Cyber Intelligence and Social Media Analytics: 
Current Research Trends and Challenges

Serena Tardelli1∗, Marco Avvenuti2, Guglielmo Cola1, Stefano Cresci1, Tiziano Fagni1, 
Margherita Gambini1,2, Lorenzo Mannocci1,3, Michele Mazza1, Caterina Senette1 and 
Maurizio Tesconi1

1 Institute of Informatics and Telematics, National Research Council (IIT-CNR), Pisa, Italy – name.surname@iit.cnr.it
2 Department of Information Engineering, University of Pisa, Italy – marco.avvenuti@unipi.it
3 Department of Computer Science, University of Pisa, Italy

Abstract

Online Social Networks (OSNs) are a rich source of data for Cyber Security and Cyber Intelligence applications, as they can reveal valuable insights into users’ behaviors, preferences, and opinions. Analyzing OSN data poses significant challenges, such as dealing with misinformation campaigns, protecting users' privacy, and extracting relevant information from large and heterogeneous datasets. The Cyber Intelligence (CI) unit of the IIT-CNR has been conducting cutting-edge research on these topics, using state-of-the-art techniques from artificial intelligence, machine learning, natural language processing, and computer vision. In this paper, we present some of the main activities of the CI group and the technologies we have developed and applied to various CI areas. In addition, we present our involvement in projects that leverage artificial intelligence technologies for the development and implementation of Cyber Security techniques and systems based on social media and online social networks.

Keywords

cyber intelligence, artificial intelligence, machine learning, deep learning, social media intelligence

1. Introduction

The Internet has become a dominant platform for communication, work and entertainment in the modern world. However, this also exposes a lot of personal and sensitive information to the public, which can be used for surveillance and prevention purposes in various domains of interest (such as terrorism, crime, etc.). Social networks are especially popular among people who use them to interact and share their views on diverse topics. The massive amount of data generated by these interactions, often referred to as "big data", demands sophisticated processing and analysis techniques to produce intelligence information that can support effective decision-making for specific areas of interest. To create advanced descriptive and predictive models, state-of-the-art machine learning and artificial intelligence (deep learning) techniques are often employed, resulting in highly accurate software solutions that can assist system users.

The Cyber Intelligence (CI) unit aims to develop innovative solutions for the analysis of social media data, in order to extract useful information for various purposes, such as security, prevention, and moderation. The CI unit applies advanced techniques of machine learning and artificial intelligence, especially deep learning, to process and analyze the big data generated by users on social networks. The unit focuses on several research topics related to social media analysis, including: botnet and fake news detection; content analysis for hate speech detection and extremist account identification; moderation intervention, evaluation, and planning; coordinated behavior and conspiracy theory diffusion analysis; metrics development to monitor the "health" of social ecosystems.

1.1. Objectives

The research activities of the CI unit are mainly focused on Cyber Intelligence and Social Media Analytics. These interdisciplinary fields combine methods and techniques from computer science, data science, social sciences, and security studies. Our group is interested in exploring various aspects of these fields, such as:

• Data collection from diverse sources, such as the Web and social media platforms.

• Analysis of large datasets ("big data") to extract useful insights and investigate the dynamics and patterns of online behaviors, such as the interactions and influence among different actors (e.g., individuals, groups, organizations) in various domains (e.g., politics, health, security).

• Development of novel algorithms, tools, and models that can describe or predict patterns in the
data, using state-of-the-art techniques from machine learning and deep learning.

- Creation of advanced and complex data visualization interfaces.
- Understanding the opportunities and challenges of using online data for intelligence purposes, such as situational awareness, threat detection and prevention, decision support, and strategic communication.

The group engages in these activities with two main objectives: (i) produce scientific outcomes that contribute to the advancement of knowledge in this field of study; (ii) enhance and update their own expertise and competencies. Below we present a more comprehensive description of each area of activity.

2. Research activities

The Cyber Intelligence (CI) research group of the IIT-CNR has been working on Cyber Security for years, focusing on research topics related to Social Media Analytics and Cyber Intelligence. The group has developed and refined skills in data collection from the most popular Social Media platforms, using both crawling techniques through native Web services and scraping techniques when official services are not available to access the data of interest. The collected data is stored and analyzed with big data technologies and exploited by applying advanced artificial intelligence techniques such as Machine Learning and Deep Learning, to create predictive or descriptive models that can support or automate specific tasks. We provide a list of some of the most significant research activities that the CI unit has conducted or is conducting as follows.

Social sensing for emergency management systems. This line of research focuses on leveraging human sensing and AI for emergency management, especially in the aftermath of disasters. The aim is to develop systems that collect social crisis data from sources such as Twitter [1], and use AI tools to enrich them with information about the damage, location, and needs of the affected people. Additional adoption of geoparsing models can help link textual mentions of places to their geographical coordinates [2, 3]. The goal is to design decision support systems that can monitor and manage catastrophic events (natural or man-made) and help authorities in the early stages of the event [4, 5, 6, 7, 8].

Detection of malicious automated accounts. Information or influence operations (IOs) on Social Media have been frequently carried out on social media to mislead and manipulate users. IOs can take various shapes, target individuals or online groups, and have a variety of goals. In this line of research, we investigate, analyze, and characterize online misbehavior in its many forms, including fake accounts [9, 10], colluding users (e.g., paid trolls) [11], and automation (e.g., social bots) [12, 13, 14, 15]. Using Machine Learning, Deep Learning, and Social Network Analysis techniques, we develop and implement cutting-edge tools and models able to detect these strategies and mitigate the influence of malicious actors who disseminate and amplify harmful information.

Analysis and detection of coordinated behavior. In the age of information warfare, for IOs to be successful, they must reach and influence a large number of users, regardless of their specific goals and methods of deception. To achieve a substantial outreach, influence and impact, campaigns often require large-scale and coordinated efforts on online social media platforms. As such, it is essential to identify patterns of coordinated activity, regardless of the intent, and authenticity. Indeed, coordinated behavior can be neutral or authentic (e.g., protesters, activists, etc.), as well as inauthentic and/or potentially harmful. As a result, we shifted the focus from single accounts to coordinated and synchronized behavior [16, 17]. We experiment with network science methods to develop novel strategies for detecting and measuring coordinated online behavior in various IOs, such as important political events [18]. We also investigate the temporal evolution of complex coordination patterns between users, how coordination evolves over time and how users adapt their behavior to changing circumstances [19, 20].

Analysis and detection of information disorder. Information disorder is a term that encompasses various forms of misleading, inaccurate, or false information that are intentionally or unintentionally spread online. It can have serious consequences for individuals, communities, and societies, such as undermining trust in democratic institutions, fueling polarization and hate speech, and endangering public health and safety. In this research area, we study information disorder in its many forms, such as misinformation, disinformation [21, 22, 23], fake news [24], malinformation, infodemic [25, 26, 27, 28], or propaganda [29, 30], depending on the source, intent, and impact of the information.

Analysis and detection of online financial and cryptocurrency discussions. Our research investigates the online ecosystem related to cryptocurrencies and financial markets, with a focus on detecting and analyzing manipulation and fraud attempts. We leveraged a range
of methods and data sources, such as social media, price data, and blockchain transactions, to study, explore, and detect different phenomena, such as: (i) online cryptocurrency manipulation (e.g., pump-and-dump, thefts, etc.) by malicious actors who seek to profit from the volatility and anonymity of the market [31, 10]; (ii) financial spam to influence the market or scam unsuspecting users and other fraudulent practices that exploit the popularity of certain companies or topics to promote less important or dubious ones [32, 33, 34, 14]. Our research aims to contribute to the understanding of the challenges, dynamics, and impacts of these phenomena, as well as to develop techniques, tools, and solutions for their detection and prevention.

Deep fake detection. Deep fake is a term that refers to the use of artificial intelligence to create realistic but fake images, videos, text, or audio of people or events. Deep fake technology can be used for multiple purposes, such as entertainment, and education. However, it also poses serious challenges for society, such as undermining trust in information sources, violating privacy and consent, and facilitating misinformation and manipulation. Therefore, it is important to develop models to mitigate and prevent potential abuse of deep fake content. As CI unit, we study and implement novel strategies to detect deepfake multimedia content, such as images, videos, and texts. Since the text generative models are increasing both in number and accuracy in resembling a human-written text, we investigate the optimal approach (in terms of data availability and training time) to detect texts written by all typologies of generative techniques, either old (e.g., RNN, Markov Chains) or new (e.g., GPT2, GPT3, GPT4, and ChatGPT), with a focus on deepfake texts written for social media [35, 36, 37].

Online extremist content detection. Any online content that promotes or incites violence, hatred, discrimination or radicalization based on ideological, religious, political or ethnic grounds can have a negative impact on individuals and societies, as it can foster intolerance, polarization and radicalization. Removing and prevent online extremism content while respecting human rights and freedom of expression is a complex and multifaceted challenge that needs a collaborative and holistic approach. As CI unit, we focus on various aspects in this area. For instance, we are interested in studying metrics for the identification of radicalization pathways, extremist users[29], and texts containing violent and hateful language (such as racial, political, etc.) [38, 39]. In addition, we focus on political polarization, examining how users’ political orientation (political leaning) and opinion (stance detection) vary according to the most salient topics in the country’s political agenda [40, 41]. Conspiracy theories can be also a part of online extremist content. Online conspiracy theories are claims that challenge the official or mainstream narratives of events or phenomena. They often involve elaborate plots, hidden agendas, secret societies, or powerful elites. They can have serious consequences for individuals and society (e.g., spreading misinformation, eroding trust, inciting violence, undermining democracy). As such, our research also examines how conspiracy theories spread on social media platforms, focusing on how to detect them and the users who propagate them.

Content moderation. Content moderation is the process of monitoring and regulating online content created and shared by users on social media platforms. Content moderation can help prevent the spread of harmful or illegal content, such as hate speech, violence, misinformation, spam, etc. However, content moderation also poses some challenges and risks, such as infringing on users’ freedom of expression, and privacy, as well as exposing moderators to psychological harm. Therefore, there is a need for a set of strategies and practices that aim to reduce the negative impacts of content moderation on both users and moderators. We survey and experiment with multiple strategies (i.e., interventions) to evaluate the effects and effectiveness of moderation interventions on social media platforms (e.g., Reddit, Twitter), both at the platform level and at the individual user level. We analyse user reactions to moderation interventions, focusing on the characteristics that might influence user reactions to interventions (e.g., user’s personality, political leaning), thus providing new knowledge and tools for mitigating widespread issues in online platforms [42, 43, 44].

3. Research Projects

The CI unit has been working on several research projects in the past years, covering different aspects of computational social science, web science, social media analysis, and cyber intelligence. In this section, we briefly describe some of the main ongoing research projects and their objectives and outcomes.

SERICS (DETERRENCE). The DETERRENCE project is part of the SERICS Foundation - Security and Rights in CyberSpace (www.serics.eu/). SERICS is funded - under the National Recovery and Resilience Plan, supported by the European Union - NextGenerationEU. The Foundation includes 10 Spokes.

Our activity is expressed within Spoke 2, Disinformation and Fake News through the project called DETERRENCE - DEcision supporT SystEm foR cybeR intelli-GENCE coordinated by the CI unit.
The traditional analysis of this data is based on the re-
comparison, reuse and integration of big data, methods 
human, social and economic sciences. It allows for easy 
research in multiple fields, such as mathematics, AI, and 
national infrastructures and opens new avenues for 
of social life. SoBigData builds on several established 
cation of social data mining to study multiple aspects 
that enables ethical and scientific exploration and appli-
cation, and analysis functionalities.

INTERROGATE  The INTERROGATE project (artifi-
cial INtelligence Text Enrichment foR impROving bG 
dAtA procEssing) is funded as part of the High Training 
projects promoted by the Italian Fondo per lo sviluppo e 
la coesione and the Tuscany Region. The project is based 
on the premise that in today’s world, every company 
has access to an enormous volume of data (Big Data). 
The traditional analysis of this data is based on the rela-
tional model, storing data in tables (structured data). 
However, today only about 20% of the data available 
for the companies is in the form of structured data, while 
the remaining 80% is unstructured and usually available as 
free text. One way to benefit from these huge amounts 
of textual data is to use text mining techniques to ex-
tract value from the data. The aim of the INTERROGATE 
project, coordinated by the CI unit, is to define a Big 
Data architecture, based on open-source solutions, that 
allows complex Text Mining models to be applied to large 
amounts of data in a scalable way. These models, based 
on the most advanced AI techniques (deep learning), will 
enrich textual resources with new structured information, 
in order to enable novel and powerful search, aggrega-
tion, and analysis functionalities.

SoBigData++ Research Infrastructure. SoBig-
Data++ is a European project funded under the Horizon 
2020 Framework Programme, which is the largest 
research and innovation program in the history of the 
European Union. The project’s goal is to create a “Social 
Mining & Big Data” ecosystem: a research infrastructure 
that enables ethical and scientific exploration and appli-
cation of social data mining to study multiple aspects 
of social life. SoBigData builds on several established 
national infrastructures and opens new avenues for 
research in multiple fields, such as mathematics, AI, and 
human, social and economic sciences. It allows for easy 
comparison, reuse and integration of big data, methods 
and services, creating an interdisciplinary research 
community. In this project, CI is mainly involved as 
leader of the Social Media Observatory, which aims to 
develop a set of tools to facilitate listening campaigns on 
social media as well as the interpretation of retrieved 
data. Developed tools include libraries to ease real-time 
data collection and the analysis of information diffusion 
on Twitter [45]. This project has led to the creation of 
two ongoing spin-off projects: SoBigData PPP and 
SoBigData.it.

SoBigData PPP. The SoBigData RI Preparatory Phase 
Project is a SoBigData++ spinoff with the goal of mov-
ing the RI forward from the simple awareness of ethical 
and legal challenges in social mining to the development 
of concrete tools that operationalize ethics with value-
sensitive design, incorporating values and norms for pri-
vacy protection, fairness, transparency, and pluralism.

SoBigData.it. The SoBigData.it project aims at 
strengthening the technological, scientific, and ethical 
 aspects of the Italian RI for Social Mining and Big Data 
Analytics. The goal is to enhance interdisciplinary and 
innovative research on the multiple aspects of social com-
plexity by combining data and model-driven approach. 
The CI unit’s main contribution is to investigate specific 
societal topics through data science, with a particular 
emphasis on analyzing Societal Debates and Misinforma-
tion across diverse domains, such as politics, health, and 
finance.

4. Conclusions

Cyber and social media intelligence is a vital but diffi-
cult domain in the information disorder era. It needs a 
comprehensive and diverse approach that considers the 
technical, social, and ethical dimensions of social media 
data. It also needs a constant change and innovation to 
match the changing social media environment. This re-
search field has a wide scope and relies mainly on public 
data that are enhanced with indicators of various aspects 
such as coordination, polarization, propaganda, etc. As 
such, the main ethical risk in this research is the poten-
tial deanonymization of the datasets, which could expose 
users’ sensitive information. To mitigate this risk, data 
protection and privacy-preserving techniques must be 
adopted to ensure that data is used and shared responsi-
ably and ethically.

The CI unit’s mission is to develop and deliver cutting-
edge cyber intelligence tools, solutions and systems for 
decision-making and research. By applying network sci-
cence, artificial intelligence, machine learning, and deep 
learning, the unit aims to identify and mitigate the next
generation of online ecosystem disruption and harmful phenomena.

Acknowledgments

We thank for the support by project SERICS (PE00000014) under the NRRP MUR program funded by the EU – NGEU; by the European Union – Horizon 2020 Program under the scheme "INFRAIA-01-2018-2019 – Integrating Activities for Advanced Communities", Grant Agreement n. 871042, "SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics"; and by project DESIRE 2.0 (DissEmination of Scientific REsults) funded by IIT-CNR.

References


[38] F. Del Vigna, A. Cimino, F. Dell’Orletta, M. Petrocchi, M. Tesconi. Hate me, hate me not: Hate speech detection on facebook, 2017.


