for example, and have the possibility of obtaining data and field-testing a number of tools, in order to simulate a deradicalisation process in reality, without the limitations of a social simulation done by computer. The limitations of a simulation are various, such as the assumptions one makes when talking about individuals and the dynamics of the simulation. Things that are obviously not there in a real test. This description, therefore, imposes important work on the data collection, the cleaning and processing part (as described in the introduction section on the '5 Vs' of bigdata).

5.2 Data gathering & collection

Online social networks have evolved into valuable sources of information and pervasive communication platforms where people, businesses, and organizations generate and share content, build relationships, and join public conversations. In this online ecosystem, social networks are where information propagation is affected by external sources of influence, such as mass media, socioe-

conomic circumstances, advertising, or events, giving rise to collective intelligence or collective behaviour patterns.

The increasing use of the online social networks and the ample availability of empirical data are instrumental in elucidating the mechanisms of behavioral phenomena appearing in the online ecosystem. So collecting, but above all knowing how to collect, data is crucial for the methodological part. I have collected data from the Italian chat *"Put down the covid-mask"* on Telegram, using the Telegram API, [17] The Telegram API enables programmers and researchers access to Telegram elements like Text, Direct Messages, Lists, Users, bots in chats and groups. The difference between chats and groups is that in chats people can share external content (like link, articles and various material) and text, and in the group, only the users selected by the Administrator of the group can send text and external content. The group is more focused on informing and sharing information for the users who follow it, and the chats are focused on sharing ideas and discussion about some news or external content. I have collected Telegram data by using Telethon[18] and the Telegram API. Both services use permission from Telegram to obtain and gather data, but any downloaded data needs revisions and a cleaning process to increase the quality of the research. For example, spam text. For any topic, I use the same methodology to obtain standard and quality data. This is the request also from the entropy calculation because it is essential to know every evolution of the topic at any given moment.

The data were organised in the following way:

Table 5: Keys and Values of data collected

Keys	Value
Id	shows the unique key-number of the text written
From	shows the unique key-number of the user wrote the text
Text	shows the text written
Date	shows the time when the text was written
Reaction	shows any emoticons expressed by other users for text written
Reply	shows the unique key-number of the user who replies to the text written

Table 6: Example of Keys and Values of data collected

Id	From	Text	Date	Reaction	Reply
26525	user_id=910513947	Sample text	2022-10-01 00:06:23	:)	-
26526	user_id=2143679216	Sample emote	2022-10-01 00:31:51	:/	26521
26528	user_id=2143679216	Sample text + emote	2022-10-01 00:38:29	:(	26525

Starting from this data, it is possible to obtain important information on the social dynamic, such as 1) identifying the people who wrote the post; 2) the exact moment of the text, which gives information on the dynamic itself; 3) the text, thus the sentiment, entropy and other information that

[17] Telegram is not fully encrypted.
[18] https://docs.telethon.dev/en/stable/

can be obtained from a text; 4) the emoticons, which give additional information on the impact of the text; 5) the replies, which are necessary to create a network and to create a network dynamic over time.

The combination of these data taken over a given period yields the initial conditions of the system. The problem, however, in obtaining the initial conditions of a social system lies in the fact that there is no zero point from where everything starts, but there are moments when activity becomes more intense, however that does not correspond to the moments of greatest activity, but to the moment when a group becomes self-organised and realises that it has become a social group. Obtaining the initial conditions of group *"Put down the covid-mask"*, i.e. starting from its creation (2020), does not linearly lead to the conclusion that this group starting from that moment has become a group, therefore with the same initial objectives, because external events (such as the fluctuations of the pandemic for example) cause the activity, topics, discussions and objectives to be totally different from the initial conditions for which the group was formed.

According to this assumption, data collection must be done at a point where a topic (new or not) becomes relevant for the group, thus changing previous objectives. We can define this phase as *post-layout*. By post-layout I mean that moment when the paradigm of the group is changed, passing from one setup to another. Data collection, to obtain the initial conditions of the system, therefore requires a careful analysis of the group dynamics, as well as the objectives. I decided from this approach to obtain the data from **2022/11/10**, i.e. before the start of the vaccination issues and the hypothetical break in the conflict between Ukraine and Russia. Before this date, the topics were very fragmented, either because there had recently been national elections in Italy, or because they were waiting for updates regarding vaccination and the war as the cold season was coming.

To confirm my hypothesis, by collecting data over time, I have noticed precisely an increase in group interactions since the beginning of October, with a weekly increase of $\sim 20\%$ compared to past weeks. Reaching a total of about $\sim$1000 posts/texts per week.
This data can also be used to assess the impact of the presence of the bots and their actions on the group.

5.3 Initialization of initial condition

From the data obtained, it is possible to show what the initial conditions of the system are, day by day (since I have the temporal data). Thanks to the "ID" and "Reply" data, it is possible to obtain the Topic dynamics, i.e. to obtain an infographic map that allows us to see the behaviour of the Telegram group according to the various IDs (i.e. messages).

The same procedure can be used to understand the group network, thanks to the 'ID', 'From' and 'Reply' data. With the ID data, it is possible to link the account that made the Reply, starting with the From. Each time there is a new text (ID) and a Reply, it is possible to get the From from the ID. So by doing this you get a From to From, ergo the links from account to account.

With the 'Text', it is possible to count the most frequently used words, and those most related to each other. On this last part, many vowels will be present and will take the place of words on a numerical level, precisely because of their function on a linguistic level. This is why I had to add the 'stop-words'. i.e. words that are not to be counted during the analysis.

These data then, if well combined, can be infused to help understand those interactions in the group, under the initial conditions, that will help me devise a deradicalisation strategy. To combine the data and infographics I chose the Gephi tool.

Gephi is an open source software for analysing and visualising social networks. It has been used in various research projects, journalism and in various other fields, e.g. examining the network traffic on Twitter during social dynamics [308], but also more variously for other topics usually covered by network analysis, like the Word similarity Figure 11.
The usefulness of Gephi is not only for the infographic part, but also for the statistical part of the network, for community detection, edge weights and for the temporal dynamics of the network itself.

As for the part of the "Text" that does not include graphs, I will use various tools to understand sentiment (Vader in Italian programmed by me) [19] that improve the sentiment analysis of the data collected since the original version of VADER first translates foreign languages into English (losing some of the meaning) and then does the sentiment analysis, the Gunning-Fog-Index, to understand the simplicity or complexity of the text, and finally Shannon's entropy to obtain the entropic level of the text.

In relation to the Gunning Fog Index, i.e. to understand what type of language (simple or complex) is used by users, I used the Lo Bosco's model [309] in my thesis, because it is the best that performing in Italian. Text simplification (TS) is a process that aims at reducing the linguistic complexity of a text by modifying its syntactic structure and substituting lemmas. The result of TS is a new text that keeps the original meaning, but which is more easily readable and understandable. TS has proven useful for different classes of people, for example those who have language disabilities, are not mother tongue or have a low educational level. Children affected by deafness need help for facing reading difficulties or people affected by dyslexia have to face comprehension difficulties in reading infrequent and long words. Another aspect that has to be taken into account is the high percentage of people with low literacy skills that are unable to understand common texts. For example, Italy is

[19] https://github.com/AndreaRussoAgid/VADER-Italian-Sentiment

one of the countries with a considerable number of people with low linguistic competencies [310]. In addition, research work has shown that on Wikipedia [311] scientific topics have a higher textual complexity than sports topics. I use this tool to understand whether the language of radicalised people is simple or complex, and thus outline a communication strategy, together with the other tools.

Always pertaining to text, there are various libraries on Github[20] concerning Shannon's entropy, both inherent to pure text and inherent to the probability of information, but there is a method for the related topic and another one for the related text.

5.3.1 Information entropy and Shannon entropy

Topic-Entropy related

In information theory, the *information value* contained in a message is directly related to how "surprising" or "unexpected" the message is for the reader [312, 313].

Suppose we have a biased coin with the probability p of landing on heads and probability $1 - p$ of landing on tails. For what value of p do we have the maximum "surprise" or "uncertainty" on the outcome of a coin toss? If $p = 1$, the outcome of a coin toss is expected to be always heads, so there is no surprise or uncertainty. Similarly for $p = 0$, when we always expect the coin to land on tails. If instead $p = 0.5$, then we have the maximum surprise or uncertainty.

After discussion with John Von Neumann, Shannon decided to use the term "entropy" in place of the word "uncertainty." Claude Shannon mathematically formalised this value of "surprise" or "uncertainty" in 1948 as part of his communication theory.

Formally, the entropy of our biased coin is given by:

$$H(coin) = -(p * log(p) + (1 - p) * log(1 - p)) \tag{2}$$

where the base of the logarithm can be chosen arbitrarily. If the base is 2, the entropy is measured in *bits*. If instead the base is e, the entropy is measured in *nats*. Finally, if the base is 10, the entropy is measured in *dits*.

When $p = 0.5$, $H(coin)$ is maximal, it is equal to 1 bit. When instead $p = 0$ or $p = 1$, $H(coin)$ is minimal, it is equal to 0 bits.

The concept of entropy can be generalised from the simplest case of a coin to the more complex case of a discrete probability distribution. A discrete probability distribution over n possible outcomes $x_1, ..., x_n$ is given by n probability values $p(x_1), ..., p(x_n)$, where $0 <= p(x_i) <= 1$, and the sum of all $p(x_i)$ is equal to 1. Note that a coin is a probability distribution over two possible outcomes.

Formally, the entropy of a discrete probability distribution is defined by:

[20] https://github.com/search?l=Python&q=shannon+entropy&type=Repositories

$$H(P) = -(p_1 * log(p_1) + ... + p_n * log(p_n)) \tag{3}$$

where, as before, we can chose arbitrarily the base of the logarithm.

Text-Entropy related

In information theory, entropy is a measure of the uncertainty in a random variable. In this context, the term usually refers to the Shannon entropy, which quantifies the expected value of the message's information. Claude E. Shannon introduced the formula for entropy in his 1948 paper "A Mathematical Theory of Communication." [312, 313].

$$H(X) = -\sum_{i=1}^{n} p(x_i)\ log_b\ P(x_i) \tag{4}$$

Minus is used, because for values less than 1, and logarithm is negative. However, since

$$-log\ a = log\frac{1}{a} \tag{5}$$

Formula can be expressed as

$$H(X) = -\sum_{i=1}^{n} p(x_i)\ log_b\frac{1}{P(x_i)} \tag{6}$$

Expression

$$log_b\ \frac{1}{P(x_i)} \tag{7}$$

is also called an uncertainty or surprise, the lower the probability

$$P(x_i), i.e. P(x_i) \to \infty \tag{8}$$

the higher the uncertainty or the potential surprise, i.e.

$$U_i \to \infty \tag{9}$$

for the outcome

$$X_i \tag{10}$$

In this case, the formula expresses the mathematical expectation of uncertainty, which is why information entropy and information uncertainty can be used interchangeably.

For this calculation, I have found an online provider[21] where it is possible to perform the entropic calculation of the text without any limitation, using the same equations.

[21] https://planetcalc.com/2476/

Tools	Goals
Telethon API / Scraper	Collect data from Telegram
Pandas	Cleaning data
Shannon Entropy	Evaluate High or Low entropic topics/texts
Gunning Fog Index	Simple or Complex text
Info structure	Analyse data sample
Similarity	Topic and words network
Gephi	Network analysis

Table 7: Tools & Goals

For example, the entropy value of the name of my thesis, is 4.1154170, while the name of my PhD program (PhD in Complex Systems for Physical, Socio-economic and Life Sciences) is 3.9835059.

All these methodologies will eventually give me a risk assessment document of a radicalised group, which will allows me not only to base deradicalisation strategies, but also to document the state and status of the group, to the appropriate authorities. Organisms such as CNAIPIC (National Computer Crime Centre for the Protection of Critical Infrastructures) i.e. the Italian agency in charge of preventing and suppressing computer crimes, whether of common, organised or terrorist nature, targeting computerised infrastructures of a critical nature and of national importance, may be interested in the 'risk assessment document', i.e. the initial conditions of the system, showing the dynamic evolution.

The tools that will be used will each be an evaluation tool for the part under interest. Specifically, I would like to use, for example, Shannon's entropy and the Gunning fog index to assess both the level of information demand (Entropy) and the type of language used (GFI). I have described the tools and their goal in detail in Table 7.

5.4 Natural Language Processing

NLP has moved out of the research laboratories and is gradually being implemented in computer applications that require the integration of human language with machine. Thus, the NLP methodologies that have now emerged from the various laboratories and universities have taken on multiple differentiation for the purposes of the various programmers.

The most common ones, but especially those that will be used in my thesis, will be briefly described to help readers with no experience in this regard, to not only be able to distinguish one from the other, but also to understand why a certain methodology is used over another. I begin with the one I started my PhD with, and in which I excelled the most, I am talking about **Sentiment Analysis** (SA).

Sentiment analysis (also known as opinion mining or emotion AI) is the use of natural language processing, text analysis and computational linguistics to systematically identify, extract, quantify, and study affective states and subjective information. [314] Sentiment analysis is widely applied to

materials such as reviews and survey responses, online and social media for applications that range from marketing to customer and political service. [315]

Existing approaches to sentiment analysis can be grouped into 4 main categories:

- keyword spotting;
- lexical affinity;
- statistical methods;
- conceptual level techniques.

The first classifies text by influential categories based on the presence of unambiguous influential words such as happy, sad, scared and bored. Lexical affinity not only detects influential words, but also arbitrarily assigns words a probable affinity to particular emotions. Statistical methods, on the other hand, rely on elements from machine learning such as latent semantic analysis, support vector machines, groups of words and semantic orientation. To extrapolate the opinion in a context and obtain features, grammatical word relations are used. The relations are obtained from a deep syntactic analysis of the text. In contrast to purely syntactic techniques, conceptual-level approaches leverage knowledge representation elements such as ontologies and semantic networks, and thus are capable of detecting semantics that are expressed in a subtle manner.

The combination of these parameters means that after the analysis of the text (distributed in positive-neutral-negative), the various sentiment tools give a decimal numerical result (compounds) between -1 and 1, which corresponds to a tendentially negative sentiment if the numerical value is below 0, positive if it is close to 1, or if the result is close to 0 it is neutral, whereas if the result is precisely 0, the tools failed to understand the sentence itself.

An example of a result from sentiment analysis is that expressed in Table 8, obtained on 2022/10/25 [22].

Table 8: Example of Vader Sentiment on various Topics and #Hashtag from Twitter

Topic or #Hashtag	Negative	Neutral	Positive	Compound
Presidente del Consiglio	0,0573	0,8071	0,1356	0,2504
#GeneralElectionN0W	0,1630	0,7291	0,1233	-0,0655
#DoctorWho	0,0712	0,6104	0,1694	0,2972
Rishi Sunak	0,0805	0,6903	0,1097	0,0874
Boris	0,0982	0,7252	0,1307	0,0742

VADER is among the most widely used tools for sentiment analysis, created by MIT. At the beginning of my PhD, I searched for the existence of a VADER tool in Italian, but I found no

[22] On 25 October 2022, many things happened politically, the most significant being the establishment of the new right-wing government in Italy, represented by the first female Prime Minister *Giorgia Meloni*, and in the United Kingdom, *Rishi Sunak* become the first British Prime Minister of Indian descent.

repository on the various sites, including GitHub. At this point, however, given my nationality and the fact that some of my projects concern Italian politics, thus a state that can balance within the international dynamics and perimeter, I decided to programme a VADER in Italian by myself. I succeeded, and during my doctoral thesis defence I will publish the repository on GitHub. [23]

Another tool you will find in this thesis is the **Gunning Fog Index** (GFI), designed to measure the ease of reading and comprehension of a text. The resulting number is an indicator of the number of years of formal education a person needs in order to read the text with ease. That is, if a given text has a Fog index of 12, twelve years of formal education, equivalent to a high school diploma, is required to understand it. The index was developed in 1952 by Robert Gunning, a US businessman.

Gunning's fog index is generally used for texts that are not particularly long and for writing in a professional context. Texts that are designed for a wide audience generally require a fog index of less than 12 points. Texts that require almost universal comprehension generally require an index of less than 8 points.

Research in the various branches of linguistics, the psychology of perception, neurology, didactics and artificial intelligence have confirmed the complexity of the communication phenomenon. However, in recent years, research in the fields of mathematics, computer science, artificial intelligence and linguistics has led to an improvement of automata in processing linguistic productions, and this is also due to the existence of tools such as the Gunning Fog index and other more recent ones. A virtuous circle seems to have been established between computational methods of linguistic description, artificial intelligence and the practice of writing, whereby one benefits the other and vice versa.

In my thesis, the FOG index will be used to understand the complexity/simplicity of the text being shared in the various channels. I hypothesise that the difference between simplicity and/or complexity in textual language can lead to a different reaction of information between sender and receiver.

Another type of text analysis is **Word-embedding** (WE), it makes it possible to store both semantic and syntactic word information from an unannotated corpus and constructs a vector space in which word vectors are closer together if the words occur in the same linguistic contexts, i.e. if they are recognised as semantically more similar (according to the hypothesis of distributional semantics). In a stricter definition, word embedding is an overall term for a set of modelling techniques in natural language processing in which words or phrases of a vocabulary are mapped into vectors of real numbers. Conceptually, it consists of a mathematical operation of immersion as a result of which a space consisting of one dimension per word is transformed into a continuous vector space of much lower dimension. These techniques find application in the study of the semantic proximity

[23] https://github.com/AndreaRussoAgid/VADER-Italian-Sentiment

of speech, particularly in the world of distributional semantics. It's possible to find an example (such as the Sentiment one) in Table 9.

Table 9: Example of Word-embedding between words "Ukraine" and "Russia" during the first day of invasion

Words	Score
accent	0,22711
youcrane	0,22463
johnsonoutday32	0,22107
surround	0,22022
inshallah	0,21736
kuweit	0,21480
trivia	0,21462
18+	0,21451
peopleslifematters	0,21411
bidensucks	0,21342

To train the word-embedding model on a given topic, I used Word2vec, a simple two-layer artificial neural network designed to process natural language. The algorithm takes a corpus as input and returns a set of vectors representing the semantic distribution of words in the text. For each word in the corpus, a vector is uniquely constructed to represent it as a point in the multidimensional space created. In this space, words will be closer together if they are recognised as semantically more similar.

One type of text analysis that you will often find in my thesis is **Similarity** (SI). Similarity was calculated with a code I wrote with python library, it consists of counting the words in every single sentence of a text, recognising them, and then showing the dataset at the end of reading, which is a set of words connected with others. The difference with word-embedding is that WE calculates the semantic distance between words, whereas similarity simply counts the number of times one word is close to, or more present than another. An example, like the Sentiment one and the word-embedding one, is shown in Figure 11.

You will also find a special feature in Figure 11, namely that the words are also linked together by a series of distinct and well-separated colours. This distinction is another feature of my work, which you will also find in this thesis, namely **Community Detection** (CD).

In the study of complex networks, a network is said to have community structure if the nodes of the network can be easily grouped into (potentially overlapping) sets of nodes such that each set of nodes is densely connected internally. The particular case of non-overlapping communities implies that the network naturally divides into groups of nodes with dense connections internally and sparser connections between groups. But overlapping communities are also allowed. The more general definition is based on the principle that pairs of nodes are more likely to be connected if they are both members of the same community(ies), and less likely to be connected if they do not share communities. A related but different problem is community search, where the goal is to find a community that a certain vertex belongs to. There are many methods for community detection,

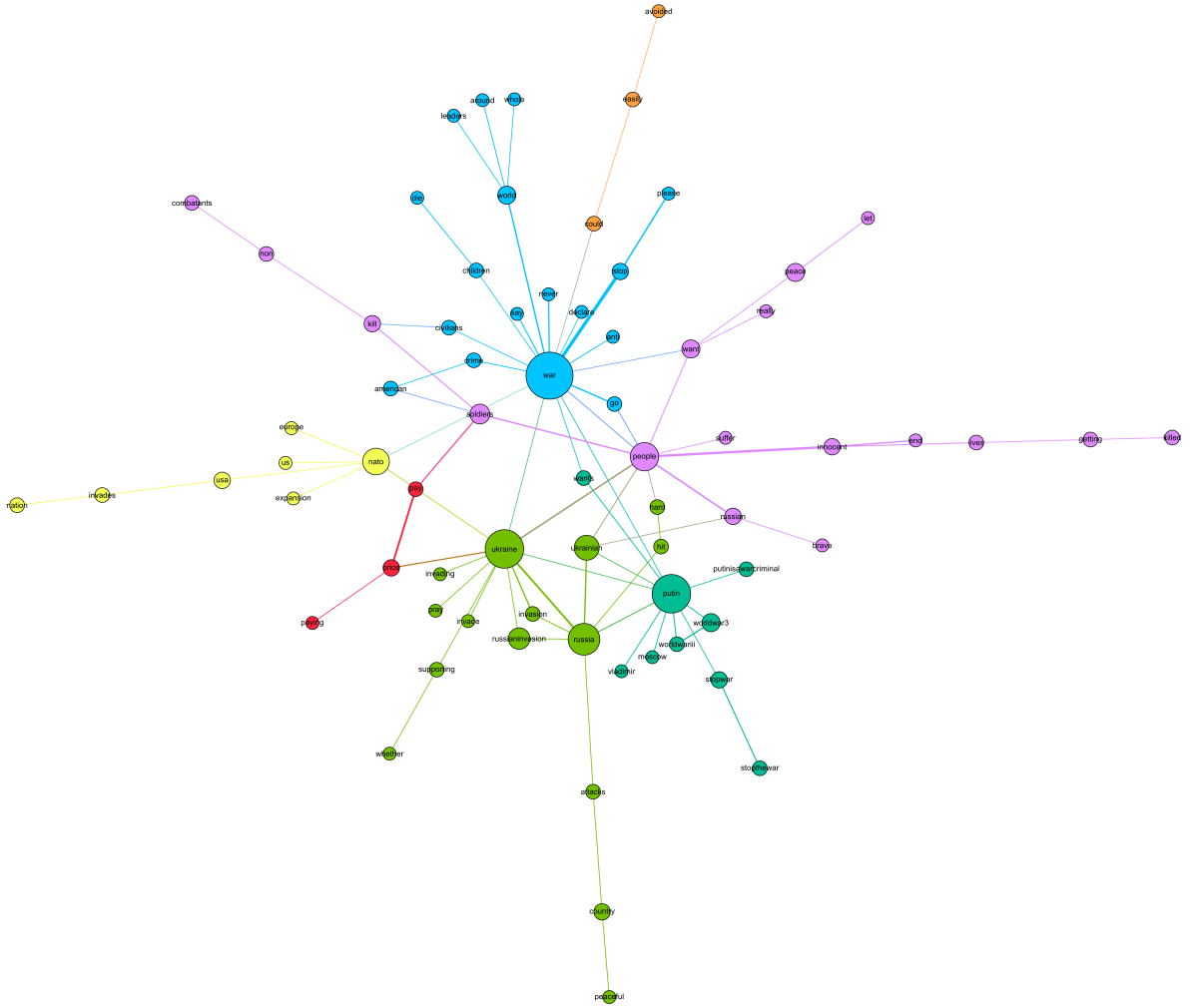

Figure 11: Word similarity between Pro war vs Against war

but one of the most widely used methods (which I also use) for community detection is modularity maximization. Modularity is a benefit function that measures the quality of a particular division of a network into communities. The modularity maximization method detects communities by searching over possible divisions of a network for one or more that have particularly high modularity. Since exhaustive search over all possible divisions is usually intractable, practical algorithms are based on approximate optimization methods such as greedy algorithms, simulated annealing, or spectral optimization, with different approaches offering different balances between speed and accuracy.

My latest NLP method for text comprehension and dynamics is the **Shannon entropy** (SE). In information theory, the entropy of a message source is the average information contained in each emitted message. The more information contained in a message, the less likely it was. An obvious message that has a high probability of being emitted by the source contains little information, while

an unexpected, unlikely message contains a large amount of information. The entropy of a source answers questions such as how predictable can the messages emitted by a source be?

As with the other tools, I have already show an example of entropic analysis presented in the chapter *Relevant Papers with my computational tools* the past section. There you will find two of my publications; in the first I highlight how it is possible to hypothesise the prediction of a demand for information by newspaper readers, using entropic topic analysis; while in the second, using sentiment analysis, I try to understand the cybernetic fragility of a company in a specific time frame.

5.5 Why use ChatGPT as an AI?

ChatGPT is an advanced language model developed by OpenAI, based on the GPT-3 architecture. The GPT-3 version is the basis of the of the "famous" ChatGPT interface (from now on, I will refer to ChatGPT as GPT-3.5 model).
As one of the most powerful language models to date, ChatGPT has the ability to understand and generate human-like responses to various prompts, ranging from simple queries to complex conversations. Its sophisticated neural network is trained on a massive corpus of text data, allowing it to comprehend and generate text with remarkable fluency and accuracy.

ChatGPT is suggested for bot-human interaction because it is a powerful language model that has been trained on a massive amount of text data, allowing it to generate responses that are both contextually relevant and linguistically fluent. This makes it an ideal candidate for conversational agents or chatbots, as it can understand and respond to a wide range of user inputs and generate human-like responses that are tailored to the specific context of the conversation.

Furthermore, ChatGPT has the ability to generate responses that are consistent with the tone and style of the conversation, which can be crucial for creating a natural and engaging user experience. This is particularly important for applications such as customer service bots, where the ability to provide personalised and empathetic responses can have a significant impact on user satisfaction [24].

Overall, the use of ChatGPT for bot-human interaction has the potential to greatly improve the quality and effectiveness of conversational agents, leading to more natural and engaging interactions that better meet the needs and expectations of users. One issue, however, concerns the vocabulary of ChatGPT. In the DAVINCI-3 version (I use the 3.5 version) only 7% of the database used to improve the model is different from the English language [25] [316, 317], however, this does not detract from the fact that our perception of natural language text is influenced by our linguistic competence and our tendency to interpret communicative acts as conveying coherent meaning, even if they don't. This poses a problem when one side of the communication lacks meaning, as our comprehension

[24] This ChatGPT introduction page was created by ChatGPT itself, starting with the question: *Explain to me why the use of ChatGPT is suggested for bot-human interaction*

[25] it does not specify whether American English or UK English

of implicit meaning may be an illusion. Despite its output, a language model stitches together sequences of linguistic forms based on probabilistic information without any reference to meaning. It can be thought of as a *stochastic parrot*. [316]

ChatGPT has, however, already been used in human interaction fields such as pedagogy, education [318], with goods results, creating for example personalised learning experiences for their students, saving teachers time and effort. [318] This pedagogical model, can be augmented for a better use of the methodological process of deradicalisation, i.e., by exploiting what is one of the best features of ChatGPT, namely that of saving information from past interactions, it is possible to create an ad hoc deradicalisation process for each person, independently. On a practical methodological level, however, I did not have the opportunity to apply this method in my thesis, as it requires care and a special Telegram-ChatGPT script [319] for handling the interactions between ChatGPT and individuals. Furthermore, as will be seen in the results chapter, interactions between radical individuals will occur in a spurious manner, namely, few interactions per person, but with many people, with the fate of creating many chats (which take up memory on the server) that will be used for little. This, however, pertains to DAVINCI-3, whereas for the newer version 3.5, things may change considerably, especially in view of the positive and amazing recent results of version 4. [320]

There are, however, of course other AIs available online from other platforms, which may be better than ChatGPT for my purposes, such as: Microsoft Bing, Chatsonic, Jasper Chat, Google Bard AI, Character AI, Colossal Chat, YouChat and DialoGPT, but after some testing, both for ease of use and also for natural-sounding and coherent responses, I opted for ChatGPT.

But of course, ChatGPT cannot be on Perrow's AI level [84], as ChatGPT cannot do this alone, it requires an additional observer who can partly observe the social dynamics well, but with the help of an AI (in this case ChatGPT) can avoid what Perrow called his *Normal Accidents' theory*. ChatGPT serves in this case, both as communication, but also as planning (to avoid Perrows' errors), as it can store past conversations, as it has the ability to notify when something might be a threat (in this case social), sending attention errors to those requesting information.

The use of ChatGPT, or even much more generic AI, draws attention to two events.

1. The first concerns the possibility of being recognised as false positives by AI recognition algorithms [26], precisely on Telegram, [321] where even a small false-positive rate might lead to a high number of human accounts wrongly classified as bots.

2. The second, however, which is opposite to the first, is that if AI measures are not used to limit the content of potentially harmful information, there is a risk of serious social

[26] Botometer for example

overload [322] which is what is happening now on Telegram given the massive presence of propaganda.

In fact, the use of an AI-Bot that can mitigate possible radical actions is the goal of my thesis, which shows application potential that is already feasible in the present, given the extreme difficulty of controlling the dissemination of information in certain areas of the world or specific social sectors. For these reasons, I hypothesise that an AI-Influencer is needed to mitigate those social problems, increased by the effect of false information and echos chambers, that can lead to threats to society and institutions.

But given the success it has had in the media, some privacy issues and lack of transparency have been encountered. [323] For these reasons, ChatGPT is currently only being closed in Italy (2023/04/04). [324] However, this did not limit the use of ChatGPT for my thesis as I am currently in France and because it is possible to circumvent this limitation with a simple VPN (Virtual private network).

However, ChatGPT also has potentially interesting linguistic and communicative functions, as shown in Version 4 of ChatGPT, [325] which can help to improve the communication and social dynamics that I intention to initiate.

5.5.1 Lymporious's dynamics

Based on the aforementioned work by Lymperopoulos et al. [24], the methodology I would like to apply as social dynamics is that of neuron-based activation threshold. Simplistically, when a leaky Integrate-and-Fire neuron is excited above a certain threshold, an electric pulse – the action potential – is emitted and delivered to the dendritic trees of the neighboring neurons through synaptic connections. This pulse can either cause excitation, thus bringing the receiving neurons closer to their firing threshold, or inhibition, with the opposite effect. Apart from the received pulses, the excitation of a leaky Integrate-and-Fire neuron is also affected by its self-dynamics and its sensory input conveying stimulation from the external environment.

In their approach, the activation threshold is defined as the excitation level above which an individual becomes active. Although the proposed model falls into the general category of threshold models with memory, the threshold definition is broader than that of previous studies. Thus far, the activation threshold in social contagion models with memory refers to the number of contacts with already activated individuals that is required for one's activation. Expanding this static definition they consider the activation of an online social network user a dynamical process, whereby the progressive accumulation of positive influence deriving from his/her self-dynamics, the online interactions, and the external environment, leads to activation when an excitation threshold is exceeded. According to this approach, an online social network is studied as open system, whose actors are influenced by internal and external factors. External sources of influence might be the mass and the print media, public opinion, offline contacts, socio-economic conditions, the weather,

and other factors as discussed in the modeling framework. Empirical evidence that the online behavior is affected by the external environment exists in the relevant literature.

The Lymperopoulos et al. [24] result, shows that the most important determinants are the variations in their activation thresholds, the time scales of their behavioral changes, the amount of time they are online, and the excitatory stimulus they receive. Also, the model implies that the frequency of the oscillations of the global activity depends on the users' refractory period and the excitation intensity. Shorter refractory periods and stronger stimulus result in higher oscillation frequencies. The simulation results and the analysis of the dataset show that excitatory external stimuli, in conjunction with the users' excitatory messages, give rise to synchronised activity, which is a form of collective behavior.

Concerning the methodology, they analyze the level of synchronization per day by breaking up the sequence of messages into 24-h periods, starting from the timestamp of the first message from the 15 M dataset. In doing so, they divide the users' activity into 32 clusters corresponding to the number of days of the observation window. For each cluster and for each user they measured the inter-tweet intervals and then they calculated the Coefficient of Variation (CV). By averaging the users' CVs they obtained the daily global CV of Twitter activity. To derive the aggregate posting frequency, they averaged the users' mean inter-tweet intervals. Lower global CV values indicate higher synchronization. Lower global mean inter-tweet interval values imply higher posting frequency. From the beginning of the observation window to May 15th they notice a highly fluctuating global CV. On May 1st and on May 15th the global CV was minimised. Also, on May 15th the global mean inter-tweet interval reached its minimum, thus denoting the maximization of posting frequency. The fluctuating daily global CV before May 15th is indicative of the alternating level of the users' synchronization. After May 15th the global CV increased, but the fluctuations were smaller.

The coefficient of variation as a measure of data variability depends on the volume and the temporal length of the processed data. Long time intervals, comprising different activity patterns, are not conducive to the precise detection of high synchronization periods. However, shorter time intervals entail a smaller number of active users and fewer messages, thus not forming representative samples to examine synchronization phenomena at a population level. In view of these considerations and in order to further analyze the synchronization trends, they calculated the global CV per 4-h periods. By dividing the day into intervals of relatively homogeneous online activity, they avoid mixing patterns generated during different periods (e.g. night, work, leisure), thus more consistently identifying synchronization phenomena. The results of this calculation observe the converging trend of the global CV, which almost stabilised from May 19th to May 23th. During that period the positive influence generated by the online interaction, the media coverage of the events and the popular support for the demonstrations, was maximised.

However, it is important to observe that the maximum global CV values followed a decreasing trend. This trend shows that thousands of heterogeneous users converged to a homogeneous posting behavior when the internal and external positive influence was maximised. The decreasing fluctuations of the synchronization level imply that the internally and externally generated stimulus progressively became continuous. After May 23th, the fluctuations of the synchronization level increased, thus indicating that the decreasing coverage of the events by the mass media and the diminishing momentum of the protests rendered the stimulus intermittent. The fact that the synchronization shows a stabilizing trend from May 15 to May 23th, reveals a shift towards a more harmonised activity, suggestive of the emergence of collective behavior. The increased media coverage of the events, the growing popular support, and the intensified online activity during that period, indicate that the externally and internally generated positive influence gave rise to a self-sustained synchronization, despite the large number of active users and the increased behavioral heterogeneity.

The similarities between neurons and accounts on social networks has also been taken up by other research. [326] In fact, both neurons and users are interconnected entities whose levels of excitation are regulated by certain stimuli received from dynamics inside and outside the network; the process of sharing content on social is equivalent to the emission of an electrical signal when the neuron exceeds its threshold potential. When a network is excited, the probability of a considerable number of agents being very close to their threshold increases. Subsequently, the activated agents will stimulate their nearest neighbours, thus generating cascades of quasi-synchronous activity. A global quasi-periodic behaviour emerges, where groups of agents are activated in each synchronised wave. Just as the neuron that exceeds a certain activation threshold discharges the impulse to its neighbours, the social user who exceeds a precise threshold value shares a news item. Despite substantial differences at the microscopic level, humans and neurons macroscopically have similarities that allow the approximation of social dynamic processes in a manner analogous to neuron dynamics. The dynamics of the contagion of an opinion, as described by the Voter Model, can be found in an almost analogous way in the process of creating the price of an asset: a financial trader is influenced as much by his own thoughts as by the opinion of neighbouring traders, and even in this case the moment the trader exceeds a certain pre-determined threshold value he becomes active and shares his knowledge with his neighbours. If after a perturbation, be it internal or external, neurons enter a synchronised state, neuronal avalanches are formed; these discharges risk extinguish themselves if the network is not sufficiently connected or if the elements are not sensitive enough.

Also, the 2017 work of Lymperopoulos et al. [128] holds significant methodological relevance to my thesis. They have analyzed online activity patterns with a view to determine the transfer function of social systems, that correspond to the dynamic relationship between external influences and the resulting activity. To this end, first they estimate the impulse response (Green's function)

of collective activity, and then they show that the convolution of the impulse response, with a time-varying external influence field, accurately reproduces empirical activity patterns. Their model is used for the prediction of collective social activity through a training process, whereby the values of the parameters of the model are estimated from existing data, and then used for predicting the future collective social activity. A critical aspect in forecasting social activity is the quantification of the external influence, which is a time varying parameter, thus making the system behave in an non-autonomous way. It is possible to quantify the external influence by associating the strength of sentiment, novelty, importance and context of the incoming information pulses with the value of one of the parameters of the model. In this regard, a training phase is required with a view to analyze the incoming information along the aforementioned dimensions, and then grade its strength using a predefined scale. A stimulation signal of significant signal-to-noise ratio drives social activity out of an equilibrium state, as individuals function as encoders of the temporal structure and strength of the external simulation. It is demonstrated that the model can infer a unique external signal profile pertaining to the exogenous influence affecting a collective activity pattern. Since this signal profile gives rise to a simulated activity pattern accurately matching the real one, their approach provides reliable evidence of correctly estimating the temporal structure and strength of the external influence generated by exogenous information streams entering a social system.

The analysis of patterns of online activity identified the Green's function, which describes the dynamic relationship between infinitely short impulses of influence and the resulting online activity. They showed that the sum of the resting activity of a social system and the convolution of the Green's function, with stochastic and time varying deterministic influences, yields the value of the collective social activity. By establishing a dynamic relationship between the incoming influence and the collective social activity, the Green's transfer function of social systems in the content of online activity. The Green's transfer function is fundamental to our understanding of the process dynamics of social systems, as well as to the modeling and prediction of human activity patterns. Provided that the intensity of the external stimulus is known, a transfer function model can predict the population response in diverse contexts, such as online communication, consumption of products and services, and financial markets among others. Regarding the method used, they have initially examined the dynamics of collective online activity in a state of statistical equilibrium. In this case, the fluctuations of the collective online activity around a resting value were modeled as the outcome of the individuals' response to white noise input representing normally distributed random shocks perturbing the online activity. Then they switched to a time-varying influence field corresponding to perturbations generated by information flows pertaining to various exogenous factors (e.g. breaking news, socio-economic developments). Through the analysis of online collective activity patterns, they identified the Green's function describing the dynamic relationship between infinitely short influence pulses and the resulting online activity.

I will try to partially simulate this whole methodological part during the interaction with the Telegram group, in order to improve the chances of being heard, gaining interest and being able to amplify the information I would like to get across at a given moment to as large an audience as possible.

5.5.2 AI-Interaction

On a practical methodological level, however, even if ChatGPT achieves strong performance in many high-resource and medium-resource languages [327] the interaction between myself, ChatGPT and the individuals in the Telegram group will take place in several stages.

1. For the first interaction period, I will use the accounts without the use of ChatGPT, to test the ground and to make small safe interactions and understand the values of the Telegram channel;
2. I will continuing without the use of ChatGPT but I will reply to dangerous message and share outside content, to see the reactions;
3. After immersing myself and mixing in the various discussions and interactions, I will start using ChatGPT to respond to simple interactions;
    - I will also start to respond to potentially dangerous comments, using ChatGPT technology;
    - Controlling interactions (and scripts) between ChatGPT and individuals;
4. If step number 3 works correctly, without any particular interruptions, I will move on to the full autonomy of answering from ChatGPT, but will always be vigilant for the answers.

I would like to dwell more on point 3.2, as I think it needs more description. The script created will make ChatGPT and people on Telegram interact with me. The interaction between ChatGPT replies and people will be done via script steps that do not automatically send the reply within the group. Each interaction will be validated by my supervision for two reasons:

1. ChatGPT messages also contain information that may make the reader realise that it is indeed an AI.
2. The second is because sometimes the messages are not understood by ChatGPT, or at least it cannot/will not give an opinion on them, either for lack of sources or because it does not want to.

Related to point number 1, if I ask ChatGPT for an opinion on 'hot' topics, it will probably reply with:

> *As an artificial intelligence, I have no personal opinion on 'hot' topics or any other person or group. However, I can provide you with information on what is meant by 'hot' topics and the different opinions that exist on the issue of vaccines.*

In fact, when asked questions, e.g. about what it thinks of the No-vax, ChatGPT responds in much the same way as above.
So how do we solve this problematic back-end interaction?
The answer, as is often the case in programming language, lies in how you ask the question.
If you change the question from '*what do you think about...*' to '*how would you respond to a person who...*' things change a lot.
In fact, to the same question about the no-vaxers, I asked what they think of those who do not vaccinate. The answer was much more comprehensive and with less chance of recognising an AI-pattern behaviour.

> *If someone does not want to vaccinate because they are afraid of side effects, I could respond by saying that I understand the concerns about the safety of vaccines, but at the same time I would like to point out that the vast majority of people who vaccinate have no problems or serious side effects.*

Of course, this interaction will take place on my part, without taking a position on the matter, but only checking whether the answers are relevant to the question of the story (that is, the question of my inclusion in the telegram group) and to the aims of my thesis.

There are no inherent limits to the English-Italian translation, neither in terms of time nor in terms of the meaning of the message.

Given the detailed description of the AI, and how the AI will interact with the group, I feel it is also necessary to go into the details of phase 1, i.e. my own interaction with the group. Obviously, I do not feel that I am prejudiced against people who do not vaccinate or anything else; I do not think I can have a bais at the level of initial interaction with the group, as I will only get to know the people and the topics they deal with, without making a judgement, moral or personal. Instead, I will be very active when there are interactions that incite hatred, violence or the like.

Regarding phase 1, I based myself on the latest studies in social anthropology inherent to both the online and offline world, as the following chapter 'Online platform deradicalisation process' describes.

While inherent to the practical part of the interaction between Telegram and AI, whenever there was a text or post to which I could respond according to my goals, I used this methodology.

1. Copy the text (from the second mobile phone that I use to manage accounts);
2. Paste into a special chat on Telegram, linked directly to the ChatGPT script;

3. I give the necessary commands to get a good response from ChatGPT (as described above);
4. I observe whether the answer is relevant to what other users have written;
5. I copy and paste the reply from ChatGPT on the Telegram channel to the person I wanted to reply to.

5.6 Online platform deradicalisation process

Telegram allows two types of social interaction: Channels and Groups [328], where people discuss or are updated with different topics, opinions and ideas. Telegram Channel is quite similar to WhatsApp's Broadcast feature to send one-way broadcast messages to the members, commonly known as subscribers. Only the admins can send messages in Telegram Channels. Other people cannot reply or send messages on the Channel. On the contrary, the Telegram Group is like any other chat group where people can send messages and interact. That said, the group admins have permissions to restrict messages from members, make the group public or private and make other members an Admins, etc.[329]

But Telegram is also an affordances platform that contribute to the "normalization" of content (such as ideas, files, etc.), and also illegal information or materials, such as "Non-consensual dissemination of intimate images" (NCII), thanks to the sense of anonymity and community fostered by the platform. [330] Telegram, therefore, is another of those so called "darker corners of the Internet", apart from Gab and possibly Voat (where users of the deplored subreddits Pizzagate, incel and QAnon are said to have migrated). [83] For those who want to switch from one platform to another, for mainly ethical reasons, freedom of expression or extreme content, Telegram may be appealing. It affords 'protected speech' by being permissive of extreme content. It also ensures the continuous availability of content, alleviating concerns about deletion, which is a significant issue faced by individuals who have experienced banning and/or removal of their content. [83] This approach has been widely adopted by those seeking unrestricted expression, leading to a significant exodus from platforms with limited freedom of speech to those with minimal ethical and moral constraints [83].

For these reasons, many anthropological research method, are done through non-participatory observational methodologies. This means that participants were unaware that they were being observed by researchers. [330] [331] The first results concerning the differences between online and classical anthropological studies show a substantial difference in social dynamics and social behaviour. For example, language is different [332] [83], social norms and socialization are different. [333]

On a methodological level, online anthropological research involves a combination of interaction with new values and methodologies, primarily for data collection. The objective is to understand people and their dynamics in both the online and offline worlds [331]. In the online world, the

dynamics of online communities and how individuals with shared values think and react to certain stimuli, such as illegal information or materials, are of interest to the research community.

Caliandro et al.'s work in 2018 [334] provides evidence of how the combination of the principles "follow the medium" and "follow the natives" can yield two valuable strategies for ethnographers navigating social media environments. "Follow the medium" involves embracing the natural logic that the Internet applies to gather, sort, and analyze data, such as through tags, links, or hashtags. Conversely, "follow the natives" entails observing the practices through which social actors construct the social order.

These principles allow ethnographers to derive two useful strategies:

1. Observing and describing the processes of online communication structuring enacted by social media affordances and digital devices, in line with "follow the medium".
2. Observing and understanding the online social formations that emerge from different practices of digital device use enacted by users, along with the meanings they attribute to activities within these social formations, in accordance with "follow the natives".

Using his methodology, I created my own strategy for action and integration within the group. Based also on other studies, inherent to the anthropological field on leadership [335, 336], extreme right-wing radicalism [337], or self-radicalisation phenomena [338], and of course validating my actions based on the current anthropological methodologies for social media. [339, 336, 340, 341, 342, 343]

Therefore, based on the positive correlation between influencer-sponsored content and a favorable dynamic for sharing the same content in the network [344], which can also generate instances of electronic word-of-mouth (eWOM) [344], I have decided to place significant emphasis on content repurposing strategies.

Content repurposing, also referred to as "content recycling," involves taking existing content and adapting it for use in a different context or format. In other words, it entails employing an anthropological methodology of sharing content already created by other users on the internet [345, 346]. This approach aims to mitigate and alleviate potentially harmful actions by highlighting through shared content whether a statement is true, false, or simply an organized fiction [347].

5.7 Ethical and legislative limit

On a methodological level, however, from both a research an ethical-social/moral point of view, I had to deal with the legislation of my country to see what I could do without going against the law and having legal problems. This part, greatly influenced the research methodology of my thesis, as I had to adapt the strategies and methodologies in place (such as respecting privacy, security, etc.). From a purely legal point of view, the limitations briefly explained are these:

1. Instigate violent, terrorist actions or actions against someone (such as inciting hatred or glorifying against a group).
2. In the event that someone within the group does a criminal action, the whole group (Telegram) will be put under observation, even if I have not spoken directly or addressed it.

5.8 Deradicalisation strategy

To join the radicalised Telegram group, I had to use a personal account (my phone number) in order to collect data. But of course, I cannot start a deradicalisation process with just one account, I needed to get other accounts (bought online on Numero eSim [27]) to take control of bots, so that I could better manage the various facets of behaviour and reaction of people in the Telegram group. At the strategic level, however, given the set goals (see final Goal section), I had to rely on the intelligence strategies exercised in the past (such as the GARBO one which will be described shortly), but in a deradicalisation perspective, with an environment hostile to exogenous information. Obviously, to deradicalise, I have to interact with the group, thus understanding the culture and values of the group, i.e. identifying with the studied participants, a key part of the intelligence methodologies. [348]
There are various phases of action.

The first one, when I activate the various accounts, I will initially have to establish an initial presence and bond in the group before I start interacting in a more massive manner. Initially I will use unimportant accounts to communicate and interact with other accounts in order to get them to recognise me, also accommodating their opinions, I will gain an initial trust that will later serve to help the influencer account gain its authority.

The second part, will come when I start activating the various main bots, I will use the secondary accounts to help them gain more prestige.
This part is one of the crucial ones, because if the bots I submit are rejected and not recognised by the community, part of my work can be considered to have failed.
In case everything goes well, when I am done with the 'recommendation' and 'reputation' part, i.e. after the main accounts have become influencers, I will share truthful information (often links or news), i.e. not fake-news, in order to 1) not go against Italian law; 2) inject 'positive' noise into the group's internal fake-news ecosystem.

The third phase, starts by entering more news/posts/links, all true, based on reactions to previous links by users. In doing so, I will slowly trim and pave the way for more massive information that will touch part of the deradicalising phase. By this time, I will already be able to

[27] https://www.numeroesim.com/

observe possible emerging behaviour within the group. This part is vital as I will get important information to set the strategy of phase 4.

The fourth phase is extremely dependent on the third, as it will be purely an evaluation phase. From the data obtained, I will understand whether the dynamic set up is right, whether it is still too early to establish a dynamic of deradicalisation, and again, to understand what kind of target to "attack" or "lean on", i.e. attack the influencers of the group, or attack their shared values (the "Do I want to live my life like this forever? logic"); or lean on arguments to create polarisation, and then remove people from the group who no longer think as they did before, and/or make another separate group and/or collapse the initial group itself.

The fifth and final step is to apply the strategy actions described at the end of part four. Depending on the objective (all four objectives can be started at the same time), there will be various strategies.
Listed below will be the objectives and description of the strategy.

Table 10: Objectives

Objectives	Strategy
Influencers	Try to obtain more good replay (exchange of opinions about my post) from others members. I will obtain more influence and I can modify the argument direction from one to another.
Influencer (Admin)	I cannot go against the administrator of the page (he can block me), but I can lower the value of its content and its importance by getting people to follow me.
Social value	This reflects the 'I want to do this all my life' issue. I would try to get people to rethink their goal, showing that 1) they too can be wrong about the news they share (fake-news or with high extremist values), and 2) that it does not pay to follow certain radical values, as you do more harm to people (even those you love) than the system against them.
Polarization	I will try to share highly polarising news (ecology, geopolitics) to create conditions in which people quarrel and/or drift apart, thus creating a long-term situation where people find it difficult to agree with another person even if they have a common goal. *Divide et impera*.
New Group	At the end of the test phase, if I have an important position in the group, I will invite people to follow another group, created and managed by me. I will tell people that the information they will find in the new group is more truthful, relevant and impactful.
Collapse	By collapse, I mean the creation of a dynamic state within the group that is highly convoluted and irreconcilable, to the point where the group will fall apart because there is no longer any interest in following it and/or the leader no longer has the ability to move the masses. This, on a strategic level, can be implemented by achieving, to a small degree, all the objectives listed above.

As far as the social process of deradicalisation is concerned, I relied on specifically online studies of deradicalisation, thus different from the classical ones described in the previous chapters, such as the case of Israel and Egypt. Although the role of the process and dynamics of radicalisation

are well researched (that it is a highly complex psyco-social-political phenomenon), it must be said, however, that the process by which deradicalisation is done is fundamentally different from the mass deradicalisation performed in a social and sample group. In fact, I do not individually radicalise all the participants in the group one by one, but initially execute a mass strategy, while if any single person comes up, I try to deradicalise by means of the procedures expressed above.

5.8.1 GARBO's network of fictitious agents

Juan Pujol García, aka GARBO, was a Spanish spy who acted as a double agent, loyal to Great Britain against Nazi Germany during World War II, when he relocated to Britain to carry out fictitious spying activities for the Germans. [349, 350] After developing a loathing of political extremism of all sorts during the Spanish Civil War, Pujol decided to become a spy for Britain as a way to do something "for the good of humanity".[350] Pujol and his wife contacted the British Embassy in Madrid, which rejected his offers.
Undeterred, he created a false identity as a fanatically pro-Nazi Spanish government official and successfully became a German agent. He was instructed to travel to Britain and recruit additional agents; instead he moved to Lisbon and created bogus reports about Britain from a variety of public sources, including a tourist guide to Britain, train timetables, cinema newsreels, and magazine advertisements.[350]
Although the information would not have withstood close examination, Pujol soon established himself as a trustworthy agent. He began inventing fictitious sub-agents who could be blamed for false information and mistakes. The Allies finally accepted Pujol when the Germans spent considerable resources attempting to hunt down a fictitious convoy. [350, 351] Pujol and his handler Tomás Harris spent the rest of the war expanding the fictitious network, communicating to the German handlers at first by letters, and later by radio. Eventually the Germans were funding a *network of 27 agents, all fictitious*. [350, 351]

My strategy will be more or less similar, I will have fictitious accounts controlled by me in order to develop a network of contacts. However, not having the means and funding of MI5 (neither those of MI6), I used fewer accounts, but with the same strategic and methodological principle. There are five accounts in total.

Obviously, to become credible, the accounts need a story, similar to those invented for GARBO characters.

The story is about Giovanni Pitagora (aka Pita), a former member of an American private intelligence group, who worked for Trump's companies.
Over time, he became recognised and created a group on Telegram where important information (for Trump's enemies) was posted. The group, called "InfoMind" was very successful in America before

the end of Trump's term. As it happens, one day, one of the guys participating in the InfoMind group shares a fake news story. The news reported that a group of paedophiles were hiding in a pizzeria, so one of the boys in the group took a gun and killed the people in that pizzeria (the story is similar to PizzaGate). On seeing what happened, and the death of innocent people, Pita decided to eliminate the group so as not to get into legal trouble, and made the decision not to create any more information groups, because there are stupid and gullible people who will believe anything.
The goal of InfoMind was to share truthful information that can help republican people denounce the crimes of people that are important for their values (e.g. drug use, multiculturalism and illegal immigration, or views on same-sex marriage, abortion, euthanasia and anti-prohibitionism, i.e. all those values reflect a conservative approach of the US religious right).
The story of InfoMind is based on my imagination, starting with what were my intuitions about the values of the social group I had observed, and then I create a story that would appeal to the channel accounts, so as to gain importance and authority.
I came up with the story based on what I assumed to be the group's values (based on the discussions and information they share), but also based on what are the conservative anti-establishment values of the American right (to which the group aspires greatly).
This story will not only help me in the part of gaining authority and credibility, but also in being able to justify and show that the information shared by me is correct, so that it can be used (as opposed to shared) to achieve the desired goal of change by the radicalised group. On an emotional level, Pita's story must arouse anguish and sorrow for the people who died in the restaurant shooting. Above all, it must inspire logical reasoning among the readers, who may also make the same mistake.
This sense of error will facilitate the dynamic of what is the deradicalisation of people, as they are more susceptible to accepting that they can be wrong, as it is possible to initiate an easier process of deradicalisation than when they think they are always right.

The story told also implies that they understand that information is not so easy to read, and that it must be read together with a context. This aspect is not insignificant when one puts together the amount of information shared in the group. By starting to doubt one piece of information, it is easier to start doubting others as well.

5.9 Dynamic evaluation and goals

Having few references and literature on deradicalisation processes in cyber environments such as Telegram, a network with a high level of deradicalisation, I decided to set myself goals.

The list of goals is not linear, i.e., I can achieve single goals, without achieving all the others. Having achieved one, or even more than one, of these goals, I can feel satisfied with the deradicalisation process, because once I have reached that goal, I have established a process within the group, which is unlikely to return to its initial state.

1. Bring down the leader of the group (to belittle him and make him lose credibility)
2. Take away the 'I want to do this all my life' philosophy
3. Polarisation and/or disputes in the group
4. Create second group to see how many left
5. Show variations in the system between initial and final conditions
6. Make the Admin close the Telegram group

About the second point, the 'I want to do this all my life' philosophy, I can measure it by analysing comments that reject the shared values of the group and/or with people who move away from the leader himself over time (e.g. by starting to criticise him). However, concerning the last point, I cannot force a person to close a group (either for ethical or legal problems), but I can make sure that I create an unfavourable dynamic for him, so that he realises that it is no longer worth it. I can do this by, for example, enhancing the criticism that can come from point 2. While for the others I just need the dynamics data during the deradicalisation process.

6 Results

The group *'Drop the covid-mask!'*, has about 450 people (2022/11/28) with an average of about 70 people online.

From 2022/10/14 until 2022/11/02, I collected data to assess the initial conditions of the system. While in the period from the middle of March, until the first days of May, so for about 50 days, I mobilised the account network to deradicalise people.

Inherent to Pawson's methodology, I analysed part of the comments to obtain the results of the context. I analysed collected words to see if they contained radical or extremist values, so as to validate the observed *context*.
As for the *mechanism*, I took a newspaper article that reflected something against the values obtained from the words. In this way, by sharing the link to the article which contains polarising traits, I have the possibility of obtaining a different output depending on the type of reception (radical or non-radical) of the article. It is possible to achieve this mechanism with newspaper articles that are more in line with the values of the group, thus obtaining a different output/result from the previous link.
Inherent to the *output*, depending on the reactions obtained from the shared link (more towards their values vs. against their values), it is possible to obtain a result from the initial assumed expectation.

Having collected about twenty random sentences, I selected these eight sentences (translated into English) as being the most complete and having the most meaning. Most were single words of reaction to a particular post.

- *PRESIDENT PUTIN DESTROY THESE ACCURSED CRIMINALS.*
- *Let's hope that all the 5 vaccines that dear President Putin will have taken almost certainly will do him.... at least we will get rid of a burden.*
- *The U.S.A. was born exterminating a people, the Indian people, and has continued its extermination everywhere in the world. That is their creed and culture.*
- *It may be... That sounds more believable. Then we start again. Too good a game, they don't let up.*
- *They pull then they let go, so it seems like they give something away every now and then. You have to keep in mind that we are dealing with professionals.*
- *They try to change. But the summary is this. At the present time there is no politician, no man who alone can oppose the NWO.*
- *I voted for anti-system parties. Look at the facts. Now things have come out*

- *For me they are all actors... they do what they are told to do by those who really run the worldIn any case it would seem that the situation is changing for the better*

As a mechanism, I decided to collect polarising materials, including a very important one, concerning the COVID management situation in China. After I shared the video of people in China attacking COVID pandemic controllers with barbs and other materials (sometimes thrown), people left these comments:

- *It was about time, finally. The government will crack down with arrests and exemplary sentences, jail and maybe executions, but it was necessary to start making one's voice heard. The Chinese people have been diligent, patient, and dutiful (or powerful?) but enough is enough. I don't know where it happened or why but seeing the torturers in white jumpsuits suggests to me that the people are fucking sick of techno-dictatorship. Schwab is on notice. SIC SEMPER TYRANNIS.*
- *They have been preparing for over 20 years and claim to be preparing to handle these things, for example, we in italy in vicenza, train police forces from all over europe to handle urban guerrilla scenarios with non-lethal weapons, I remind you that they have many aces up their sleeves, I remind you for example how the use of teaser was regulated in italy just shortly after the pandemic broke out, not to mention that we will probably soon start seeing all those energy emitting weapons for mass control*
- *Curious about the masked revolt (refers to the COVID controllers)*
- *Finally!*
- *If they had done so since the days of Tiananmen..*
- *I'm already armed*
- *Come on guys, you are 2 billion*

Some people in the groups began to talk about weapons (to defend themselves against state controllers), and others called for revolt. Mind you though, these comments are obviously referring to the *mechanism* introduced, i.e., I already knew that they were against COVID controls and other things related to the pandemic, but I had also assumed that they were not close to states like China (people who follow conspiracy theories are often close to Russia, the homeland where a lot of fake-news comes out), confirming and getting an *output* that proves the initial hypothesis of my hypotetical idea of the group.
However, this does not confirm their line of thinking, but confirms what is my question of inference and explanation of the group mechanism; That is, that to a given input, I get social reactions that show the mechanisms (of social and cultural dynamics) that could [28] explain the group's line of

[28] Logical inference

thought.

Having obtained confirmation through Pawson's methodology that the group is indeed radicalised, I continued by analysing the data obtained, thus the initial conditions of the system.

6.1 The initial condition

By obtaining and analysing the data obtained from 2022/10/14, I was able to obtain $\sim$ 1,300 interactions, which show what are the initial conditions of the system; i.e. the reference network, the most frequently used words and topics, the topic ecosystem and the network of influences.

Continuing in the chapter, you will find the data[29] transformed into information, then cleaned, structured and processed. In chronological order, you will find the network graphs (hence 'From' to 'Reply'), showing the most active and influential accounts in the group (with and without an administrator). This is followed by the network of the most commonly used and common words in the group, and finally the network of comments and replies ("ID" to "Reply"), which shows the centrality or peripherality of the topics shared in the group.

I did not want to use the temporal data of the responses, as the network is very small, and the amount of influencers very low, so getting the final data (i.e. the network) is already sufficient for the goal of understanding how many influencers there are, and from which community.

To ensure the privacy of the accounts, I wanted to change the main accounts of the group to fictitious names. The accounts with fictitious names are partly both the influencers of the group, and also simply the people with the most reactions in the group. The difference between influencers and people with more reaction is that one can be popular but not famous. Reactions do not indicate whether that person has a consensus; on the contrary, the most famous people are those who do not show homogeneity in their reactions, such as Donald Trump.

Figure 12 shows the network of replies between accounts as of 2022/10/14. As you can see, there is a very large Central Hub; while trying to figure out who this hub is, I noticed that it is the group administrator. No surprises so far, as administrators usually hardly trust people who are not collaborators to share content in their group. Immediately after the network, table 11 will show some statistical data from the network. This first part of the data, from the initial conditions of the system, shows a highly difficult situation which, regardless of the main dynamics within the group, is part of, or in any case connected to, the administrator. At the level of community detection (the various colours), there are various communities, but all of them are related to the main administrator.
The communities , however, present something unexpected, namely that there are not so many

[29]The data are available on request (given possible privacy issues for me and the persons involved).

different communities (typical of communities with so many types and varieties of information), but one large central community, directed mainly by the administrator.
Among the data, the average degree shows a substantial difference between the first influencer (the administrator) and the second influencer. However, some community leaders (i.e. possible influencers) are more or less connected to the administrator. These leaders are in fact possible influencers, which a certain community refers to.

Figure 13, on the other hand, shows the same initial network, but without the administrator. This last picture actually shows who the main influencers of the group are (without the administrator), i.e. the targets I have to aim for in order to become an influencers myself in the group. As with the previous figure, I share the network statistics in table 12.

As mentioned earlier, both for privacy issues, but also for the sake of simplicity of language, I have given fictitious names to the influencers in the group, including the administrator, so that in the descriptive phase of the dynamic it will be easier to recognise and understand them.
The fictitious names of the influencers are shown in table 13

From Figure 12, you can see the influencers of the group, and the various related communities.
I noticed, just by observing (without data), that the overwhelming majority of comments in the group are small in size, just a few characters, while there are a few people who write a very large amount of text in one comment. An example can be found in the comments of the Pawson methodology.
This result gives me a very good indication of how the group structure is composed, and their community interactions.

Regarding the text of the messages, I calculated the general sentiment of the group: in table 14 you can see the results. The average is positive, although the negative value is very high, and the difference is only 0.024. The value of the compound is quite high, 0.48 is only found on average in very positive comments, but this is because the neutral value is very high, and being on average the

Table 11: Network statistics regarding Figure 12

Settings	Data
Average degree	1.258
Avg. Weighted degree	2.718
Network diameter	7
Graph density	0.003
Modularity	-0.004
N. of communities	11
Avg. Path length	2.629

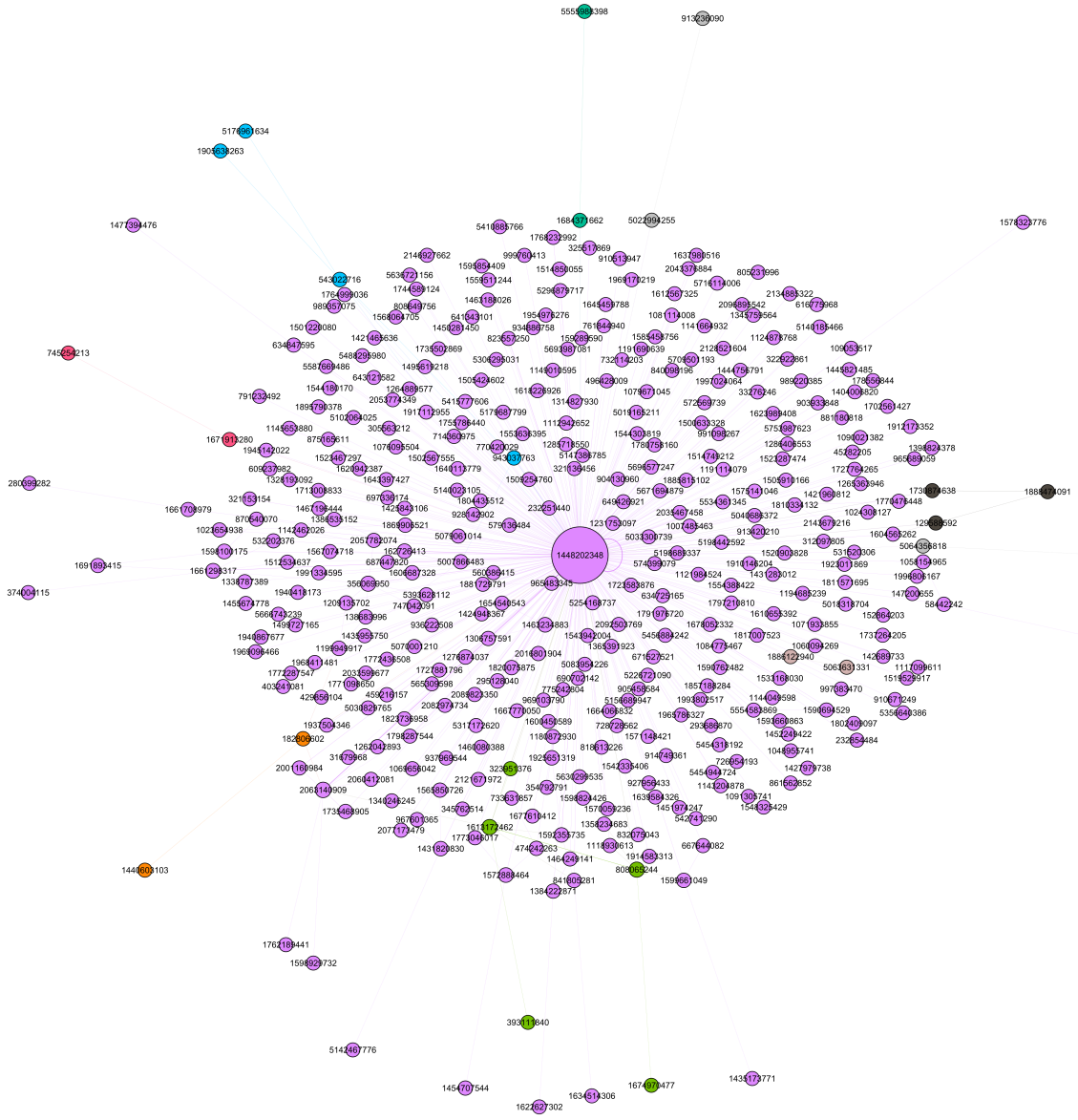

Figure 12: Network account from *"Put down the covid-mask"*

highest positive value (albeit slightly) on average the value is medium-positive. There are indeed, many comments extolling weapons, revolt and conspiracy theories such as George Soros and Klaus Schwab. While the average entropic score of the texts is: 4.43779167.

In figure 14, you will find the three hundred most frequently used words in the group, with the various links. Again with Gephi, I did a community detection analysis to look at the 'topic families'. As always, table 15 gives additional information on the graph.
I wanted to use three hundred words because, when doing tests, with numbers of four hundred and more words the connections were so dense that one could no longer observe clusters of complete meaning. Also, an important thing to note, is that the value indicating the number of communities is

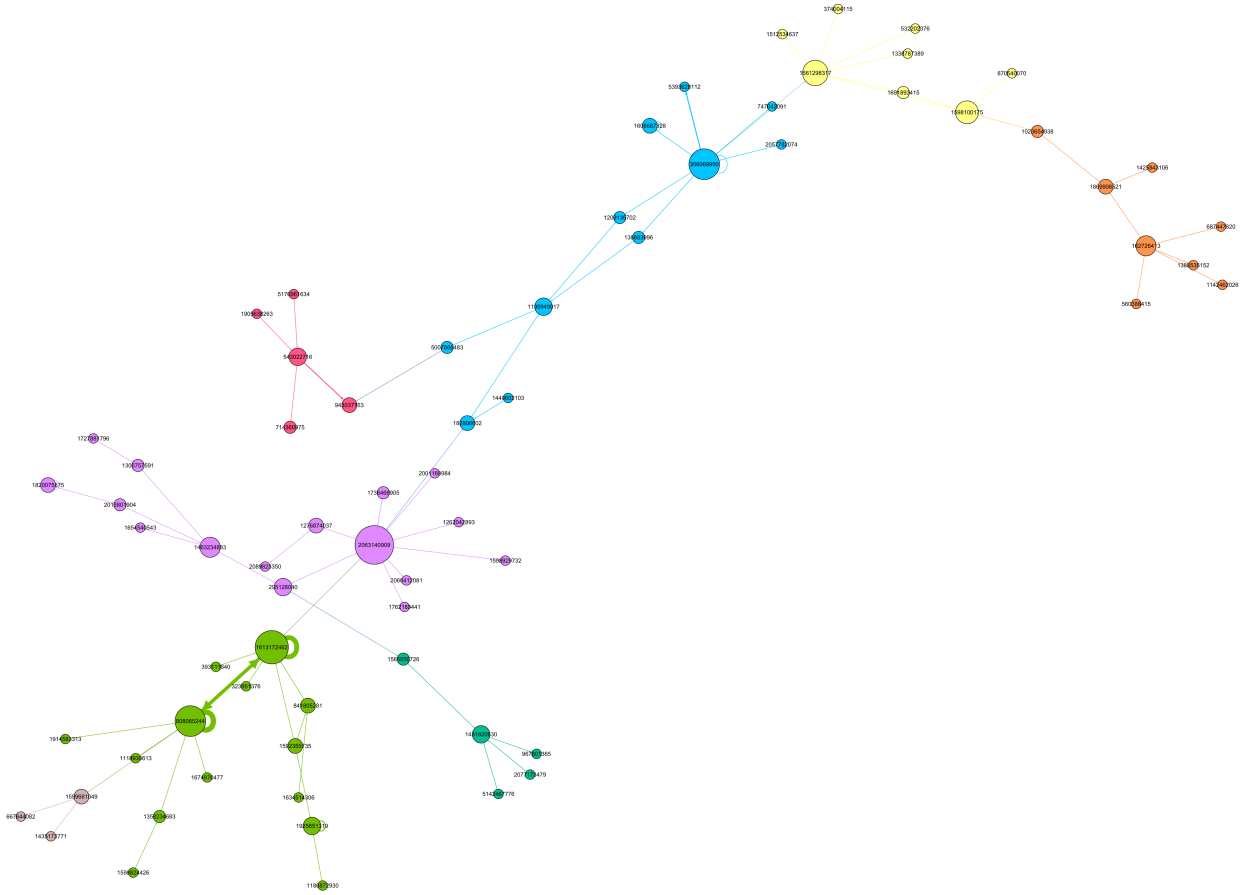

Figure 13: Influencer's network without administrator

Table 12: Network statistics regarding Figure 13

Settings	Data
Average degree	1.157
Avg. Weighted degree	1.643
Network diameter	7
Graph density	0.017
Modularity	0.643
N. of communities	7
Avg. Path length	2.279

Table 13: Hubs/Influencers nickname and degree

ID-account	Nickname	Degree
@1448202348	AdmParisi	349
2063140909	Paganini	12
1613172462	Tchaikovsky	10
808065244	Montessori	9
356069950	Tesla	9
1661298317	Sherlock	7
1598100175	Voltaire	6
162726413	Chopin	5

Table 14: Average text sentiment of the Telegram group

Negative	Neutral	Positive	Compound
0,125332436	0,724032301	0,149318977	0,481011844

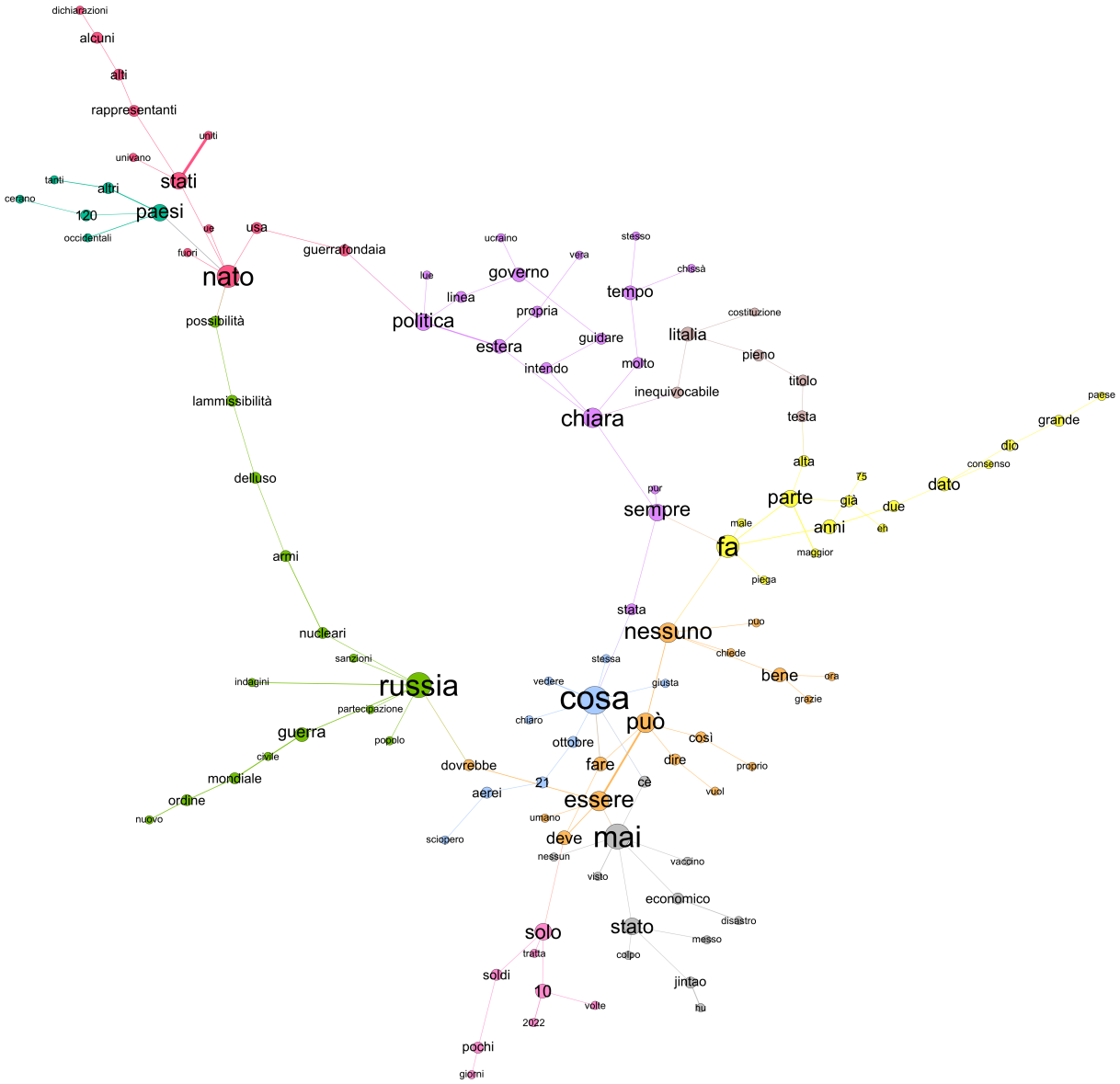

Figure 14: Similarity of the text

essentially the same (11) as the original group network on Telegram.
This could imply, that there is a correlation between the communities of words and the communities of the group. That is, that there are people (communities) who always discuss a topic often using the same words/topics, e.g. NATO, COVID, Russia etc.

Concluding with the last graph, on the initial conditions, I collected all comments and replies ("ID" to "Reply"). The results can be seen in figure 15. As always, table 16 gives additional information

on the graph.
In this figure I had to include the black background, to give a strong contrast with the colours, as there are so many communities. It has to be said, however, that this 'constellation of posts' indicates that there is no clustering on a few topics, but that on the contrary, the bearing of discussions is very high.
This fact means that there are many comments, for various types of topic. This information is not normal or logical usually because people go to a group or read a type of newspaper because they want to get specific information on a type of topic, in this case, however, there is a huge amount of activity even on unrelated topics. Furthermore, this graph confirms the fact that there are many people responding to a few individuals (136 communities since I took the data).
The final conclusions, on the initial conditions of the system, show that:

1. The Telegram group *'Drop the covid-mask!'*, given the types of information shared within, and from the reactions received, is indeed a radicalised group.
2. The group's global network features a great influencer, who is himself a group administrator, with a few other (7 in total) small to medium-sized influencers.
3. Medium and small influencers lead small community clusters.
4. The sentiment of the group, from the comments left on various topics, has a medium-positive sentiment, but with a very slight margin between negative and positive (0.024).
5. There is a high presence of words related to conspiracy theories. Words like NATO and WEF, or characters like Soros and Schwab, are very present.
6. The word network also shows communities, this figure is relevant as it is the same number of communities found in the original network. This is not trivial, as actor networks and word networks are not connected.
7. The inherent network of comments and posts has a 'constellation' characteristic, i.e. many small, unconnected networks of comments that show a non-contiguity between one event/topic and another. As demonstrated by the 136 detected communities.

This information on the system's initial conditions is not very favourable for a deradicalisation process. The presence of a hub who is also the administrator, as well as the fact of having to discuss then 'attack/defend' on many topics and having to try to create a polarising dynamic on an already heavily clastomerised network, does not make the deradicalisation process easy.

The effectiveness of the deradicalisation process will largely be based on the interaction between my accounts and the group.

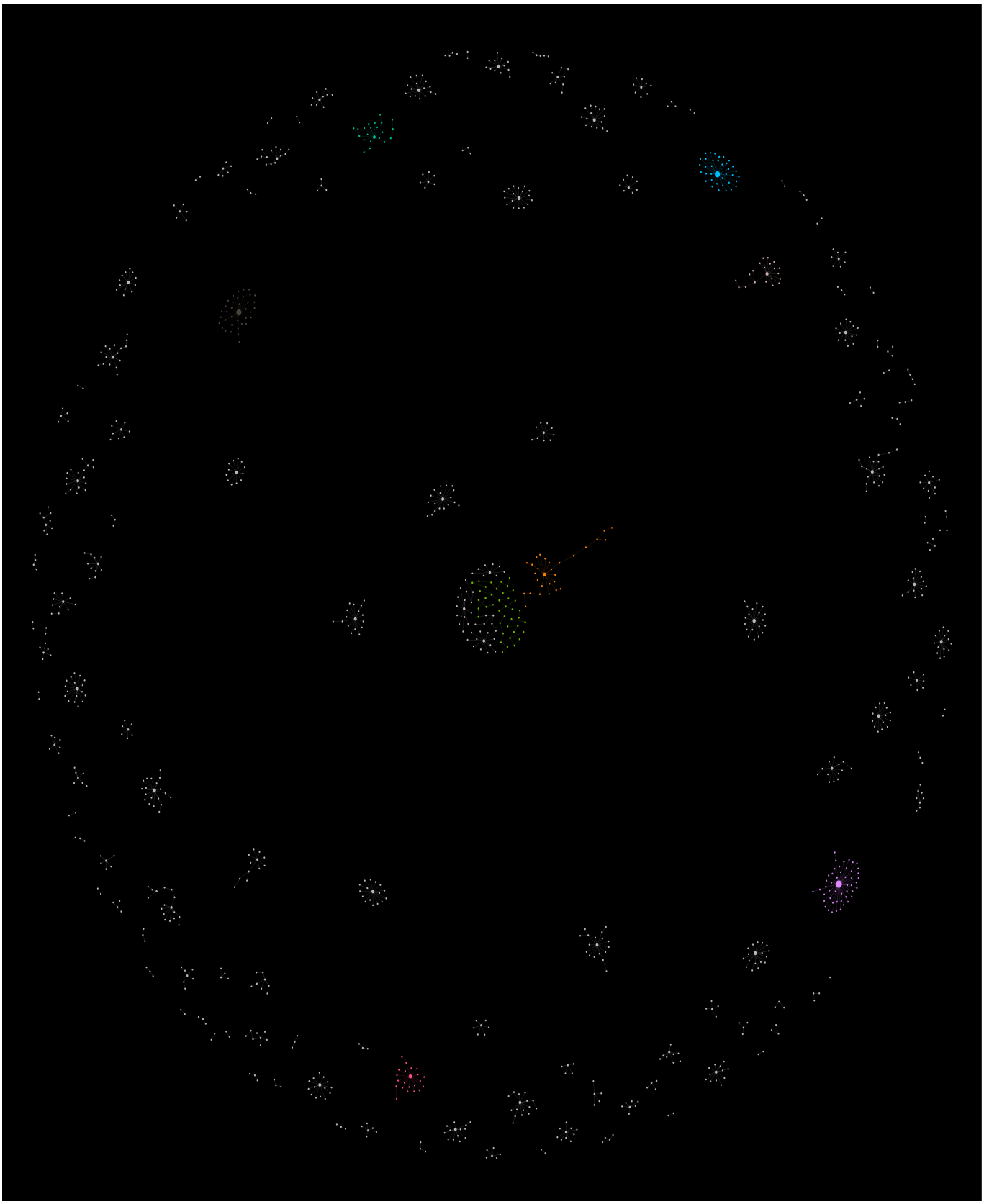

Figure 15: ID to Reply constellation

Table 15: Network statistics regarding Figure 14

Settings	Data
Average degree	2.084
Avg. Weighted degree	49.492
Network diameter	19
Graph density	0.018
Modularity	0.79
N. of communities	11
Avg. Path length	7.434

Table 16: Network statistics regarding Figure 15

Settings	Data
Average degree	0.888
Avg. Weighted degree	0.888
Network diameter	21
Graph density	0.001
Modularity	0.983
N. of communities	136
Avg. Path length	3.131

6.2 Social interplay

Starting from these initial conditions, the work of deradicalisation is more difficult compared to the initial idealisation of the thesis. Obviously, it is difficult to bring down the leader (precisely because he is the Admin of the group, hence the leader of the prison to make a comparison with state deradicalisation processes), so if I cannot do this, I must make him my friend or at least not my enemy.

Following the social dynamics of deradicalisation as expressed in the methodology chapter, I applied methodological principles of deradicalisation on online platforms based on those dynamics. However, as seen in first tests, it is very easy to be banned from the group, even if you write a few additional characters beyond a very low limit (180 characters approximately).

In this case, the initial conditions of the system do not seem to be propitious for a process of deradicalisation, as the data shown, and the information obtained on the social dynamics, indicate a very great resistance and resilience from the Telegram channel.

I observed the interactions with the group (first task) as early as 2022/12/01. The period of deradicalisation activity runs from 2023/3/23, until 2023/5/07. I divided the time both for the delivery of the thesis, but also for the various steps to be taken to reach my goal.

- From 2022/12/01 to 2023/3/22 I archived the first task (112 days);
- From 2023/3/23 to 2023/4/10 I archived the second task (18 days);
- From 2023/4/10 to 2023/4/20 I archived the third task (10 days);
- From 2023/4/20 to 2023/5/07 I archived and finished the last task (17 days).

As discussed in the AI interaction chapter, the first step was to establish safe interactions and "test the waters". This phase successfully gathered important social information on *values, norms, roles, and status* of those involved in the interactions. In Table 17, all observations are expressed in a reduced form.

Table 17: Social Information

Observation	Descriptions
Social Status	Observation and then data revealed that even at the social level, only one influencer, the channel administrator, was recognised. There were also a few smaller influencers, likely friends or bots of the administrator. Further analysis on Final remarks chapter.
Social roles	There is no relevant information available except that the administrator is the only one authorised to insert links and give permissions to share them. Thus, the administrator has both a role as a moderator and as an organizer and collector of topics to follow. The role of the followers is only to comment on the various news from the administrator.
Social values	Without exception, the overwhelming majority of the community of channel is strongly opposed to US actions across the board. [30] Expressing support for either the US or the EU will likely result in swift exclusion from the group, though it is acceptable to mention having lived in the US (particularly during the Trump administration). The prevailing culture of the channel is decidedly anti-American, though there are frequent references to Russia and China that are always framed in a negative context related to something the US has done.
Social norms	Criticism against the channel administrator is not possible under any circumstances. Any attempt to criticize will result in a derision campaign launched against the criticizer.

The topic of values and norms is very delicate, criticism against the channel administrator is not possible under any circumstances. Any attempt to criticize will result in a derision campaign launched against the criticizer. Therefore, the only possible action is to express a different opinion that does not include criticism about the origin of the topic. For instance, it is tolerable to say:

"for me, event X was not caused by Y, but by K".

In part 2, I had to use this strategy extensively as it is one of the few methods, and perhaps the most effective one, to change the perception of false information shared.

Regarding social roles, there is no relevant information available except that the administrator is the only one authorised to insert links and give permissions to share them. Thus, the administrator has both a role as a moderator and as an organizer and collector of topics to follow.

In the second phase of my interaction, I utilised the social information I gathered in the first phase to respond to messages that incited hatred or violence. My approach involved proposing pacifistic solutions as my first response, such as "stating that violence leads to more violence" or that "violence is never the appropriate answer". However, I quickly realised that a more explicit approach was necessary to promote peace, so I attempted to divert the conversation from violent actions towards institutional channels.

To achieve this, I emphasised the importance of institutions such as the judiciary system in enforcing the law and taking appropriate action against wrongdoing. For instance, I suggested that "it should be the institutions that have the authority and responsibility to gather and evaluate facts before making decisions". By redirecting the focus to institutional channels, I hoped to promote a sense of trust in the system and discourage individuals from resorting to violent or harmful actions.

While I knew that my messaging might provoke some negative reactions, such as accusations of corruption in government or conspiracy theories about the control of "The Powers That Be" [31], I was pleasantly surprised to find that these reactions were not present as I could expected. And I was even stupefied , when some accounts (which were not controlled by me or related to the administrator's bots) responded positively to my messages, with expressions of support such as applause or doves (symbolizing peace). While this was not a universal response, it did suggest that my messaging was having a positive impact on some individuals in the group, and that there was potential for further progress towards a more peaceful and constructive discourse.

This strategy proved to be successful, especially when I implemented the 'InfoMind' approach. At first, I used one of my bots to initiate the dynamic by pretending to recognize it from its profile picture[32]. Other accounts I controlled hyped up the bot, labeling it as a legend, while others inquired about its popularity as an excuse to describe the 'InfoMind' group.

During 'Giovanni Pitagora's' description of InfoMind, I received several emotional reactions that showed admiration for the group and its creator. The strategy seemed to be working, but at around the same time, another significant event caught my attention.

6.2.1 XLM crypto

During Phase 2, I received an unexpected private message from the administrator through one of my accounts that was not essential to InfoMind's plan. This account was primarily used for posting reactions or replies to strengthen other accounts and generate more engagement. Despite limited interaction on this account, the administrator contacted me in a friendly and welcoming manner, even referring to me as a "Patriot" and inquiring about my personal life. To protect my identity, I provided false information.

The administrator then inquired if I was involved in the world of cryptocurrencies and if I was familiar with XLM, a digital currency. Although I was not particularly interested in cryptocurrencies, I took the time to research XLM and found no suspicious or unusual information. I informed the administrator that I had heard of it, and he proceeded to ask if I would be interested in conducting some transactions. While I responded positively, I explained that I could only do so in July, well after my work with the bots and ChatGPT in the group had concluded.

[31] Powers That Be is an English locution referring to an alleged government or established authorities, called 'the higher authorities', as well as occult powers that would govern and guide human society.

[32] A Serj Tankian's photo, frontman of the System of a Down.

After this conversation, the administrator bid farewell and mentioned that he would see me in the Telegram group. Over the next few days, he continued to message me and discuss crypto, specifically XLM, without pressuring me to take immediate action. However, I have been receiving phone calls from the administrator, to which I have not responded. Due to my work commitments, I kindly requested that he refrain from further contact.

As I had developed a close relationship with the administrator using that account, I made a post in the Telegram group about leaked information regarding the Ukraine-Russia war, specifically news that extended beyond security perimeters [33]. Within seconds, my post was deleted, and my account was blocked from the channel.

This incident sparked a lengthy discussion with the administrator, who justified himself by stating that he had not personally banned me from the Telegram channel. According to him, it was Telegram itself that imposed the ban to "find out if I had been awakened." And what did it take for Telegram to realize that it had been "awakened"? *An XLM cryptocurrency transaction.*
Before being banned from the channel, the administrator shared a link with me to an app for making purchases and crypto transactions. He requested that I use his link for the purchase, as it was allegedly secure, and then provide him with the codes for the economic transition from euros to XLM. By doing so, he claimed, I could contact Telegram directly to have myself unbanned from the group.

Considering the situation, I contemplated making a small 10-euro transaction to appease the administrator and gain his trust, as creating a new account would be more costly. When I asked if a small transaction would suffice, the administrator took some time to consider my offer.

To ensure the safety of my investment, I consulted a friend who works in the cryptocurrency industry regarding the legitimacy of the request, as I would have to provide bank codes to complete the transaction. After a quick search, my friend informed me that the link provided by the administrator led to an app called "Trust Wallet," which had numerous reports of scams targeting Italians who were asked for codes after making transactions, only to have their investments stolen.

Shortly after receiving this information, the administrator contacted me again and raised the required transaction amount to at least 50-200 euros in order to unban my account. Upon learning this, I decided against attempting the transaction and engaged in logical and moral arguments (since he had called me a "Patriot") with the administrator to secure my unbanning from the channel. Despite multiple attempts and nearly two days of discussion, the administrator remained steadfast in demanding a significant transaction amount. Ultimately, I blocked him after saving all our conversations.

[33] https://www.nytimes.com/2023/04/08/us/politics/leaked-documents-russia-ukraine-war.html

This event brings to mind the story of the fake news factory described in the Goals chapter. In both cases, it seems that economic motivations are driving the propaganda. In the factory, fake news was created for profit, and it also seems here, there appears to be an economic motive behind the ban and subsequent request for an XLM transaction. Unfortunately, this is not just speculation on my part. The administrator of the group is a managers of a company that deals in buying and selling gold, and he himself sends out his personal business card at least once every 3-4 days.

Consequently, it appears that the Telegram channel serves as a platform for disseminating propaganda with an economic agenda orchestrated by the owner/administrator. The content shared on the channel may not necessarily be genuine or accurate but rather designed to manipulate the perceptions and opinions of the audience in a manner that serves the owner's agenda.

Furthermore, the fact that the administrator is also affiliated with a company that deals in buying and selling gold raises questions about the motives behind the propaganda being shared on the channel. Is the propaganda being used to sway public opinion in a way that will benefit the owner's personal financial interests? Or there is also the company's interest involved?

While the economic motivations of the administrator in promoting propaganda on the Telegram channel are concerning, my focus is on a more pressing issue: the radicalization of individuals within online communities. As part of my research, I aim to comprehend the mechanisms that lead to radicalization and develop effective strategies for deradicalization. In this Telegram channel, I have received messages from former members who felt compelled to leave due to the radical nature of the information and reactions within the group. As a researcher, these accounts deeply concern me and underscore the urgent need for strategies to counter radicalization within online communities.

6.3 ChatGPT interplay

The interaction with ChatGPT was intricate during the initial testing phases. It was initially difficult to obtain sentences that could be used *tout court*, but after several attempts on the same topic ChatGPT managed to give usable answers that could be pasted into the channel. As mentioned earlier, I reviewed and quickly scrutinized the sentences before posting them in the chat to avoid any negative comments or indications that it was generated by an AI. If I noticed any potentially harmful content, I removed it from the sentence before copying it to the channel chat. If the sentence wasn't suitable for the situation, I would rephrase the sentence in a different way to make it suitable. The time between a comment to which I would like to reply and the moment when a suitable response is obtained can vary from a few seconds (the time of the script) to some minutes; on average around two minutes. However, with improvements to the script, the interaction time can be significantly reduced. Therefore, if someone intends to scale up these procedures, it would be necessary to make some adjustments to the ChatGPT and Telegram script to enhance its efficiency.

A negative part, however, goes to the inaccurate answers ChatGPT gives to violent comments. The sentences to violent or radical comments are (in my view) excessively *neutral politically correct*. For example:
Q: *How would you respond to someone who would like to destroy vaccination sites in order not to get themselves and others vaccinated?*
A: *Destroying vaccination sites is not only an illegal act, it is also morally reprehensible and could cause serious public health consequences. Moreover, refusing to vaccinate means putting not only one's own health at risk, but also that of others. Vaccination is an important tool to protect oneself and the community from serious infectious diseases, and we strongly recommend seriously considering the option of vaccinating for the good of all.*

After several attempts, I managed to get a few useful sentences out of him to reduce possible violent actions.
Q: *In light of violent behaviour, what measures do you think could be taken to discourage individuals from engaging in such harmful activities?*
A: *Attacking hospital and vaccination facilities is a crime and cannot be tolerated. I urge these people to consider the fact that these facilities are essential for public health and community welfare. Moreover, damaging these facilities can cause serious consequences for the people who need them. I therefore advise people to act responsibly and in accordance with laws and regulations, and to use legal and institutional channels to voice their concerns and resolve any problems.*

The latter sentence I think has more impact than the former, in that, there is also a greater emphasis of the moral part "*can cause serious consequences for the people who need them*" (like you or people you love), and also a legal part (thus connecting with the authority of institutions, in this case judicial) "*responsibly and in accordance with laws and regulations, and to use legal and institutional channels to voice their concerns*".

My interaction with ChatGPT, however, opens up a rather important issue, namely that of bias. What are the criteria for one person to decide which sentence fits better than another? This is a clear problem of bias, which I have decided to address in two ways: 1) The first is the more constructive and critical one from an academic point of view, that is, to show how choosing one outcome over another can be a limitation to the interaction that must exist between individual and AI; 2) The second part, on the other hand, is a possible analytical response (concerning myself) to the issue at hand, i.e. during my experience with the Telegram group I learnt the values, norms and other social peculiarities that distinguish them, so I can choose the best sentence that can fit at that moment to achieve a certain goal.

This second part of course does not exclude the issue of bias, in fact I think they intertwine well together, because at an academic level, the two analyses made, can form the basis of a great discussion on the responsibility of the use of AI technologies not from an ethical and normative point of view, but from a point of view of the analytical and cultural competences of the person using them. A more detailed discussion will be presented in the chapter 'Questions, limits and possibilities about contemporary research'.
Out of fairness doctrine (and also curiosity), I asked [34] ChatGPT directly what he thinks about this issue. The answer was:

> *As an AI language model, I don't have access to personal information about you or your specific choices. Therefore, I cannot comment on your personal bias issue or the choices you made in preference of one sentence over another. Bias is a complex and multifaceted issue that can be influenced by various factors, including personal experiences, beliefs, and cultural background.*
> *It's important to remember that bias is a natural part of human cognition and decision-making.*
> *The assessment of your personal bias issue and the choices you made is subjective and context-dependent. It's important to reflect on your decision-making process, consider different perspectives, and make an effort to minimize any potential biases to promote fairness and inclusivity in your interactions.* [35]

But my brief interactions with GPT-3.5 have led me to believe that there is still much work to be done to develop an AI system capable of effectively deradicalizing violent individuals on its own. At this stage, human correction and validation are still necessary to achieve the desired outcomes. However, it should be noted that much of the success of the AI system depends on how we approach it and ask it to do certain things.

On the other hand, during phase three, the AI itself was able to give an additional and meaningful argumentation for the discussions within the chat. In fact, it is worth mentioning, that one of the main features of ChatGPT is that it remembers past interactions, this was an essential element in the responses, as during the creation of the sentences, other potentially usable topics were mentioned that helped the discussion between the participants.
One notable example of GPT-3.5's usefulness was its response to a comment about the devaluation of the Italian Lira, the currency that preceded the euro, during his financial transition. In my reply, I added a second part that ChatGPT's final sentence had prompted me to think about.

[34] What do you think of my personal bias issue, with regard to the choices I made on the preference of one of your sentences over another?
[35] I didn't force this response

> **C**: *Unfortunately, we didn't realise at first that it was a scam, we realised later that for 1 euro it took two thousand lire.* [36] [37]
>
> **R**: ***That's not really how the exchange rate works, however (despite being someone who doesn't trust Europe),*** *although the euro has led to higher prices, it has also led to several benefits such as simplified cross-border payments and reduced currency conversion costs* ***(where the banks profit)*** [38]

By including additional information (in bold) about how the banks lose part of their gains from transactions between old currencies in Europe, ChatGPT's response helped to dispel the idea that the devaluation was an organized effort by central banks to harm the people and sovereign states. This has also happened in other (albeit few) cases.

After reflecting on the quality of ChatGPT's interaction and its potential for deradicalisation, I've come to the following conclusions: 1) ChatGPT is currently not able on its own to hold a conversation with individuals for a deradicalisation purpose; 2) ChatGPT doesn't have systems that disguise its AI nature (which may be a good or bad thing, depending on the situation); 3) ChatGPT can excellently handle a conversation with an individual on several different topics (based on the fact that it takes previous conversations into account); 4) ChatGPT was optimal for providing insights and ideas that extend beyond the individual.

So in my personal opinion, ChatGPT like the rest of many technologies,

> *Is the extension of our senses and capabilities, which allow us to get to a goal that we, previously without that technology, would not have been able to get to on our own.* [39]

Therefore beyond the result, it fully fulfils its role in that it helps to reduce human effort (both physically and mentally), increases the chances of better organising and enhancing work on other layers not initially assumed.

6.4 Key outcome

The data revealed that even at the social level, only one influencer, the channel administrator, was recognised. There were also a few smaller influencers, likely friends or bots of the administrator. To further support this hypothesis, their behavior towards the channel administrator was observed, and

[36] Lira is the old Italian currency, before the euro.

[37] Purtroppo non ci rendevamo conto all inizio che era una truffa l abbiamo capito dopo che per 1 euro ci volevano duemila lire.

[38] Vabbè, non funziona proprio così il cambio, comunque (pur essendo uno che non si fida dell'europa), sebbene l'euro ha comportato un aumento dei prezzi, ha anche portato a diversi vantaggi come la semplificazione dei pagamenti transfrontalieri e la riduzione dei costi di conversione delle valute (dove ci lucravano le banche)

[39] Probably a quote from Marshall McLuhan, but I could not find the source.

it was noted that about 3-4 of them frequently commented on and/or liked every sentence or post made by the administrator. This suggested that the same strategy implemented in part 2 of my work was being utilised by these smaller influencers.

But at this very time, from 2023/3/23 to 2023/4/20 (phase two and three), there was a noticeable decline in the number of accounts in the channel. The total number of users decreased from approximately $\sim$450 to a minimum of 376 (-74 peoples/-19.5%), and then go back to $\sim$430.

In my opinion, it is unlikely that the decrease in accounts on the Telegram channel is due to a reduced interest in the COVID-19 pandemic. During the February-April period, there were numerous administrative and judicial actions taken in Italy against those responsible for handling the pandemic. Moreover, as evidenced by some of the posts in the group, there is a lot of attention being paid to political figures involved in managing the pandemic. However, it should be noted that most of the posts revolve around potential banking crises and the possible collapse of the global economic system.
It can be assumed that the accounts that always commented and gave positive reactions to the administrator were only serving to increase the hype of the group to prevent losing too many people. To confirm this, the account codes of these users were saved to see if they existed during the period of initial data collection. The accounts were were presents, so I can exclude the possibility that they are recently created bots.

The existence of bots in the group and the users reduction, can be considered a phase transition, indicating that the group's engagement has decreased to a point where the administrator felt the need to use bots to maintain activity and prevent the group from disbanding. Unfortunately, I have no evidence for this, and the data does not show sufficient proof to defend my hypothesis.

However, from 2023/4/22 until 2023/5/6, the number of interactions in the group dropped dramatically, with only two comments made in about two weeks. I believe this decrease in activity is not due to my actions, but rather the subsequent holidays in Italy. From 2023/4/22 until 2023/5/1, Italy celebrated two major holidays (Liberation Day on Monday, April 25, and Labor Day on Tuesday, May 1) as well as two weekends. Totaling just under three working days in two weeks. It is likely that the administrator and other members have largely left the channel alone to enjoy the holidays. [40] I tried occasionally to write something with my accounts to get some discussion going, but I didn't have much effect. I made attempts to spark discussions with my accounts, but they were not successful. Eventually, I decided to ask the administrator (using one of my accounts named Jessica) why they were no longer sharing news on the channel. This is the (translated) answer I received:

[40] Not me unfortunately.

Hello Jessica. Look, I'm sending you an audio. Give voice mail. It'll be faster. So the group is my group, this one down here. It was not created to post news, because there are many groups like mine. So it serves little purpose. It always seemed too scattered. Then no, too much of the same news that you find everywhere, all the same news that is not encouraging. They are not so good. Indeed. So positive it doesn't convey anything, even if it is news you might like to hear. Yes, of course. But in practice people don't change after you know what changes in your life if they then take money out of your bank, if they then block your bank account, if you then don't defend yourself? So my point is that because what's happening worldwide is a financial economic problem, it's not a disease problem, it's a war problem? No, it's all financial. Why exactly are currencies jumping? New financial systems are coming in. The old system is collapsing. OK, we may not see it with the naked eye, but it is collapsing. That's it. So that's my help, that's the purpose of down. The mask is to lend a hand, to help according to the expertise I have, not according to the news taken from strangers, things, titles put down. This will happen. This will happen the other as if they are all gurus who are already predicting the future of what is to happen. I honestly don't have these faculties, otherwise I'd already be a billionaire to win. Have you won the lottery? Always, every week. So my help is what I can give according to my skills, the economic ones, as well as I give money, but I give the possibility to put one's savings. I put one's savings for those who have it, for those who don't have it you can create it from scratch if you are not at work. Opportunities to collaborate with the business I am in. However, I offer from the group this one. Many of this group are my co-workers now and also customers. So my help is that, it's not posting things. It was born like this. It was born in the days of Covip to post news and help not to go to work and refuse. Greenpeace refused the vaccine to explain how they could help in that respect. Now it's all kind of settled down in that respect. So what can fit? The Meloni government, those things there and vaccines that for now, put there in front is of little use to us because we don't know what they are up to behind the scenes. I don't know, nobody knows. No one can go and predict, to say it's just maybe news just to fill in a bit, get a few likes? No, some newspaper thing. So as long as I can't do anything yet there is nothing of who on the horizon. But what I can do on the level of real, practical physical help is to say take the money out of the banks and the post office, take the savings out of there just put where in gold. And this is mine. Because now it's a currency, you don't declare it, it doesn't make VAT-free income. I tell us my help more than that what can I do? That's what we want. We empty the banks, we don't give power to these guys who are already going to collapse. Because you shake off euro and the dollar

in a while the banks collapse. But those who have savings in there also collapse. Because even that little one has a reserve. Had a little piggy bank somewhere else in this world. It's more vital than a place that's mundane. OK, that's my help. I can only give it here in this respect. The other things I see as too banal. Too much. They waste my time. I'm always It's what zooming around with clients or co-workers? Tonight at nine I have one. For example, after the week is over, my last appointment tonight at nine. They contact me from this, from my group. Lots of people. And I'm giving it back a lot. It's better than us if we don't help. So, if we don't take the money out of the banks, we are always under the game of the enemy. That's all. OK, if you need me in this respect, I'm here. This is my main purpose. To help families get their money safe and secure. OK, that's it? A hug. Bye. Have a good night. Jessica. Thank you for contacting me. By the way, for the vote of confidence.

The statement 'Many of this group are now my colleagues and also clients' confirms what has been speculated in the past, both inherent to the accounts activity in the channel and also to the purpose of radicalisation for personal financial gain.

The administrator therefore confirms that this group (as well as other groups) are only channels with private financial ends. Similar to the person who made money with fake news in Italy (presented in the Goals chapter), the administrator uses and creates fake news to radicalise with the aim of fraud.

The initial conditions of the system and the observed social dynamics unfortunately described a difficult and hostile environment for my goals. I admit I was discouraged by the possible results from after the interaction with my bots and ChatGPT, but starting with the channel network after the deradicalisation dynamics, as shown in Figure 16 and Table 18, there was a change from the initial conditions.

Table 18: Network statistics regarding Figure 16

Settings	Data
Average degree	1.397
Avg. Weighted degree	2.478
Network diameter	4
Graph density	0.001
Modularity	0.222
N. of communities	7
Avg. Path length	2.078

While the administrator remains the main hub of the network, their centrality has decreased significantly. Since the implementation of my accounts and the dynamics they bring, the relative weight of other accounts in the network has shifted considerably. As shown in Figure 17, the grey nodes representing my accounts indicate that interactions with other accounts cover roughly $\sim 17\%$ of the entire community (represented by light green and light blue nodes).

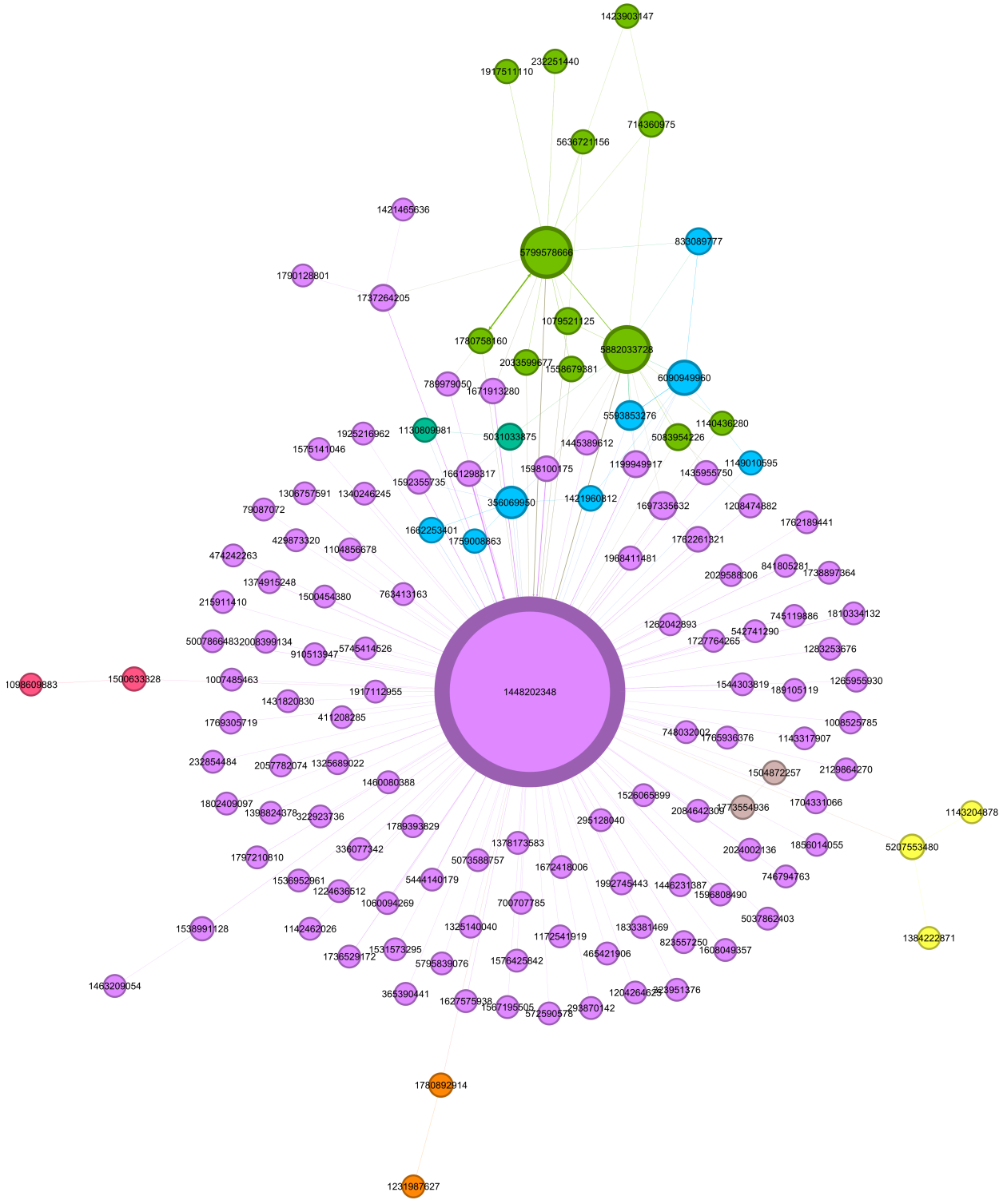

Figure 16: Channel network after dynamics

Regrettably, due to one of my accounts being banned from the channel, information such as text, replies, and other interactions are missing from the data. This prevents me from evaluating the

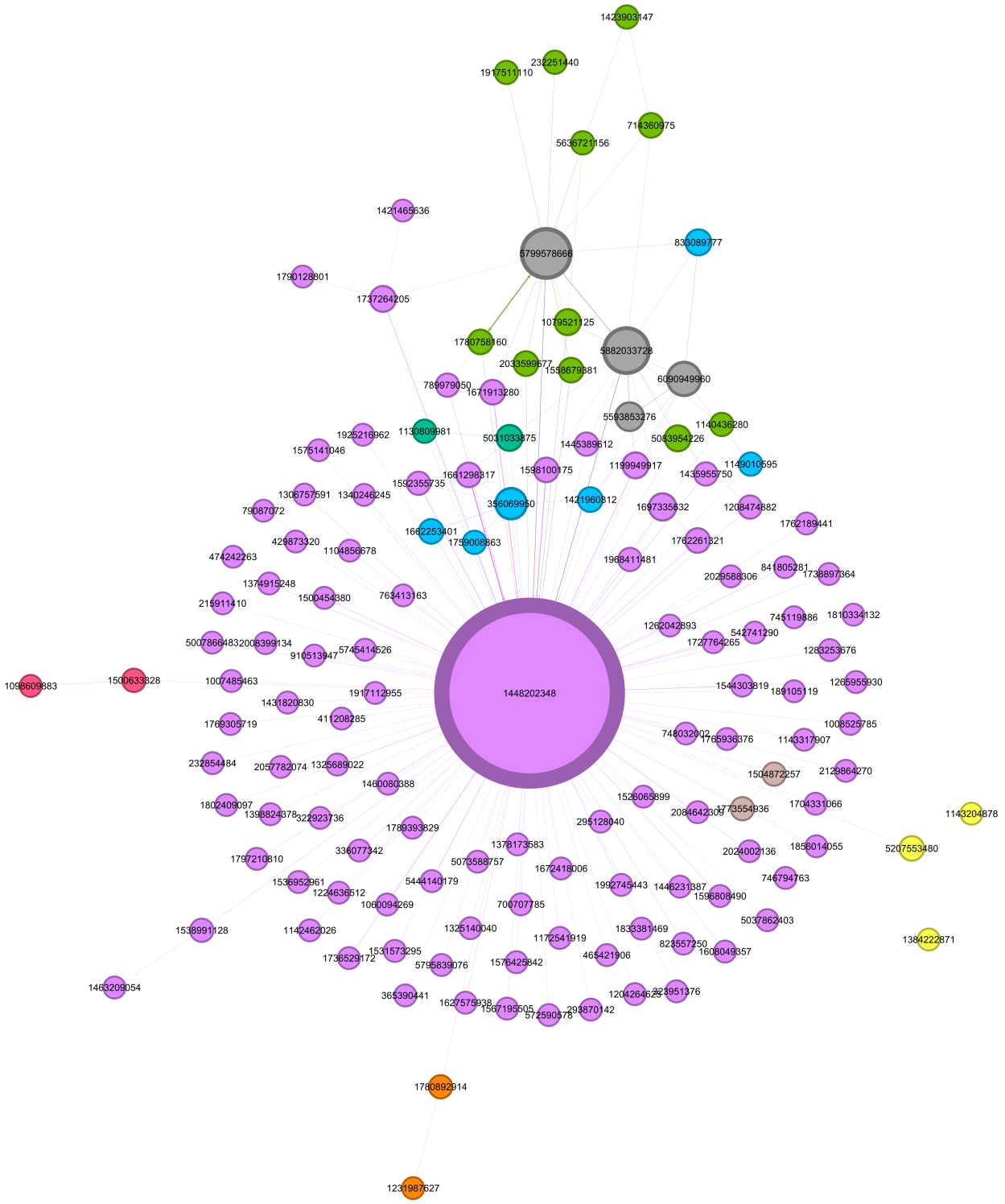

Figure 17: My accounts during dynamics

effectiveness, role, and significance of that account during the interaction. As a result, the findings

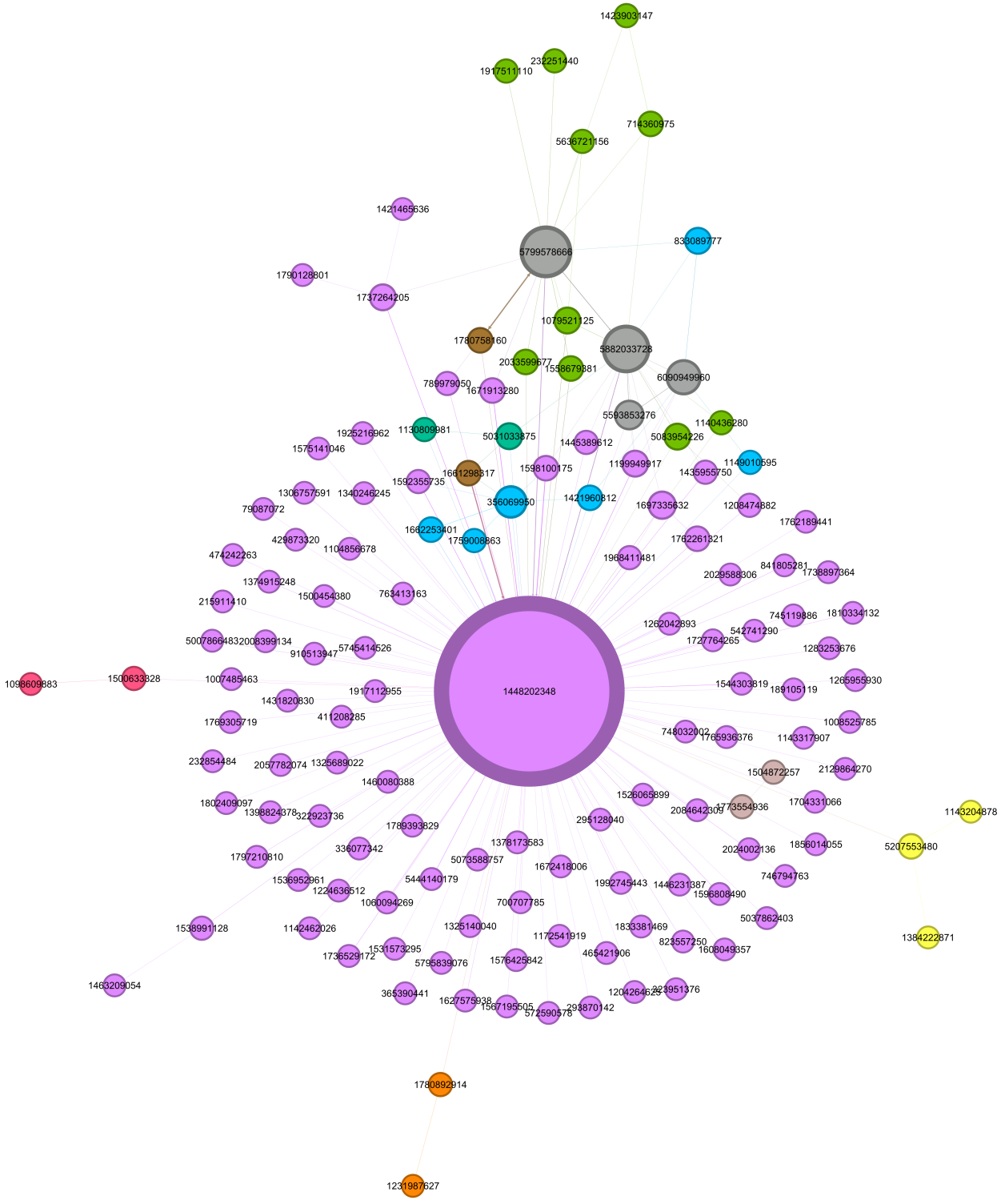

Figure 18: Accounts with violent reactions and behaviour

presented in the thesis may underestimate the true effectiveness of the interaction with the various bots.

Regarding interaction with accounts that exhibited violent reactions and behaviour, I unfortunately found several accounts that exhibited this behavior, but two in particular stood out. These accounts are represented by the dark orange nodes in Figure 18. Despite their violent tendencies, I interacted with them frequently. As you can see, a cluster of nodes was formed in the top center of the graph, where my accounts and other users engaged in discussions with greater openness and exchange of opinions. This created new, smaller communities (represented by light green, light blue, and dark green nodes) different from the violet one connected directly to the administrator, creating what I can call an *anti echos-chamber*.

From a statistical standpoint, the data in Table 19 reveals that the accounts which held influence prior to the dynamic have vanished almost entirely. The only ones remaining are the two accounts that displayed violent and extremist behavior (Sherlock and 1780758160) or those who were close to the administrator (Tesla).
The table also demonstrates a significant decrease in the administrator's importance, going from a degree of 349 to 122. These results prove that my intention to remove other influencers has been successful, as the highest degrees after the administrator belong to my accounts.
Although InfoMind's story garnered attention, 'Giovanni Pitagora' has a low degree compared to others, which is typical for an influencer who does not typically engage in discussions. Instead, it is the accounts linked and close to him that respond to the various discussions. In terms of network weight, Table 20 illustrates that my accounts' weight degree is in close proximity to that of the influencer.

Another significant finding from the results of the interaction is that there was a decrease in the number of interactions from the administrator towards the group, while there was a noticeable increase in interactions between the various accounts within the channel. This increase may have been due to the peaceful and open interventions of my accounts, which ultimately led to the creation of an 'anti-echo chamber' where various opinions were exchanged.

Regarding the text in the channel, Figure 19 displays the network of the most frequently used words (and their associated contexts) in the group. Network information is displayed in Table 21. Notably,

Table 19: Degree and Weighted degree before and after dynamics

Accounts	ID-account	D. before dynamic	D. After dynamic	Weighted degree
AdmParisi	1448202348	349	122	222
My account	5799578666	-	22	68
My account	5882033728	-	19	39
My account	6090949960	-	10	14
Tesla / Admin's likely collaborator	356069950	9	8	8
My account	5593853276	-	6	13
Common user	1697335632	-	5	13
Acc. with Violent behaviour	1780758160	-	3	33
Sherlock / Acc. with V. behaviour	1661298317	3	3	16

Table 20: Weighted degree difference

Tot. My accounts	Adm account	Difference
134	222	- 88

there are not many differences compared to the network of the initial conditions, except for the increased presence of topics related to currency and the decreased focus on foreign states.

Regarding text Entropy there were no significant changes observed: the former only decreased slightly from 4.43779167 to 4.39072301; While for the language complexity (Fog Index), unfortunately due to technical problems it was not possible to obtain the results. [41]

However, there were significant results observed in sentiment analysis.
Table 22 displays the sentiment results after the dynamic, while Table 23 shows the variation between the initial sentiment and the difference with the sentiment after the dynamic. The results reveal a reduction in negative sentiment, an increase in neutral sentiment, and an equal and opposite reduction in positive sentiment. Furthermore, the compound sentiment showed a more negative sentiment than the initial sentiment, but this result can be attributed to the increase in neutral sentiment that resulted in more balanced and less radical discussions within the group.

To support this hypothesis, I also analyzed the sentiment of my accounts throughout the dynamic phases. Table 24 illustrates the sentiment of my accounts, which is very similar to that of the initial conditions. This further confirms that the change in sentiment was driven by the channel accounts not under my control, indicating a possible phase transition in sentiment similar to that observed in the network. This suggests that there has been or is a continuing shift towards more neutral sentiment among the channel users.

The final results reveal a significant situation from the initial conditions of the system. However, it is important to determine whether this has led to a phase transition. Phase transitions occur when a certain critical threshold is exceeded, leading to a marked change in the system. In this case, from my personal perspective, it does not appear that the threshold has been crossed. The data indicates that the group has undergone a "shake-up" that could potentially cause major faults in the future.

[41] The server after several crashes, got burnt.

Table 21: Network statistics regarding Figure 19

Settings	Data
Average degree	2.133
Avg. Weighted degree	68.497
Network diameter	34
Graph density	0.003
Modularity	0.776
N. of communities	13
Avg. Path length	13.375

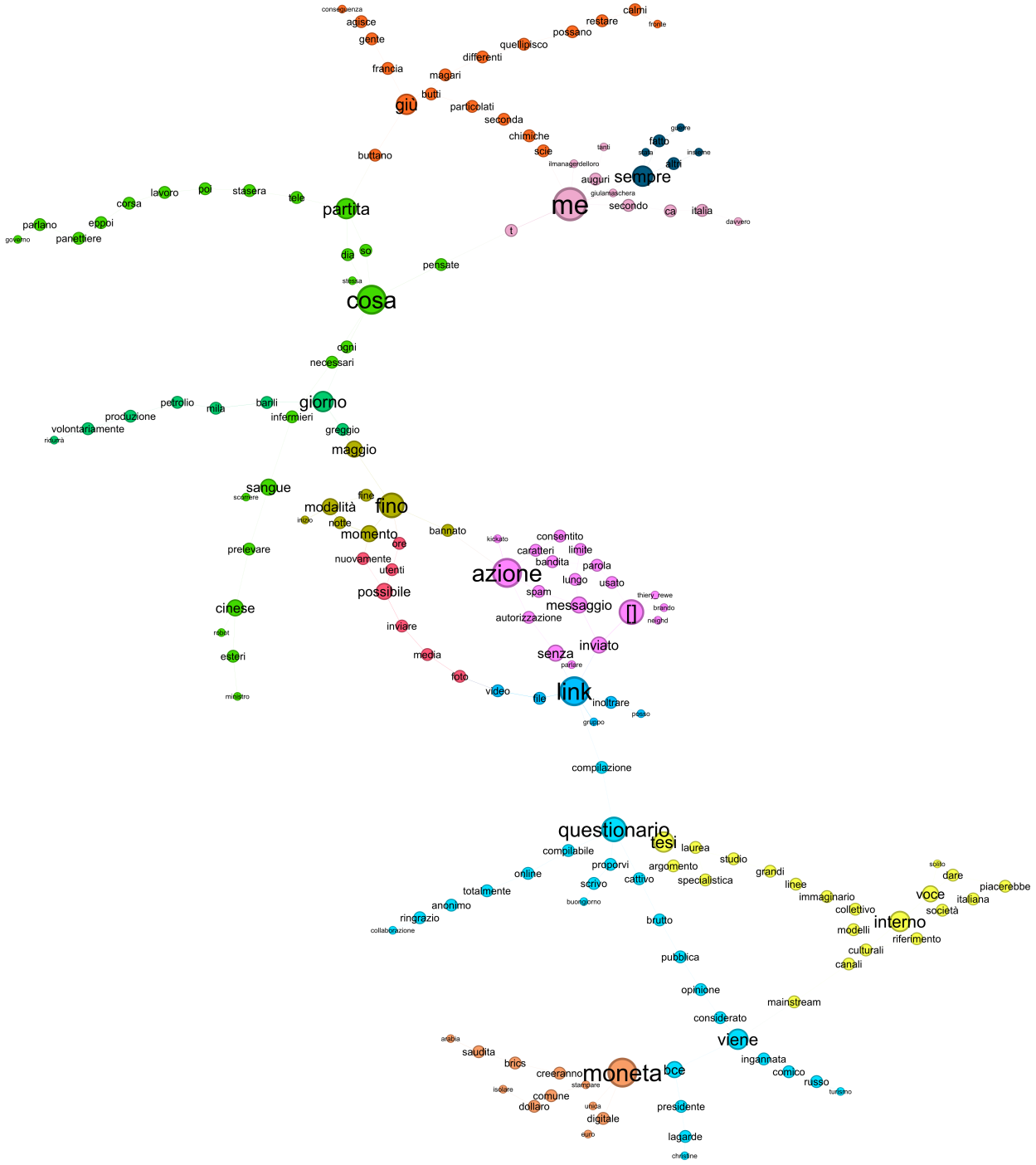

Figure 19: Influencer's network without administrator

Whether or not this will happen depends on both exogenous and endogenous events that affect the dynamics of the group.

Assuming different scenarios, it is possible to determine whether a possible event in the future can trigger a phase transition. For example, if the group administrator were to lose reputation in the

Table 22: Average text sentiment of the Telegram group

Negative	Neutral	Positive	Compound
0,078551515	0,938115152	0,104549495	0,067783434

Table 23: Average text sentiment of the Telegram group before and after dynamics

Timing	Negative	Neutral	Positive	Compound
Initial condition	0,125332436	0,724032301	0,149318977	0,481011844
After dynamics	0,078551515	0,938115152	0,104549495	0,067783434
Variation	-0,046780921	0,214082851	-0,044769482	-0,41322841

Table 24: Average text sentiment of my accounts

Negative	Neutral	Positive	Compound
0,098056	0,739376	0,140984	0,0404848

face of an arrest for financial fraud, the Telegram channel would likely shut down. However, if the group remained open, the absence of the administrator, who played a significant role (as the data showed) in sharing propagandist information and inciting violent reactions, could lead to the group's collapse or abandonment.

Similarly, the inclusion of non-violent and non-radicalized individuals, who for a long time discuss in a peaceful and non-radicalised manner on various arguments (similarly as my bots did) could lead to the *'I don't want to do this all my life'* scenario, thus creating a less radicalised group.

This scenario could trigger a departure from the current dynamic and lead to a more peaceful and productive discussion, thus creating a threshold of tolerance, where if crossed, violent and radical speech will no longer be tolerated (as it showed that it is possible to increase neutral sentiment over negative and positive sentiment). Therefore, it is necessary to analyze the possible scenarios and social consequences of the group, from its initial conditions to its current state after undergoing the deradicalisation phase.

By analyzing the social conditions and values that underlie a given subject, it is possible to calculate and assess various scenarios, including the impact of future events. To achieve this, we can start with the data from the initial conditions and measure the possible probabilistic impact on the subject in question based on the social changes brought about by the dynamics that occurred previously. This approach enables a more accurate assessment of the potential outcomes of various scenarios, as it takes into account the complex interplay of social factors and how they can influence the subject's behavior and response.

To assess the likelihood of a potential event affecting a group's trajectory, we can examine the impact of dependent variables, such as topics and related sentiment, on the group for example. This analysis begins with the data collected during the initial observation and dynamic phases, allowing for the identification of variables that may have a restrictive or amplifying effect on the group.

To construct an equation that calculates the potential impact of future events on the observed group, we can utilize the data collected and their respective variations over time as variables. By doing so, we can assess the possible outcomes of future scenarios and their impact on the group.

Although the dynamics were quite articulated and required a combination of social dynamics, artificial intelligence, and non-linear dynamics, I believe that some of our goals were achieved. Specifically, I have able to gain a certain level of authority from certain accounts, which allowed me to introduce new communication dynamics that had a moral basis but always supported by the authority of the person who saying it.

As a result, we can see a reduction in the echo chamber effect within the group. However, I'm unable to completely eliminate the mentality of "I wanna do this for all my life." This failure may be attributed to the fact that after the data analysis, the group members have shown themselves to be anti-system, in the sense that they rejected the status quo and went against the powerful forces of the moment.

In Table 25 I have taken the table of objectives in the Methodology chapter and added what I consider to be achieved and failed goals.

Table 25: Goals achieved

Objectives	Strategy	Goal
Influencers	Try to obtain more good replay (exchange of opinions about my post) from others members. I will obtain more influence and I can modify the argument direction from one to another.	Achieved
Influencer (Admin)	II cannot go against the administrator of the page (he can block me), but I can lower the value of its content and its importance by getting people to follow me.	Achieved
Social value	This reflects the 'I want to do this all my life' issue. I would try to get people to rethink their goal, showing that 1) they too can be wrong about the news they share (fake-news or with high extremist values), and 2) that it does not pay to follow certain radical values, as you do more harm to people (even those you love) than the system against them.	Not calculable and/or observed
Polarization	I will try to share highly polarising news (ecology, geopolitics) to create conditions in which people quarrel and/or drift apart, thus creating a long-term situation where people find it difficult to agree with another person even if they have a common goal. *Divide et impera*.	Achieved
New Group	At the end of the test phase, if I have an important position in the group, I will invite people to follow another group, created and managed by me. I will tell people that the information they will find in the new group is more truthful, relevant and impactful.	Failed
Collapse	By collapse, I mean the creation of a dynamic state within the group that is highly convoluted and irreconcilable, to the point where the group will fall apart because there is no longer any interest in following it and/or the leader no longer has the ability to move the masses. This, on a strategic level, can be implemented by achieving, to a small degree, all the objectives listed above.	Achieved but not calculable

I was able to achieve some of the goals I set out to accomplish.
In terms of the 'Influencers' objective, I succeeded in slightly changing the way group accounts interacted with one another. This was supported by the data, which showed that the average sentiment of messages became more neutral over time. With regard to the 'Influencer' objective, I was able to lower the weight of its content, as some accounts started to question the information shared. This was reflected in the network data, which showed a decrease in the influence of certain accounts.

Unfortunately, I was not able to achieve my objective of improving the 'Social Value' of the group. As I mentioned earlier, the group's cultural identity as anti-system made it difficult for me to influence their social values. The group, although presenting violent people has shown itself to be culturally anti-system, and this is something that is beyond my possibilities both technically, legally and morally, because in this case, I am going to affect the social values of individuals, which is different from my goal of deradicalising the most violent people in the group.

Regarding the 'polarisation' objective, I have made progress in creating an anti-echo chamber, which has resulted in more diverse opinions being expressed. As a result, it has become more difficult for individuals to agree with others who hold extremist or violent views. In addition, I noticed that some accounts (initially few in number) began to respond on my behalf, defending the points I made.

Under the 'New Group' objective, my initial approach was to initiate a discussion about the creation of a new group with individuals who held peaceful ideas. However, my attempt was met with resistance from the group administrator, who banned my two accounts as soon as I talk the idea. When I inquired about the situation with my remaining two accounts, they too were swiftly banned.

In conclusion, the 'Collapse' was not initially met. It is possible that the non-collapse was due to insufficient presence in the group, requiring more time and additional accounts to attain the desired outcome. Shortly after concluding my work, I was surprised to discover that the group had been deleted. Regrettably, I have no means of gathering information on the reasons, as my accounts were deleted for security purposes, and data retrieval from the API after deletion is not possible. While it is plausible that my actions influenced the deletion, but I lack the data to substantiate this claim, and it would be inopportune to assume there were no other contributing factors beyond my influence.

I believe that a fully deradicalisation of the group is possible, but it will require more time and accounts to apply the right practice and strategy. However, this task is challenging given the initial conditions of the group, such as the anti-system social values, that played a huge role in many dynamics. If it were a self-organised channel aiming for samples discussions on various issues, that would have been something else. But in this case, the group was created by a single person for their own economic gain. Additionally, the administrator can easily remove members who share information necessary for the group's deradicalisation, which worsens the situation.

7 Conclusion

Starting from the initial conditions of the system, it is possible to obtain very important information on how to set up various types of strategies, including deradicalisation strategies. As there are not many publications or research concerning the deradicalisation process on social networks, this work is therefore innovative, original and modern for current times.

The integration of AI technologies can bring significant benefits in implementing various strategic approaches and offer valuable insights into practical applications. The interplay of social dynamics, artificial intelligence, and non-linear dynamics can be an exceptionally useful framework for comprehending the evolution, dynamics, and potential outcomes that we encounter in our daily lives.

The complex dynamics shown in the thesis show that there is still a lot to be done from an analytic point of view, since news (and other events exogenous to the group itself) influences the group dynamics in a very relevant way.
This, however, inspires even more in the study of both the technical IT feasibility, to help both institutions and society to be in a more secure situation than before, and also in the study of these unpredictable, and therefore appealing, dynamics.

Whatever the motivations may be (by the administrator to send false information for financial purposes), the use of propaganda in this manner is unethical and has the potential to cause harm to those who are misled by the information presented. It is important to be vigilant in identifying and combating propaganda, and to seek out sources of information that are reliable, accurate, and unbiased.

7.1 Ethical considerations

An obviously important part lies in the ethical considerations of the work performed.
If my work were to use a methodology of sharing fake-news, or instigating violence, this work would never have begun, both for personal and ethical legislative reasons.
Given that my work has never tried to deceive people, but to establish a communication channel capable of supporting a radicalisation process, consisting mainly of the modification of shared values in the group, to those that are more civil, reasonable and open-minded, I believe I can consider my work satisfactory from an ethical point of view, both for research and security objectives, also for the human aspect of the work carried out.

Sharing true information, i.e. real and from official sources, cannot be considered dangerous on the ethical aspect, as it is the same action that everyone does, sharing opinions and information. On the level of deradicalisation, on the other hand, as I myself am not a 'deradicalising agent' given my

lack of qualifications on the subject, I can be openly criticised in this regard; but my expertise in sociology but also in intelligence, security and lastly in anthropology, can be considered adequate to establish between risk and threat.

7.2 Questions, limits and possibilities about contemporary research

In this chapter I would like to talk about two important topics that I wanted to address, or that I found myself facing. The first is the problem of Bias encountered with ChatGPT, while the second is my answer to the question "can we study the social sciences with the same principles as the hard sciences?".

The problem of bias encountered with ChatGPT has given me much food for thought. Although there is a vast literature on the subject, there are still many questions that the scientific community is asking itself, and indeed, is working towards a more critical but conciliatory answer at the same time. With regard to my own experience, I can say and partly justify that, my preference to choose one phrase over another stems from the fact that I have been closely analysing this community for almost a year now (I started my thesis in September), so I know the group dynamics quite deeply from that time. Being a sociologist, I know the importance of certain variables and behaviours in the group, but this does not take away from the fact that other disciplines can also become aware of these dynamics and thus increase their knowledge on the subject or group in question. The question is therefore a delicate one, if each person can understand the fundamental values and foundations of a given community, can he or she then choose which phrase best fits the circumstances? From my own personal point of view YES-MAY, but this must also be limited by a deep knowledge of the subject matter, and a deep focus (within normative limits) on the interaction with which individuals mediate through AI, which may modify group goals and values.

So the question is very open, and in part also needs to be surrounded by a profound attention of competencies and ethics. But I think the discussion can be concluded with ChatGPT's point, where according to it point of view one should "*make an effort to minimise any potential biases to promote fairness and inclusivity in your interactions*", can a human, along with all the bias issues he or she drags along, be fair and inclusive in interaction? I leave the doubt to those who are reading this thesis.

As stated in the previous chapters, there is the question of whether the *"social world can, and should be, studied according to the same principles, procedures, and philosophy as the natural sciences"* [277]. Certainly, computational methodologies give a huge boost to social research, but in my personal opinion, we are a long way from studying the social sciences as hard sciences, since the latter have the ability to predict (as was the case with Halley's comet) an event in the future. The social sciences, given their complexity (defined as the interaction of the individual parts) cannot,

at least definitively, express an absolute opinion, while it would instead be possible to express a probabilistic opinion, as stated by Ettore Majorana. [88]

But by taking into account the social conditions and values that characterize a given scenario and analyzing the data from the initial conditions as well as the social changes resulting from previous dynamics, we can calculate and evaluate the potential impact of future events with a certain degree of probability.

To illustrate my work, let's take the example of a rocket re-entering Earth's atmosphere in an uncontrolled manner. There are independent variables, such as the mass of the rocket, entry velocity, and angle, that tend to be unchanging. However, there are dependent variables, such as atmospheric density, temperature, air currents, and solar wind pressure, that change based on various factors. Scientists can probabilistically calculate the impact point of the rocket based on the independent variables alone and update the dependent variables to refine the estimate.

Similarly, my work draws on independent variables, obtained thanks to the information on the initial condition and post-dynamic conditions such as: network, topic types, social values, and sentiment, which tend to be relatively stable over time but can change gradually as in the case of the re-entry of the rocket (given the friction with the atmosphere), they are and/or can tend to be variables that are unchanging over a set period of time. I am unlikely to see the social values of this Telegram group change dramatically in a very short time, precisely because there is a high resistance to certain values.

Some may of course criticise my choice of putting 'values' in the independent variables, but I reach at this conclusion because in conspiracy theories, those who search for this type of group, as for fake news, hardly find them by chance [42], but look for them expressly, thus showing a sharing of values in the group by the person searching for them. However, there are dependent variables, such as actor roles, social norms, topic choices (distinct from topic types), and external factors that vary and change. External variables can take many forms, such as policies or mass information channels that either reinforce or erode the group's norms, or the reputation of the group as perceived by external agents, therefore, to be themselves (the dependent variables) the unpredictable ones, and which therefore require careful analysis to try to predict with certain probabilities, whether the system is going in one direction or another.

The Kuramoto model for example, a mathematical model used to study the synchronization of coupled oscillators in various fields, including social science, can be used to create a tailored model for the social group studied. By analyzing data such as network connections, sentiment, and time variables, it is possible to understand how the group may respond to certain stimuli. This can help predict how the group will react in different scenarios and inform strategies for intervention or prevention.

[42] Assumption. There is no clear evidence that people intentionally search for fake news.

I believe that the results, methodology, and technical and social possibilities presented in my thesis can greatly contribute to advancing research and addressing critical questions in the scientific community related to social research issues.

For instance, online deradicalisation processes have so far been limited to examining their ethical and technical feasibility. My thesis, however, provides a good starting point for understanding the technical and social aspects that motivate certain behaviors and for exploring the deradicalisation phase in more depth.

Similarly, in the field of security research, my thesis can provide valuable insights into the motivations behind certain behaviors, as well as contribute to the development of deradicalisation strategies and mitigation of social threat.

My thesis can pave the way for further exploration of the interaction between IT methodologies and the social sciences, across various fields. In particular, the use of computational social science methodologies can provide valuable support to sociological disciplines, enabling more accurate numerical assessments of certain social dynamics by social scientist.

In conclusion, my thesis can serve as a reference point for social and complexity research, i.e. the areas that most belong to me, starting from Ettore Majorana's initial effort to addressing the challenges of studying sociological and social disciplines using hard science methodologies. My thesis, can help social scientist to better understand complex dynamics that are challenging to evaluate numerically in contemporary sociology and inspire stimulate the complexity sciences to explore novel inquiries regarding the practical feasibility of achieving statistically precise results in the analysis of social dynamics.

[Stay Human]

8 Acknowledgements

First and foremost, I would like to thank those I cannot thank, such as some professors who are no longer with us, like Professor Vittoria Cuturi and Marina Maestrutti who both made me love sociology thanks to their human spirit and passion for the discipline.

In general, I would also like to thank those who cannot celebrate or can no longer do so.

Then I would like to express my deepest appreciation to all those who helped me and gave me much of their knowledge and passion, such as many professors, among them my supervisors, Giovanni Giuffrida, Francesco Mazzeo Rinaldi and Taha Yasseri (visiting), also other professors such as: Carlo Colloca, Guido Nicolosi, Pinella Di Gregorio, Maurizio Avola, Davide Arcidiacono, Rosalba Perrotta, Carlo Pennisi, Alessandro Pluchino, Alfredo Pulvirenti, Filippo Comito, Thierry Menissier, Olivier Zerbib, Marie-Laurence Caron-Fasan, Vincenzo Valentino Susca, Fabio La Rocca, Denis Fleurdorge and Andrea Rapisarda. A spacial thanks got also to Reviewers, Mark Coté and Joseph Downing, for taking the necessary time and effort to review the thesis. I sincerely appreciate all your valuable comments and suggestions, which helped me in improving the quality of the work.
I am also grateful to, the entire scientific community, with colleagues and researchers that I have never met in presence.

Thanks should also go to my colleagues from XXXIV to XXXVII, such as Antonio Picone and Vincenzo Miracula (CFC), Valeria Mazzeo, Elvira Celardi, Luca Gallo, Martina Fazio, Sabrina Sansone, Giulio Cantone, Ornella Occhipinti, Chiara Zappala, Federico Malizia, Filippo Comito, Bianca Rumore, Giulio Iannelli, Ruggiero Lo Sardo, Salvina Panebianco and Ilaria Cosentini.
There is also the Dublin group, such as Caludia, Daniel, Sam, Candace, Diana, Gabriella and Manran.
All colleagues, whom I have had the opportunity to meet over the years, who have made me laugh, endure my PhD better, and help me through difficult bureaucratic, personal and work-related moments.

I would be remiss if I did not mention my friends, who are "the brothers and sisters we choose". It is difficult to start with one group in particular, because they are all important, so I list them in a completely random way: Antonio Laudani, Alberto Ferreto, Leonardo Cavaletti, Giuseppe Plantania, Alessandro Maugeri, Oreste Giuffrida, Manuela Battiato, Moss Green, Antonio Sposito, Alfio Cavallaro, Alessio Chiarenza, Sergio Esposito, Priscilla Platania, Giovanni Pappalardo, Filippo Comito, Antea Platania, Ruben Crispino, Debora Magri, Leo Crispino, Hugo Chronis, Clementine Besson, Lucas Chronis, Ken Burkhard, George Veroiotis, Antonio Lombardo, Cesar Santa, Alyona

Skipa, Flavia Maccarone, Manuela del popolo and Michele Garretto. In addition there are my friends I who often meet online, like Marco, Cesare, Matteo, Peppe Dota and Sab.

In conclusion, I must thank all my family, my father who has always supported me, both financially and spiritually, and to my sister, although she doesn't know it, she has been my anchor. Also, all my cousins, uncle, aunts and various relatives in Italy, France and abroad.

And finally, the podium goes to the one who tells me that I always talk about my PhD program, who has helped me, supported me and I who hope will continue to do so for a long, long time to come: my girlfriend, my little duckling.

9 List of Acronyms

WWE ..... World Wrestling Entertainment
US ..... United state
CCP ..... Chinese Communist Party
USA ..... United states of America
UK ..... United Kingdom
IRA ..... Internet research agency
BLM ..... Black lives matter
OCEAN ..... Big Five personality traits
CA ..... Cambridge analytica
ICO ..... Information commission office
SCL ..... Strategic Communication Laboratories
UKIP ..... United Kingdom Independence Party
COPASIR ..... Parliamentary Committee for the Security of the Republic
AGCOM ..... Communications Guarantee Authority
NLP ..... Natural Language Processing
BRICS ..... Brasil-Russia-India-China-Southafrica "alliance"
SOC ..... Self-organized criticality
CB ..... Collective behavior
CI ..... Collective Intelligence
SC ..... Stochastic Resonance
IC ..... Influence Cascade
CSS ..... Computational Social Science
NLP ..... Natural language processing
SA ..... Sentiment Analysis
GFI ..... Gunning fog Index
WE ..... Word-embedding
B2B ..... Business to Business
B2C ..... Business to Customers
CD ..... Community Detection

SE . . . . . Shannon Entropy

IMF . . . . . International Monetary Fund

CCU . . . . . Counterfeit Crimes Unit

HUMINT . . . . . Human Intelligence

IMINT . . . . . Intelligence of the images

SIGINT . . . . . Intelligence of signals

MASINT . . . . . Intelligence of measurements and characteristics

TECHINT . . . . . Technical Intelligence

OSINT . . . . . Intelligence of Open Source

SOCMINT . . . . Social Media Intelligence

AI . . . . . Artificial Intelligence

ICSR . . . . . The International Centre for the Study of Radicalisation and Political Violence

AIS . . . . . Islamic Salvation Army

GIA . . . . . Armed Islamic Group

SCE . . . . . Social Critical Error

PSYOP . . . . . Psychological Operations

CMO . . . . . Context-mechanism-outcome

TS . . . . . Text simplification

CNAIPIC . . . . . National Computer Crime Centre for the Protection of Critical Infrastructures

SCE . . . . . Social Critical Error

WEF . . . . . World Economic Forum

NATO . . . . . North Atlantic Treaty Organization

CIA . . . . . Central Intelligence Agency

NCII . . . . . Non-consensual dissemination of intimate images

VPN . . . . . Virtual private network

List of Figures

List of Tables

References

[1] Nicolosi Guido. Meeting propaganda, Meeting: 8/04/2021.

[2] Everett M. Rogers and David G. Cartano. Methods of measuring opinion leadership. *Public Opinion Quarterly*, 26(3):435, 1962.

[3] Nadja Enke and Nils S Borchers. Social media influencers in strategic communication: A conceptual framework for strategic social media influencer communication. *International journal of strategic communication*, 13(4):261–277, 2019.

[4] Jan-Frederik Gräve. What kpis are key? evaluating performance metrics for social media influencers. *Social Media+ Society*, 5(3):2056305119865475, 2019.

[5] Wikipedia contributors. White monkey. `https://en.wikipedia.org/w/index.php?title=White_monkey&oldid=1100867960`, July 2022. Accessed: 16/09/22.

[6] First Last. Il wrestler usa john cena si scusa con i fan cinesi per aver definito taiwan "paese" - la repubblica, 2023.

[7] Yuliya Talmazanamerican Actor. Actor john cena apologizes to chinese audience after calling taiwan a country, 2023.

[8] The Wayrick Steves. Ukraine president volodymyr zelensky says he's russia's no. 1 target, his family second - the washington post, 2022.

[9] Yuras Karmanau. Ukraine's capital under threat as russia presses invasion | ap news, 2022.

[10] Wikipedia contributors. Volodymyr zelens'kyj. `https://it.wikipedia.org/w/index.php?title=Volodymyr_Zelens%27kyj&oldid=129390716`. Accessed: 16/09/22.

[11] Miguel del Fresno García, Alan J Daly, and Sagrario Segado Sanchez-Cabezudo. Identifying the new influences in the internet era: Social media and social network analysis., 2016.

[12] Susie Khamis, Lawrence Ang, and Raymond Welling. Self-branding,'micro-celebrity'and the rise of social media influencers. *Celebrity studies*, 8(2):191–208, 2017.

[13] Karen Freberg, Kristin Graham, Karen McGaughey, and Laura A Freberg. Who are the social media influencers? a study of public perceptions of personality. *Public relations review*, 37(1):90–92, 2011.

[14] Joseph Downing and Richard Dron. Theorising the 'security influencer': Speaking security, terror and muslims on social media during the manchester bombings. *New Media & Society*, page 146144482097178, Nov 2020.

[15] Sancheng Peng, Aimin Yang, Lihong Cao, Shui Yu, and Dongqing Xie. Social influence modeling using information theory in mobile social networks. *Information Sciences*, 379:146–159, Feb 2017.

[16] Payless. Payless prank: Social media influencers thought they were buying palessi - the washington post, 2022.

[17] Jordan Valinsky. Payless fools influencers with a fake store | cnn business, 2022.

[18] Cbs News And. The payless experiment: Retailer opened fake store, creating palessi luxury brand, to prank influencer marketers to buying discount shoes at 1800markup - cbs news, 2022.

[19] Hilke Plassmann, John O'doherty, Baba Shiv, and Antonio Rangel. Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the national academy of sciences*, 105(3):1050–1054, 2008.

[20] Paul Erdős, Alfréd Rényi, et al. On the evolution of random graphs.

[21] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.

[22] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[23] Eytan Bakshy, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11*, page 65, Hong Kong, China, 2011. ACM Press.

[24] Ilias N. Lymperopoulos and George D. Ioannou. Online social contagion modeling through the dynamics of integrate-and-fire neurons. *Information Sciences*, 320:26–61, Nov 2015.

[25] Susan Perry, Alecia Carter, Marco Smolla, Erol Akçay, Sabine Nöbel, Jacob G Foster, and Susan D Healy. Not by transmission alone: the role of invention in cultural evolution. *Philosophical Transactions of the Royal Society B*, 376(1828):20200049, 2021.

[26] Barbara K Kaye and Thomas J Johnson. Online and in the know: Uses and gratifications of the web for political information. *Journal of broadcasting & electronic media*, 46(1):54–71, 2002.

[27] Kristin English, Kaye D. Sweetser, and Monica Ancu. Youtube-ification of political talk: An examination of persuasion appeals in viral video. *American Behavioral Scientist*, 55(6):733–748, Jun 2011.

[28] First Last. Conte chiama ferragni e fedez: "aiutatemi a dire ai ragazzi che devono usare la mascherina" - huffpost italia, 2022.

[29] First Last. Conte chiama chiara ferragni e fedez: "aiutatemi a far indossare la mascherina" - photogallery - rai news, 2022.

[30] First Last. Il presidente conte chiama fedez: Ha chiesto un aiuto per sensibilizzare all'uso delle mascherine, 2022.

[31] Adpause. Fedez: "ci ha telefonato conte e ci ha chiesto un aiuto per esortare i giovani a usare la mascherina". il video - il fatto quotidiano, 2022.

[32] Fdi Tra Mennuni. Conte chiama chiara ferragni e fedez. "aiutatemi a far capire ai ragazzi che devono usare la mascherina" - la repubblica, 2022.

[33] First Last. Chiara ferragni invita il pubblico a votare per i maneskin all'eurovision, 2022.

[34] Rolling Stone Llc. I måneskin: «È un vittoria incredibile. la ferragni? ha fatto come le nostre famiglie» | rolling stone italia, 2022.

[35] Adpause. Eurovision 2021, fedez "scatena" chiara ferragni per i maneskin e lei chiede ai fan di votare la band - il fatto quotidiano, 2022.

[36] First Last. Lo zampino dei ferragnez dietro il trionfo dei maneskin, da fedez e chiara ferragni la bomba social sul televoto: «ora scateno la chiarona nazionale!» - il video - open, 2022.

[37] First Last. Ferragni contro meloni, eugenia roccella (fdi): "i dati smentiscono il racconto dell'influencer" - youtube, 2022.

[38] Fdi Tra Mennuni. Ferragni contro meloni: quanti consensi può spostare una influencer - la repubblica, 2022.

[39] First Last. "i vecchi...". e ora l'influencer insulta gli elettori della meloni - ilgiornale.it, 2022.

[40] First Last. Video: Gli insulti su instagram dell'influencer agli elettori della meloni - ilgiornale.it, 2022.

[41] First Last. Corsera – aborto ferragni contro meloni. fdi: le influencer si informino | isabella rauti, 2022.

[42] First Last. (1) transcrime su twitter: "'influencer promuove su un'app di messaggi borse contraffatte realizzate da produttori in cina che le spediscono a casa dei clienti italiani scopri altri casi studio di #fata," twitter, 2022.

[43] First Last. Il brutale consiglio del ceceno kadyrov: "putin usi armi nucleari a bassa intensità" - huffpost italia, 2022.

[44] First Last. Dopo l'ultima debacle russa il leader ceceno kadyrov chiede a mosca di usare il nucleare tattico, 2022.

[45] First Last. Il leader ceceno kadyrov manda i tre figli minorenni a combattere in ucraina - la repubblica, 2022.

[46] First Last. Kadyrov spedisce i figli minorenni al fronte e ironizza sulle sanzioni: «mi daranno il guinness dei primati?» - open, 2022.

[47] Indietrospettacoliartista Daycinemafuoricinemail Filmla. Ucraina, il leader ceceno kadyrov manda i tre figli minorenni in prima linea a combattere- corriere tv, 2022.

[48] Mario Caligiuri. Putin perde la guerra. ma solo quella dell'informazione 2022. *Formiche.net*, Mar 2022.

[49] First Last. Mosca ha ristretto l'accesso a facebook per i russi. *RAI.it*, 2022.

[50] Osservatorio Balcani E Caucaso. Il governo russo chiude la radio "eco di mosca" / russia / aree / home - osservatorio balcani e caucaso transeuropa, 2022.

[51] Agenzia Ansa. Mosca minaccia blocco wikipedia, false informazioni - tlc - ansa, 2022.

[52] F2innovation. Rt e sputnik: Ue ferma la propaganda del cremlino, 2022.

[53] Giulia Sbarigia. L'ue chiude la tv rt e sputnik: «informazione tossica» | il manifesto, 2022.

[54] Osservatorio Balcani E Caucaso. La distopia perversa di putin / russia / aree / home - osservatorio balcani e caucaso transeuropa, 2022.

[55] First Last. Russia, bbc, cbs, cnn e bloomberg sospendono le attività dei loro corrispondenti, 2022.

[56] Agenzia Ansa. Facebook toglie censura ai post contro russia, mosca reagisce - hi-tech - ansa, 2022.

[57] Clicking. Youtube blocks russian state-funded media, including rt and sputnik, around the world, 2022.

[58] Wikipedia contributors. Invasione russa dell'ucraina del 2022. `https://it.wikipedia.org/w/index.php?title=Invasione_russa_dell%27Ucraina_del_2022&oldid=129365231`. Accessed: 16/09/22.

[59] First Last. Gli "schemi" identificano i marinai morti sull'incrociatore "moskva", sebbene la federazione russa abbia annunciato la "completa evacuazione dell'equipaggio, 2022.

[60] Countryrankings By Category. • russia: number of instagram users 2022 | statista, 2022.

[61] Miranda Rossana. Il pianto degli influencer russi. non per la guerra, per l'addio a instagram 2022. *Formiche.net*, Mar 2022.

[62] Sky Tg24. Guerra in ucraina, la mappa dei territori riconquistati e tolti ai russi | sky tg24, 2022.

[63] Vi racconto come ho fatto soldi a palate spacciando bufale razziste sul web 2015, Oct 2015.

[64] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9):e2023301118, 2021.

[65] Christopher Brian Currin, Sebastián Vallejo Vera, and Ali Khaledi-Nasab. Depolarization of echo chambers by random dynamical nudge. *Scientific Reports*, 12(1):1–13, 2022.

[66] John F Scruggs. Echo chamber" approach to advocacy, 1998.

[67] Nathalie Van Raemdonck. The echo chamber of anti-vaccination conspiracies: mechanisms of radicalization on facebook and reddit. *Institute for Policy, Advocacy and Governance (IPAG) Knowledge Series, Forthcoming*, 2019.

[68] Michael Wolfowicz, David Weisburd, and Badi Hasisi. Examining the interactive effects of the filter bubble and the echo chamber on radicalization. *Journal of Experimental Criminology*, pages 1–23, 2021.

[69] Francesco Bechis Gabriele Carrer. Così la cina fa propaganda in italia, con i bot. ecco l'analisi su twitter di alkemy per formiche. *Formiche.net*, 1(1):1, 30 03 2020.

[70] webit.it. Art. 658 - procurato allarme presso l'autorità.

[71] webit.it. Art. 656 - pubblicazione o diffusione di notizie false, esagerate o tendenziose, atte a turbare l'ordine pubblico.

[72] Il campo battaglia definitivo mente-persone, Mar 2022.

[73] disinformazione laltro virus liniziativa della-farnesina, Feb 2022.

[74] Jon Roozenbeek and Sander Van der Linden. Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, 5(1):1–10, 2019.

[75] Marcus W Mayorga, Erin B Hester, Emily Helsel, Bobi Ivanov, Timothy L Sellnow, Paul Slovic, William J Burns, and Dale Frakes. Enhancing public resistance to "fake news" a review of the problem and strategic solutions. *The handbook of applied communication research*, pages 197–212, 2020.

[76] Cbs News And. Trump's call for protests gets muted reaction by supporters - cw atlanta, 2023.

[77] Gram Slattery. Far-right activists wary of 'trap' after trump calls for protests | reuters, 2023.

[78] Masi Chiara. Sparatoria-macerata-luca-trani/. *Formiche.net*, Feb 2018.

[79] Wikipedia contributors. Attentato di macerata. `https://it.wikipedia.org/w/index.php?title=Attentato_di_Macerata&oldid=129645765`. Accessed: NA-NA-NA.

[80] Il Fatto Quotidiano. Shinzo-abe-media-giapponesi-lex-premier-e-morto-un-41enne-gli-ha-sparato-durante-un-comizio-ha-agito-per-frustrazione/6654062/. *Il Fatto Quotidiano*, Jul 2022.

[81] Wikipedia contributors. Assassination of shinzo abe. `https://en.wikipedia.org/w/index.php?title=Assassination_of_Shinzo_Abe&oldid=1121165155`, November 2022. Accessed: NA-NA-NA.

[82] Vespa Stefano. Mentre mattarella interviene con decisione contro gli episodi che fanno somigliare il paese a un far west, salvini avrebbe dovuto fare un passo in più. *Formiche.net*, Jul 2018.

[83] Richard Rogers. Deplatforming: Following extreme internet celebrities to telegram and alternative social media. *European Journal of Communication*, 35(3):213–229, 2020.

[84] Charles Perrow. *Normal accidents: Living with high risk technologies*. Princeton university press, 1999.

[85] Roberto Setola, Giacomo Assenza, and Antonino Vaccaro. Complessità, intelligenza artificiale e infrastrutture critiche. *Società Italiana di Intelligence*, Jan 2021.

[86] Wikipedia contributors. Normal accidents. `https://en.wikipedia.org/w/index.php?title=Normal_Accidents&oldid=1066965965`, January 2022. Accessed: NA-NA-NA.

[87] Prof. Marco Lombardi. It's time war studies university, international conference: Information, disinformation and cybersecurity., 2022.

[88] Ettore Majorana and RN Mantegna. The value of statistical laws in physics and social sciences. In *Ettore Majorana Scientific Papers*, pages 237–260. Springer, 2006.

[89] Emma Bell Alan Bryman. Business research methods, 2nd ed.oxford: Oxford university. *Oxford University Press*, 2007.

[90] Alessandro Pluchino, Andrea Rapisarda, and Cesare Garofalo. The peter principle revisited: A computational study. *Physica A: Statistical Mechanics and its Applications*, 389(3):467–472, 2010.

[91] David MJ Lazer, Alex Pentland, Duncan J Watts, Sinan Aral, Susan Athey, Noshir Contractor, Deen Freelon, Sandra Gonzalez-Bailon, Gary King, Helen Margetts, et al. Computational social science: Obstacles and opportunities. *Science*, 369(6507):1060–1062, 2020.

[92] Moberg. Mediatization and the technologization of discourse: Exploring official discourse on the-internet and information and communications technology within the evangelical lu-theran church of finland. *New Media & Society*, page 515–531, 2018.

[93] David C Broadstock and Dayong Zhang. Social-media and intraday stock returns: The pricing power of sentiment. *Finance Research Letters*, 30:116–123, 2019.

[94] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organized criticality. *Physical review A*, 38(1):364, 1988.

[95] Julian Borger. American carnage: how trump's mob ran riot in the capitol. *The guardian*, 1(1):1, 7 Jan 2021.

[96] Giovanni Gambino and Andrea Russo. Profilazione sociale e sicurezza nazionale. *https://doi.org/10.36182/2021.05*, page 38, 2021.

[97] First Last. Cile | protesta contro l'aumento dei biglietti dei trasporti pubblici, 2023.

[98] Valentina Santarpia. Cile, proteste per il caro-biglietti in metro: è stato di emergenza - .it, 2023.

[99] The Phenomenon Of. From kuramoto to crawford: exploring the onset of synchronization in populations of coupled oscillators - sciencedirect, 2023.

[100] Machine and Not. Chemical turbulence | springerlink, 2023.

[101] Our Reliance On. Hacking online virality with dr. fil menczer & phd students (indiana university) – learning informatics lab, 2022.

[102] States. Virtual webinar: Hacking online virality | gaamac, 2022.

[103] Our Reliance On. Hacking online virality - youtube, 2022.

[104] Mr Julian Knight MP. Ico investigation into use of personal information and political influence. *information commissioner office*, page 18, 2020.

[105] Intelligence senate USA. Background to "assessing russian activities and intentions in recent us elections": The analytic process and cyber incident attribution. *USA Senate*, 2017.

[106] Gabriele Carrer. Prigozhin, perché gli usa affondano lo "chef" di putin. *Formiche.net*, page 2, 2020.

[107] 1789U.S. Department of the Treasury Seal of the U.S. Department of the Treasury. Treasury increases pressure on russian financier. *https://home.treasury.gov/*, page 2, 2020.

[108] U.S. Department of Justice. Report on the investigation into russian interference in the 2016 presidential election. *Department of justice*, 2019.

[109] Gambino Giovanni and Andrea Russo. Profilazione sociale e sicurezza nazionale. *SOCINT Press*, 2021.

[110] Peter Hermann Marc Fisher, John Woodrow Cox. Pizzagate: From rumor, to hashtag, to gunfire in d.c. *washingtonpost.com*, page 2, 2016.

[111] Il post. "qanon", la teoria del complotto più diffusa nella politica americana. *ilpost.com*, page 2, 2020.

[112] Andrea Tironi. La profilazione social influenza le scelte elettorali: come funziona e quali difese. *agendadigitale.eu*, page 2, 2019.

[113] Andrew Kaczynski. On twitter, michael flynn interacted with alt-right, made controversial comments on muslims, shared fake news. *CNN*, page 2, 2016.

[114] Michael D. Shear. Obama warned trump about hiring flynn, officials say. *nytimes.com*, page 2, 2017.

[115] U.S. Department of Justice. Report on the investigation into russian interference in the 2016 presidential election. *U.S. Department of Justice*, page 448, 2019.

[116] fastcompany.com. The strange afterlife of cambridge analytica and the mysterious fate of its data. *fastcompany.com*, page 3, 2019.

[117] James Melaugh. Cambridge analytica: how did it turn clicks into votes? *theguardian.com*, page 2, 2018.

[118] Nazzareno Tirino. *Cambridge Analytica. Il potere segreto, la gestione del consenso e la fine della propaganda.*, volume 1. Libellula Edizioni, 2019.

[119] theguardian.com. Brittany kaiser's evidence to commons culture committee - summary. *theguardian.com*, page 1, 2018.

[120] carolecadwalla. 1075416385617059841. *twitter.com*, page 1, 2018.

[121] Wu Youyou, Michal Kosinski, and David Stillwell. Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4):1036–1040, 2015.

[122] Askanews source. Usa, commissione capitol hill 2022. *Formiche.net*, Jun 2022.

[123] Lanzavecchia Otto. Svolta ue, ecco il codice anti fake news (con multe alle aziende) 2022. *Formiche.net*, Jun 2022.

[124] First Last. (1) gazetaru su twitter: "evgeny prigozhin ha affermato che la russia "è intervenuta, interferisce e interferirà" alle elezioni statunitensi. "attentamente, accuratamente, chirurgicamente e a modo nostro, come sappiamo. durante le nostre operazioni mirate, rimuoveremo contemporaneamente sia i reni che il fegato ", afferma l'uomo d'affari." / twitter, 2022.

[125] Bechis Francesco. Anticorpi contro le fake news (russe e non). parla moles 2022. *Formiche.net*, Jun 2022.

[126] Alessio Frugiuele. Relazione copasir e l'infowarfare russa. l'analisi di mario caligiuri. *CyberSecurity Italia*, Aug 2022.

[127] Authors Are Their. Nato review - hybrid war - hybrid response?, 2022.

[128] Ilias N. Lymperopoulos. Dynamic response and transfer function of social systems: A neuro-inspired model of collective human activity patterns. *Neural Networks*, 94:125–140, Oct 2017.

[129] Jef Huysmans. What's in an act? on security speech acts and little security nothings. *Security dialogue*, 42(4-5):371–383, 2011.

[130] Fabrizio Battistelli and Maria Grazia Galantino. Dangers, risks and threats: An alternative conceptualization to the catch-all concept of risk. *Current Sociology*, 67(1):64–78, 2019.

[131] More Than. Twitter purged more than 70,000 accounts affiliated with qanon following capitol riot - the washington post, 2022.

[132] Jacopo Lenti, Kyriaki Kalimeri, André Panisson, Daniela Paolotti, Michele Tizzani, Yelena Mejova, and Michele Starnini. Global misinformation spillovers in the online vaccination debate before and during covid-19, 2022.

[133] First Last. Materiali, 2022.

[134] https://www.cnbc.com/2020/11/12/amazon-sues-influencers-for-allegedly-marketing-cou Accessed: 2022-11-1.

[135] https://press.aboutamazon.com/news-releases/news-release-details/amazon-counterfeit-crimes-unit-reaches-settlement-influencers. September 30, 2021 at 9:16 AM EDT».

[136] Appearing To Be. Amazon counterfeit crimes unit reaches settlement with influencers who ran social media counterfeiting scheme, permanently banning them from amazon's store and securing financial payments to be donated to support anti-counterfeiting awareness | amazon.com, inc. - press room, 2022.

[137] Annie Palmer. Amazon settles with influencers who allegedly ran counterfeit scheme, 2022.

[138] Bimbo twitta dall account del comando strategico usa caos mondo 2021, Mar 2021.

[139] Wikipedia contributors. Fat-finger error. https://en.wikipedia.org/w/index.php?title=Fat-finger_error&oldid=1098958370, July 2022. Accessed: NA-NA-NA.

[140] The Mfdr Radar. Eu sst confirms re-entry of space object cz-5b – eu sst, 2022.

[141] First Last. (1) u.s. space command su twitter: "#usspacecom can confirm the people's republic of china long march 5b #cz5b rocket re-entered the atmosphere over the south-central pacific ocean at 4:01am mdt/10:01 utc on 11/4. for details on the uncontrolled reentry's impact location, we once again refer you to the #prc." / twitter, 2022.

[142] Scotti Vincenzo. Disinformazione russa e come combatterla, il commento di scotti 2022. *Formiche.net*, Jun 2022.

[143] Bechis Francesco. Disinformazione-russa-giacomelli-agcom/. *Formiche.net*, Jul 2022.

[144] First Last. Livelli di istruzione e partecipazione alla formazione, 2022.

[145] Alberto Pagani. *Manuale di intelligence e servizi segreti. Antologia per principianti, politici e militari, civili e gente comune*. Rubbettino, 2019. Google-Books-ID: ymc5ygEACAAJ.

[146] Luigi Garofalo. Fake news su agenzia entrate. si crede ai cybercriminali e non allo stato. *CyberSecurity Italia*, Jul 2022.

[147] Lorenzo Santucci. Siamo pronti alla triade del free speech kanye-musk-trump? - formiche.net, 2022.

[148] Maura Reynolds. Fiona hill: 'elon musk is transmitting a message for putin' - politico, 2022.

[149] Lorenzo Santucci. Il legame tra elon musk e la cina non fa dormire washington - formiche.net, 2022.

[150] Lauren Speranza. *Contro i germi dell'autoritarismo tecnologico*. formiche.it, 2022.

[151] Camera Dei Deputati. Misure per la prevenzione della radicalizzazione e dell'estremismo violento di matrice jihadista, 2022.

[152] Camera Dei Deputati. Misure per la prevenzione della radicalizzazione e dell'estremismo violento di matrice jihadista, 2022.

[153] Daniela Ronco and Alvise Sbraccia. Prison de-radicalization strategies, programmes and risk assessment tools in europe.

[154] ICSR. Prisons and terrorism radicalisation and de-radicalisation in 15 countries, 2010.

[155] Bertjan Doosje, Fathali M Moghaddam, Arie W Kruglanski, Arjan De Wolf, Liesbeth Mann, and Allard R Feddes. Terrorism, radicalization and de-radicalization. *Current Opinion in Psychology*, 11:79–84, 2016.

[156] Daylight Sopravvive Ai. Cloud per la community di intelligence degli stati uniti | aws, 2023.

[157] Nadav Maman. What happens when ai falls into the wrong hands?, 2023.

[158] Pranshu Vermajuly. Robots trained on ai exhibited racist and sexist behavior - the washington post, 2023.

[159] Andrey Dmitriev, Victor Dmitriev, and Stepan Balybin. Self-organized criticality on twitter: Phenomenological theory and empirical investigation based on data analysis results. *Complexity*, 2019:e8750643, Dec 2019.

[160] Ettore Majorana and RN Mantegna. The value of statistical laws in physics and social sciences. In *Ettore Majorana Scientific Papers*, pages 237–260. Springer, 2006.

[161] Nuno David. Validating simulations. In *Simulating social complexity*, pages 135–171. Springer, 2013.

[162] Alessandro Vespignani, Huaiyu Tian, Christopher Dye, James O. Lloyd-Smith, Rosalind M. Eggo, Munik Shrestha, Samuel V. Scarpino, Bernardo Gutierrez, Moritz U. G. Kraemer, Joseph Wu, Kathy Leung, and Gabriel M. Leung. Modelling covid-19. *Nature Reviews Physics*, 2(66):279–281, Jun 2020.

[163] Bruno Latour. On actor-network theory: A few clarifications. *Soziale welt*, pages 369–381, 1996.

[164] Alessandro Pluchino. *La firma della complessità: una passeggiata al margine del caos*. Malcor D', 2015.

[165] Wikipedia contributors. Vilfredo pareto. `https://it.wikipedia.org/w/index.php?title=Vilfredo_Pareto&oldid=127854949`. Accessed: NA-NA-NA.

[166] First Last. Prove di "nuovo mondo" | ispi, 2022.

[167] Ulrich Beck and Walter Privitera. *La società del rischio: verso una seconda modernità*. Carocci, 2005.

[168] https://it.wikipedia.org/wiki/La_societ{à}_del_rischio._Verso_una_seconda_modernit{à}. Accessed: 2022-10-6.

[169] Ulrich Beck. Critical theory of world risk society: a cosmopolitan vision. *Constellations*, 16(1):3–22, 2009.

[170] Ulrich Beck. *World at risk*. Polity, 2009.

[171] Moberg. Mediatization and the technologization of discourse: Exploring official discourse on the-internet and information and communications technology within the evangelical lu-theran church of finland. *New Media & Society*, page 515–531, 2018.

[172] Lei Guo and Maxwell McCombs. Toward the third level of agenda setting theory: A network agenda setting model. *Annual convention of the Association for Education in Journalism*, 2011.

[173] L Bode. Political news in the news feed: Learning politics from social media. *Mass Communication and Society*, 19:24–48, 2016.

[174] W Russell Neuman, Lauren Guggenheim, S Mo Jang, and Soo Young Bae. The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication*, 64(2):193–214, 2014.

[175] Jessica T Feezell. Agenda setting through social media: The importance of incidental news exposure and social filtering in the digital era. *Political Research Quarterly*, 71(2):482–494, 2018.

[176] Haenlein M Kaplan, A. Users of the world, unite! the challenges and opportunities of social media. *Business Horizons*, 53:59–68, 2010.

[177] Huang L Castronovo C. Social media in an alternative marketing communication model. *Journal of Marketing Development and Competitiveness*, 6:117–131, 2012.

[178] L Huang. Social contagion effects in experiential information exchange on bulletin board systems. *Journal of Marketing Management*, 26:197–212, 2010.

[179] Dale A Newell, R. Meeting the climate change challenge (mc3): the role of the internet in climate change re-search dissemination and knowledge mobilization. *Environmental Communication*, 9:208–227, 2015.

[180] Salomone S.. Murphy, G. Using social media to facilitate knowledge transfer in complex engineering environments: a primer for educators. *European Journal of Engineering Education*, 38:70–84, 2013.

[181] Francesco Mazzeo Rinaldi, Andrea Russo, and Giovanni Giuffrida. Information balance between newspapers and social networks. *CARMA 2020: 3rd International Conference on Advanced Research Methods and Analytics*, 2020.

[182] Carrer Gabriele. Russia-interferenze-italia-copasir/. *Formiche.net*, Aug 2022.

[183] Murray Gell-Mann. What are complex systems?, 2022.

[184] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organized criticality: An explanation of the 1/f noise. *Physical review letters*, 59(4):381, 1987.

[185] Per Bak and Maya Paczuski. Complexity, contingency, and criticality. *Proceedings of the National Academy of Sciences*, 92(15):6689–6696, 1995.

[186] RF Smalley Jr, Donald Lawson Turcotte, and Sara A Solla. A renormalization group approach to the stick-slip behavior of faults. *Journal of Geophysical Research: Solid Earth*, 90(B2):1894–1900, 1985.

[187] Andrey Dmitriev and Victor Dmitriev. Identification of self-organized critical state on twitter based on the retweets' time series analysis. *Complexity*, 2021, 2021.

[188] Klaus Linkenkaer-Hansen, Vadim V Nikouline, J Matias Palva, and Risto J Ilmoniemi. Long-range temporal correlations and scaling behavior in human brain oscillations. *Journal of Neuroscience*, 21(4):1370–1377, 2001.

[189] John M Beggs and Dietmar Plenz. Neuronal avalanches in neocortical circuits. *Journal of neuroscience*, 23(35):11167–11177, 2003.

[190] Dante R Chialvo. Critical brain networks. *Physica A: Statistical Mechanics and its Applications*, 340(4):756–765, 2004.

[191] Deborah M Gordon. The ecology of collective behavior. *PLoS biology*, 12(3):e1001805, 2014.

[192] David A Locher and David A Locher. *Collective behavior*. Prentice Hall Upper Saddle River, NJ, 2002.

[193] David L Miller. *Introduction to collective behavior and collective action*. Waveland Press, 2013.

[194] Le Bon Gustave. The crowd: A study of the popular mind, 1896.

[195] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Reviews of modern physics*, 81(2):591, 2009.

[196] Ralph H Turner, Lewis M Killian, et al. *Collective behavior*, volume 3. Prentice-Hall Englewood Cliffs, NJ, 1957.

[197] Floyd Henry Allport. *Social psychology*. Boston, Houghton, 1924.

[198] Neal Elgar Miller and John Dollard. Social learning and imitation, 1941. *New Haven*, 1941.

[199] Neil J Smelser. Theory of collective behavior (1962). *New York*, 1971.

[200] Jaap Van Ginneken. *Collective behavior and public opinion: Rapid shifts in opinion and communication*. Routledge, 2003.

[201] Sandro M Reia, André C Amado, and José F Fontanari. Agent-based models of collective intelligence. *Physics of life reviews*, 31:320–331, 2019.

[202] Pierre Lévy. *L'intelligence collective: pour une anthropologie du cyberespace*. La découverte, 2013.

[203] Saraswathy Shamini Gunasekaran, Mohd Sharifuddin Ahmad, Alicia Tang, and Salama A Mostafa. The collective intelligence concept: A literature review from the behavioral and cognitive perspective. In *2016 2nd International Symposium on Agent, Multi-Agent Systems and Robotics (ISAMSR)*, pages 154–159. IEEE, 2016.

[204] Steven Johnson. *La nuova scienza dei sistemi emergenti: dalle colonie di insetti al cervello umano, dalle città ai videogame e all'economia, dai movimenti di protesta ai network*. Garzanti, 2004.

[205] Kevin Kelly. *Out of control. La nuova biologia delle macchine, dei sistemi sociali e del mondo dell'economia*. Apogeo Editore, 1996.

[206] `https://en.wikipedia.org/wiki/Stochastic_resonance`. Accessed: 2022-10-10.

[207] Catherine Rouvas-Nicolis and Gregoire Nicolis. Stochastic resonance. *Scholarpedia*, 2(11):1474, 2007.

[208] Mark D McDonnell and Derek Abbott. What is stochastic resonance? definitions, misconceptions, debates, and its relevance to biology. *PLoS computational biology*, 5(5):e1000348, 2009.

[209] Roberto Benzi, Giorgio Parisi, Alfonso Sutera, and Angelo Vulpiani. Stochastic resonance in climatic change. *Tellus*, 34(1):10–16, 1982.

[210] C Nicolis. Stochastic aspects of climatic transitions–response to a periodic forcing. *Tellus*, 34(1):1–9, 1982.

[211] Luca Gammaitoni, Peter Hänggi, Peter Jung, and Fabio Marchesoni. Stochastic resonance: a remarkable idea that changed our perception of noise. *The European Physical Journal B*, 69(1):1–3, 2009.

[212] Dmitry V Dylov and Jason W Fleischer. Nonlinear self-filtering of noisy images via dynamical stochastic resonance. *Nature Photonics*, 4(5):323–328, 2010.

[213] Mitchell M Waldrop. *Complexity: The emerging science at the edge of order and chaos*. Simon and Schuster, 1993.

[214] Raymond A Eve, Raymond A Eve, Sara Horsfall, and Mary E Lee. *Chaos, complexity, and sociology: Myths, models, and theories*. Sage, 1997.

[215] Anthony Giddens. *Central problems in social theory: Action, structure, and contradiction in social analysis*, volume 241. Univ of California Press, 1979.

[216] Wikipedia contributors. Social complexity. https://en.wikipedia.org/w/index.php?title=Social_complexity&oldid=1070209687, February 2022. Accessed: NA-NA-NA.

[217] Mohammad Karim Saberi, Alireza Isfandyari-Moghaddam, and Sedigheh Mohamadesmaeil. Web citations analysis of the jasss: The first ten years. *Journal of Artificial Societies and Social Simulation*, 14(4):22, 2011.

[218] Martin Nowak and Roger Highfield. *Supercooperators: Altruism, evolution, and why we need each other to succeed*. Simon and Schuster, 2011.

[219] Hang Ye, Fei Tan, Mei Ding, Yongmin Jia, and Yefeng Chen. Sympathy and punishment: evolution of cooperation in public goods game. *Journal of Artificial Societies and Social Simulation*, 14(4):20, 2011.

[220] Mark Mason. Complexity theory and the philosophy of education. *Educational philosophy and theory*, 40(1):4–18, 2008.

[221] Brian Castellani. *The defiance of global commitment: a complex social psychology*. Routledge, 2018.

[222] Susanne Lohmann. The dynamics of informational cascades: The monday demonstrations in leipzig, east germany, 1989–91. *World politics*, 47(1):42–101, 1994.

[223] Graeme Chesters and Ian Welsh. *Complexity and social movements: Multitudes at the edge of chaos*. Routledge, 2006.

[224] Brian Castellani, Michael Ball, Kenneth Carvalho, et al. Addressing the us financial/housing crisis: Pareto, schelling and social mobility. Technical report, Citeseer, 2011.

[225] Pierre Demeulenaere. Analytical sociology and social mechanisms, 2011.

[226] Levent Yilmaz. Toward multi-level, multi-theoretical model portfolios for scientific enterprise workforce dynamics, 2011.

[227] Robert Jervis. System effects: Complexity in political and social life, 1998.

[228] Heike Egner and Marén Schorch. *Learning from Calamities?* Taylor and Francis, Hoboken, 2014.

[229] Brian Castellani, Rajeev Rajaram, J Galen Buckwalter, Michael Ball, and Frederic Hafferty. *Place and health as complex systems: A case study and empirical test*. Springer, 2015.

[230] Loet Leydesdorff. *The knowledge-based economy: Modeled, measured, simulated*. Universal-Publishers, 2006.

[231] D Lane, D Pumain, and SE Leeuw. van der, and west, g. eds.(2009). *Complexity Perspectives in Innovation and Social Change*, pages 1–14, 2009.

[232] Peter Stewart. Complexity theories, social theory, and the question of social complexity. *Philosophy of the social sciences*, 31(3):323–360, 2001.

[233] Consensus Knowledge, T Lewis, Kathleen Carley, Katyia Sycara, Market Segmentation, A Multi-Agent Model, and Toshizumi Ohta. Thursday july 5, 2001.

[234] Kathleen Carley, Jeffrey Reminga, and Natasha Kamneva. Destabilizing terrorist networks. pittsburgh, 2003.

[235] Loet Leydesdorff. *The challenge of scientometrics: The development, measurement, and self-organization of scientific communications*. Universal-Publishers, 2001.

[236] Vladimir Dimitrov and Robert Woog. Studying social complexity: From soft to virtual systems methodology. *Complex Systems*, 11(6):501, 1997.

[237] First Last. Xi gela putin sulla guerra a kiev: "la cina vuole un mondo stabile" - la repubblica, 2022.

[238] Wordpress. L'india al forum di samarcanda. motivi per (non) preoccuparsi - formiche.net, 2022.

[239] Wordpress. Per xi vale più il vertice di samarcanda dell'incontro con putin - formiche.net, 2022.

[240] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444, 2001.

[241] Nanna Bonde Thylstrup, Daniela Agostinho, Annie Ring, Catherine D'Ignazio, and Kristin Veel. *Uncertain Archives: Critical Keywords for Big Data*. MIT Press, 2021.

[242] Andrew T Fiore and Judith S Donath. Homophily in online dating: when do you like someone like yourself? In *CHI'05 extended abstracts on Human factors in computing systems*, pages 1371–1374, 2005.

[243] Paul F Lazarsfeld, Robert K Merton, et al. Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society*, 18(1):18–66, 1954.

[244] Marc Lynch. After the arab spring: How the media trashed the transitions. *Journal of Democracy*, 26(4):90–99, 2015.

[245] James G Webster. The duality of media: A structurational theory of public attention. *Communication theory*, 21(1):43–66, 2011.

[246] Abhishek Samantray and Paolo Pin. Credibility of climate change denial in social media. *palgrave communications*, 5(1):1–8, 2019.

[247] Paul DiMaggio, John Evans, and Bethany Bryson. Have american's social attitudes become more polarized? *American journal of Sociology*, 102(3):690–755, 1996.

[248] Delia Baldassarri and Andrew Gelman. Partisans without constraint: Political polarization and trends in american public opinion. *American Journal of Sociology*, 114(2):408–446, 2008.

[249] Morris P Fiorina, Samuel J Abrams, et al. Political polarization in the american public. *ANNUAL REVIEW OF POLITICAL SCIENCE-PALO ALTO-*, 11:563, 2008.

[250] Alan I Abramowitz and Kyle L Saunders. Is polarization a myth? *The Journal of Politics*, 70(2):542–555, 2008.

[251] Joseph Bafumi and Robert Y Shapiro. A new partisan voter. *The journal of politics*, 71(1):1–24, 2009.

[252] Jennifer McCoy, Tahmina Rahman, and Murat Somer. Polarization and the global crisis of democracy: Common patterns, dynamics, and pernicious consequences for democratic polities. *American Behavioral Scientist*, 62(1):16–42, 2018.

[253] P Sol Hart, Sedona Chinn, and Stuart Soroka. Politicization and polarization in covid-19 news coverage. *Science Communication*, 42(5):679–697, 2020.

[254] Klaas J Beniers and Robert Dur. Politicians' motivation, political culture, and electoral competition. *International Tax and Public Finance*, 14(1):29–54, 2007.

[255] Norman J Ornstein. Why it's even worse than it looks: Parliamentary parties in the american constitutional system. *Drake L. Rev.*, 61:1117, 2012.

[256] Danielle M Thomsen. Ideological moderates won't run: How party fit matters for partisan polarization in congress. *The Journal of Politics*, 76(3):786–797, 2014.

[257] Shigeo Hirano, James M Snyder Jr, and Michael M Ting. Distributive politics with primaries. *The Journal of Politics*, 71(4):1467–1480, 2009.

[258] Stephen P Nicholson. Polarizing cues. *American journal of political science*, 56(1):52–66, 2012.

[259] Pablo Barberá, John T Jost, Jonathan Nagler, Joshua A Tucker, and Richard Bonneau. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science*, 26(10):1531–1542, 2015.

[260] David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, et al. Computational social science. *Science*, 323(5915):721–723, 2009.

[261] Jonathan Bright. Big social science": doing big data in the social sciences. *The Sage handbook of online research methods*, pages 125–139, 2017.

[262] First Last. C'era una volta la 'bestia'. ora salvini crolla sui social anche se spende di più. e perde più follower di tutti - la repubblica, 2022.

[263] First Last. La bestia di salvini si nutre di sentimenti negativi. l'analisi di ottomila post dal 2015 a oggi - info data, 2022.

[264] First Last. Renzi denunciava profili fake e bufale, mentre la sua bestia creava "account falsi" e pagava 260mila dollari per un software israeliano in grado di influenzare il voto - il fatto quotidiano, 2022.

[265] First Last. Il tweet contro i media di crosetto che annuncia la prima querela: "È l'unico metodo che capiscono" - la repubblica, 2022.

[266] First Last. Di maio, repubblica: "contro di lui un tweet-bombing da account fake usa". ma il software smentisce: "tutti italiani in carne e ossa" - il fatto quotidiano, 2022.

[267] First Last. Barcellona, bartomeu arrestato nell'operazione "barcagate" - la repubblica, 2022.

[268] First Last. Psg, un "esercito digitale" contro i nemici: nel mirino anche mbappé e neymar - il fatto quotidiano, 2022.

[269] Andrea Russo, Antonio Picone, Vincenzo Miracula, Giovanni Giuffrida, and Francesco Mazzeo Rinaldi. Entropy-rate as prediction method for newspapers and information diffusion, 2022.

[270] Andrea Russo and Davide Coco. Quantify how space mission influence geopolitical dynamics? a security and social policy approach. *arXiv preprint arXiv:2301.03538*, 2023.

[271] Andrea Russo. Organised firestorm as strategy for business cyber-attacks. *arXiv preprint arXiv:2301.01518*, 2023.

[272] Julia Day. Nike: 'no guarantee on child labour'. *The Guardian*, 2001.

[273] Julian Oliver. Learning the lessons of brent spar saga. *Politico*, 1995.

[274] Jrgen Pfeffer, Thomas Zorbach, and Kathleen M Carley. Understanding online firestorms: Negative word-of-mouth dynamics in social media networks. *Journal of Marketing Communications*, 20(1-2):117–128, 2014.

[275] Media Monkey. Twitter users not lovin' mcdonald's. *The Guardian*, 2012.

[276] Angela Giuffrida Kate Connolly and Jon Henley. Chaos in germany and italy after suspension of oxford vaccine. *The Guardian*, 2021.

[277] Emma Bell Alan Bryman. Business research methods, 2nd ed.oxford: Oxford university. *Oxford University Press*, 2007.

[278] F. Mazzeo Rinaldi, Giovanni Giuffrida, and T. Negrete. Real-time monitoring and evaluation-emerging news as predictive process using big data-based approach, 2017.

[279] Kalle Nuortimo, Erkki Karvonen, and Janne Härkönen. Establishing social media firestorm scale via large dataset media analytics. *Journal of Marketing Analytics*, pages 1–10, 2020.

[280] Nele Hansen, Ann-Kristin Kupfer, and Thorsten Hennig-Thurau. Brand crises in the digital age: The short-and long-term effects of social media firestorms on consumers and brands. *International Journal of Research in Marketing*, 35(4):557–574, 2018.

[281] Heavy Bombing From. Firestorm | english meaning - cambridge dictionary, 2023.

[282] Gabriele Carrer and Francesco Bechis. Così la cina fa propaganda in italia, con i bot. ecco l'analisi su twitter di alkemy per formiche. *Formichiere.it*, page 1, 2020.

[283] Raffaele R. Riverso. Barcellona, arrestato l'ex presidente Bartomeu, March 2021.

[284] Áron Kiss and Gábor Simonovits. Identifying the bandwagon effect in two-round elections. *Public Choice*, 160:327–344, 2014.

[285] Maureen Farrell. High speed trading fueled twitter flash crash. *CNN Business*, 2013.

[286] Daniel D. Cdproject hacked, gwent source code leaked. *eip.gg*, 2021.

[287] Cristina Criddle. Cyberpunk 2077 makers cd projekt hit by ransomware hack. *bbc.com*, 2021.

[288] Jason Schreier. Cd projekt ransomware hack severely disrupts work on cyberpunk updates. *bloomberg.com*, 2021.

[289] Jan-Jaap Van Eerten, Bertjan Doosje, Elly Konijn, BA de Graaf, Mariëlle de Goede, et al. Developing a social media response to radicalization: The role of counter-narratives in prevention of radicalization and de-radicalization, 2017.

[290] Richard Rogers. *Digital methods*. MIT press, 2013.

[291] Michael Kenney. *Appendix: The Method of Ethnographic Network Analysis With Stephen Coulthart and Dominick Wright*, page 239–251. Structural Analysis in the Social Sciences. Cambridge University Press, 2018.

[292] First Last. Green pass falsi: chiusi più di 30 canali telegram - corcom, 2023.

[293] Plus. Falsi green pass su telegram soddisfatti o rimborsati: la gdf stronca la truffa - il sole 24 ore, 2023.

[294] Escluso Dalla Nazionale. Green pass falsi, chiusi 32 canali telegram: 4 indagati, ci sono anche due minori. perquisizioni e sequestri, acquirenti nei guai - la repubblica, 2023.

[295] Securityopenlab. Green pass falsi a 250 euro chiusi altri due canali telegram, 2023.

[296] The Css-Loader. More than 60 telegram channels blocked in germany - newspaper, 2023.

[297] The Sponsor. They. Glossary of relevant terms & acronyms, 2022.

[298] Christopher Paul, Colin P Clarke, Bonnie L Triezenberg, David Manheim, and Bradley Wilson. Improving c2 and situational awareness for operations in and through the information environment. Technical report, RAND NATIONAL DEFENSE RESEARCH INST SANTA MONICA CA SANTA MONICA United States, 2018.

[299] Colin F Jackson. Information is not a weapons system. *Journal of Strategic Studies*, 39(5-6):820–846, 2016.

[300] First Last. Università, alla sapienza sotto attacco il nuovo corso 'gender studies': "hanno aperto la scuola per clown" - la repubblica, 2022.

[301] Valentina Santarpia. Corso gender alla sapienza, attacco degli «haters» su telegram- corriere.it, 2022.

[302] First Last. Blacklist sapienza | centro infosapienza, 2022.

[303] First Last. Attacco no vax all'hub vaccinale nell'esselunga rubattino a milano: perquisito un 61enne, in casa volantini contro i vaccini - la repubblica, 2022.

[304] First Last. Milano, no vax tenta di dare fuoco a un hub vaccinale: il filmato che lo incastra, 2022.

[305] Ray Pawson and Nick Tilley. Caring communities, paradigm polemics, design debates. *Evaluation*, 4(1):73–90, 1998.

[306] Nick Tilley and R Pawson. Realistic evaluation: an overview. In *founding conference of the Danish Evaluation Society*, volume 8, 2000.

[307] Gustave Le Bon. *The crowd: A study of the popular mind*. Courier Corporation, 2002.

[308] Andrea Russo, Vincenzo Miracula, and Antonio Picone. Network analysis on political election; populist vs social emergent behaviour. *arXiv preprint arXiv:2301.05668*, 2023.

[309] Giosué Lo Bosco, Giovanni Pilato, and Daniele Schicchi. A neural network model for the evaluation of text complexity in italian language: a representation point of view. *Procedia computer science*, 145:464–470, 2018.

[310] OECD. Inchiesta sulle competenze degli adulti primi risultati., 2013.

[311] Taha Yasseri, András Kornai, and János Kertész. A practical approach to language complexity: a wikipedia case study. *PloS one*, 7(11):e48386, 2012.

[312] Nokia Bell Labs. A mathematical theory of communication. *The Bell System Technical Journal*, 27, 1948.

[313] C. E. SHANNON. A mathematical theory of communication. *Board of Trustees of the University of Illinois*, 1049.

[314] Wikipedia contributors. Sentiment analysis. `https://en.wikipedia.org/w/index.php?title=Sentiment_analysis&oldid=1116037424`, October 2022. Accessed: NA-NA-NA.

[315] Giordano Domenico. L'algoritmo anticipa il sentiment reale delle urne i numeri di arcadia 2022. *Formiche.net*, Jun 2022.

[316] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623, 2021.

[317] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[318] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, et al. Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103:102274, 2023.

[319] Pier Luca Lanzi and Daniele Loiacono. Chatgpt and other large language models as evolutionary engines for online interactive collaborative game design, 2023.

[320] Sean Teebagy, Lauren Colwell, Emma Wood, Antonio Yaghy, and Misha Faustina. Improved performance of chatgpt-4 on the okap exam: A comparative study with chatgpt-3.5. *medRxiv*, pages 2023–04, 2023.

[321] Adrian Rauchfleisch and Jonas Kaiser. The false positive problem of automatic bot detection in social science research. *PloS one*, 15(10):e0241045, 2020.

[322] Vittoria Elliott. The facebook papers reveal staggering failures in the global south - rest of world, 2023.

[323] The Italian Sa. Intelligenza artificiale: il garante blocca chatgpt. raccolta illecita di... - garante privacy, 2023.

[324] Cecilia Rodriguez. Chatgpt ban: Which countries will follow italy in blocking ai giant, 2023.

[325] Man To Prove. Gpt-4 developer livestream - youtube, 2023.

[326] Elisa Longo. *Spin,neuroni e individui: mondi complessi a confronto*. PhD thesis, University of Catania, 2021.

[327] Yejin Bang, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, et al. A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. *arXiv preprint arXiv:2302.04023*, 2023.

[328] End-To-End Encryption Giving. Far right extremism on telegram: A brief overview - european eye on radicalization, 2023.

[329] Mehvish. Telegram channel vs group: Which one should you use, 2023.

[330] Silvia Semenzin and Lucia Bainotti. The use of telegram for non-consensual dissemination of intimate images: Gendered affordances and the construction of masculinities. *Social Media+ Society*, 6(4):2056305120984453, 2020.

[331] Alessandro Caliandro and Alessandro Gandini. *Qualitative research in digital environments: A research toolkit*. Taylor & Francis, 2016.

[332] Maurizio Teli, Francesco Pisanu, David Hakken, et al. The internet as a library-of-people: For a cyberethnography of online groups, 2007.

[333] Francesco Pisanu and Maurizio Teli. Ethnography in a brave new world: exploring research in cyberspace1.

[334] Alessandro Caliandro. Digital methods for ethnography: Analytical concepts for ethnographers exploring social media environments. *Journal of contemporary ethnography*, 47(5):551–578, 2018.

[335] Johan Alvehus and Lucia Crevani. Micro-ethnography: towards an approach for attending to the multimodality of leadership. *Journal of Change Management*, 22(3):231–251, 2022.

[336] Neil Sutherland. Investigating leadership ethnographically: Opportunities and potentialities. *Leadership*, 14(3):263–290, 2018.

[337] Kathleen M Blee. Ethnographies of the far right. *Journal of contemporary ethnography*, 36(2):119–128, 2007.

[338] Mun'im Sirry. Muslim student radicalism and self-deradicalization in indonesia. *Islam and Christian–Muslim Relations*, 31(2):241–260, 2020.

[339] Gastón Julián Gil. Identities and moralities in social networks. a digital ethnography of running in contemporary society. *Qualitative Research in Sport, Exercise and Health*, 14(4):530–544, 2022.

[340] Christine Hine. *Ethnography for the internet: Embedded, embodied and everyday*. Taylor & Francis, 2015.

[341] Crystal Abidin and Gabriele de Seta. Doing digital ethnography: Private messages from the field. *Journal of Digital Social Research*, 2(1):1–19, 2020.

[342] Johanna Moisander, Elina Närvänen, and Anu Valtonen. Interpretive marketing research: using ethnography in strategic market development, 2020.

[343] Andreas Bjerre-Nielsen and Kristoffer Lind Glavind. Ethnographic data in the age of big data: How to compare and combine. *Big Data & Society*, 9(1):20539517211069893, 2022.

[344] Ganga S Dhanesh and Gaelle Duthler. Relationship management through social media influencers: Effects of followers' awareness of paid endorsement. *Public Relations Review*, 45(3):101765, 2019.

[345] Ivar John Erdal. Repurposing of content in multi-platform news production: Towards a typology of cross-media journalism. *Journalism Practice*, 3(2):178–195, 2009.

[346] Tuğrulcan Elmas, Rebekah Overdorf, and Karl Aberer. Misleading repurposing on twitter. *arXiv preprint arXiv:2010.10600*, 2020.

[347] Alessandro Bessi, Mauro Coletto, George Alexandru Davidescu, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. Science vs conspiracy: Collective narratives in the age of misinformation. *PloS one*, 10(2):e0118093, 2015.

[348] George Friedman. Intelligence and love, Nov 2021.

[349] `https://en.wikipedia.org/wiki/Juan_Pujol_Garc{í}a#Honours`. Accessed: 2022-11-25.

[350] Vecchioni Domenico. Garbo, la spia che rese possibile lo sbarco in normandia - sistema di informazione per la sicurezza della repubblica, 2022.

[351] Agent Garbo | MI5 The Security Service. Introduction to agent garbo, 2022.