Social Media Intelligence as a Tool for Conducting Intelligence Activities

Antonio Teti(✉)

University "G. d'Annunzio" Chieti-Pescara, 66100 Chieti, Italy
antonio.teti@unich.it

Abstract. The information security of an organization must be based on the ability to guarantee the confidentiality and integrity of information relating to intellectual property, production processes, technological processes, business plans and in particular the data held regarding potential threats. In such a context, the massive and indiscriminate use of social media by company employees represents a huge vulnerability for information security. Accordingly, social espionage can be defined as an attempt to gain a competitive advantage by acquiring all publicly and semi-publicly available information in social media. In this regard, Social Media Intelligence activities, which are essentially based on the use of platforms capable of searching, collecting, filtering and assembling a considerable amount of useful information from social media, have taken on a decisive role. In this paper, we aim to analyze the phenomenon of social espionage and focus on two relevant ways of conducting it in the context of social media intelligence, briefly reviewing the main characteristics of both of them. The first one is the use of Social Media Intelligence (SOCMINT) platforms to capture social data streams that can be employed to learn useful information, for example, about the company where someone is employed. The second one is related to fake social profiles suitably constructed to capture the information held in different social conversations.

Keywords: Social Media Intelligence · SOCMINT · Web Intelligence

1 Introduction

An organization's information can be divided into auditable information released by the company and non-auditable information that can be acquired online, regardless of the company's wishes. This information may be collected directly by making search queries, creating alerts and filters, and interacting with competitors and customer communities (e.g., following Twitter profiles, subscribing to a forum, liking a competitor's Facebook page), or indirectly through various social media aggregation.

In this regard, the information security of an organization must be based on the ability to guarantee the confidentiality and integrity of information relating to intellectual property, production processes, technological processes, business plans and in particular the data held regarding potential threats competitive. Consequently, in this particular scenario, the protection of trade secrets assumes the connotation of the vital element

T. Guarda et al. (Eds.): ARTIIS 2023, CCIS 1937, pp. 281–291, 2024.
https://doi.org/10.1007/978-3-031-48930-3_21

on which to build the future of the company. The ability to access the most valuable information of a company or organization, as we have been able to understand, represents a fundamental objective for anyone who wants to understand what are the vulnerabilities, needs, trends, set objectives and expectations of their competitors.

When the productive activities of a company insist on several geographically dislocated locations even in different countries, the risk of conducting internal espionage activities increases considerably. Even the assignment to external companies (outsourcing) of some company services (e.g. IT services, cybersecurity, training, etc.) can represent a serious problem for the confidentiality of company data. The massive and indiscriminate use of social media by company employees represents a huge vulnerability for information security. The ignorance of the population on the correct use of social media also plays a role. Every day and incessantly, information of all kinds is entered on the Internet, and in particular on social media, from the more private personal ones to those relating to the work that takes place within one's workplace, without excluding those relating to one's friends and co-workers. Conversations conducted on Facebook, Twitter, Instagram, including those apparently restricted to a WhatsApp group, remain private only if a user decides not to spread them.

In such a context, social espionage can be defined as an attempt to gain a competitive advantage by acquiring all publicly and semi-publicly available information in social media. It should be noted that social espionage is one of the methodologies that can be used to conduct IT espionage activities, which are based on the use of security hackers and malicious software (Trojan horse, Spyware, Rootkit and Backdoor) that allow, often through the use of social engineering methodologies, to intercept the vulnerabilities of a computer system to access it and obtain the desired information [1]. It should be noted that social espionage does not assume the connotation of theft of information, given that it is publicly available (or semi-publicly) on social media.

In Fig. 1, it is possible to deduce the trends that tend to promote social espionage.

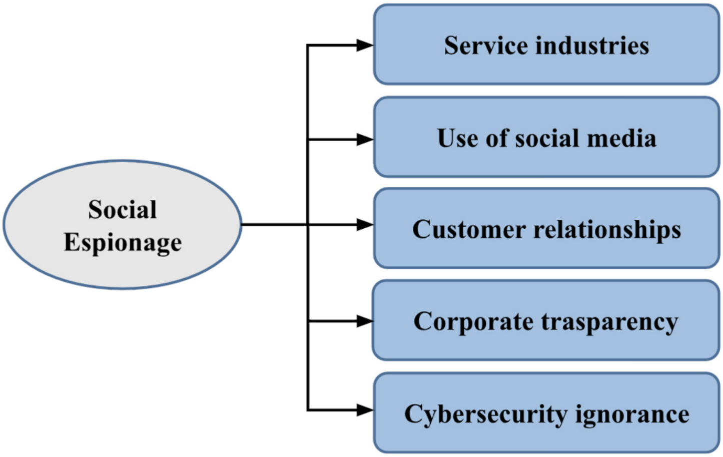

Fig. 1. Trends promoting shift to social espionage.

Social espionage emerges from the three meeting points of enterprises (suppliers, competitors, customers) as actors, social media (technology platform) as an asset, and competitive intelligence as an activity that connects actors and resources together in a harmonious way [2–4]. The lack of training courses for personnel working in the organization represents a further element of risk for the dissemination of company information. It is known that most of the espionage activities are conducted on the basis of the users' superficiality in the use of information technologies and on the lack of adoption of cyber security procedures in the workplace [1].

Social Media Intelligence (SOCMINT) is the process of capturing, collecting and analyzing data from social media platforms to obtain data and information of specific interest useful for satisfying the information needs of the top decision-makers of an organization. This can require the adoption of a multiform data collection process from multiple social media platforms such as, for example, Facebook, Twitter, Instagram, LinkedIn, to monitor the trends of the masses, identify communication patterns, propaganda and psychological conditioning and to obtain information, for example, on the behavior of its customers, on the reputation of the brand or the service offered or on the "sentiment" referred to the reference market. SOCMINT is essential for any organization (public, private, institutional) that intends to constantly evaluate the presence of information within the social digital ecosystem of Cyberspace. It can involve using advanced analytics tools and techniques for data mining, such as natural language processing and machine learning, to process the vast masses of data constantly produced by social media. It should be noted that SOCMINT finds its most natural application for the enhancement of Competitive Intelligence (CI). CI focuses on obtaining information about strategic capabilities, intellectual properties, product formulations, technological processes, business plans, and potential competitive threats [5]. According to [6], CI includes the search for information on competitors and target markets, as well as data relating to competitors in relation to business opportunities and possible weaknesses of competitors. Based on this evidence, the protection of trade secrets assumes strategic importance for the company and for its very survival.

It should be emphasized that running a SOCMINT activity requires different skills that can guarantee an accurate selection of the information to be acquired, a deep and detailed analysis of the selected information and a correct assembly of the selected data that can translate into a "product of intelligence" useful for the top decision-makers of the organization. Typically, among the techniques adopted, the following are highlighted: (i) Social Network Analysis, used to analyze the relationships and interactions between individuals and groups on social media platforms. It allows you to identify the leading influencers, the connections between individuals or groups and behavioral patterns; (ii) Natural Language Processing, allowing you to analyze the sentiment, tone and context of social media posts, comments and conversations. They allow you to identify underlying themes, opinions, and emotions related to specific topics or issues; (iii) Geolocation, to identify the physical location of social media users/groups based on their posts or the profile information that produced them. It allows you to identify the geographical context in which specific events or incidents are taking place and to trace the movement of individuals or groups.

The techniques and methodologies used can be different and must be identified and personalized on the basis of the reference target. In the literature, at the moment, being a particularly innovative technique and above all modelable on the basis of the cognitive needs of the "customer", detailed examples of platforms of this type are not available. However, it is possible to cite two examples of real application of SOCMINT, both referable to the current ongoing Russian-Ukrainian conflict. The first can be traced back to an activity conducted by Radio Free Europe/Radio Liberty (RFE/RL) in collaboration with the Conflict Intelligence Team (CIT), which made it possible to highlight the deployment of troops by the Russian army a few weeks before the 2022 invasion of Ukraine[1]; the second refers to the use of a SOCMINT model which made it possible, through the analysis of images, posts and other information taken from social media, to identify the perpetrators of a massacre carried out in Ukraine by Russian soldiers[2].

In this paper, we aim to analyze the phenomenon of social espionage and focus on two relevant ways of conducting it in the context of social media intelligence, briefly reviewing the main characteristics of both of them. The first one is the use of Social Media Intelligence (SOCMINT) platforms to capture social data streams that can be employed to learn useful information, for example, about the company where someone is employed. The second one is related to fake social profiles suitably constructed to capture the information held in different social conversations. In this case, it would be the victim of a human intelligence methodology, better known as Virtual Human Intelligence VHUMINT [7].

To the best of our knowledge, it is the first work where social espionage is studied and analyzed in the context of social media intelligence.

The paper is organized as follows. Section 2 presents the main characteristics of a typical SOCMINT platform and describes a state-of-the-art SOCMINT architecture. Section 3 introduces the concept of persuasive communication in social media which is connected to the VHUMINT. Finally, Sect. 4 draws conclusions about the previously described concepts and outlines future work directions.

2 Social Media Intelligence Platforms

Let's assume we want to acquire from the most common social media information attributable to the activities conducted by a company and the level of exposure to the dangers of an espionage activity. The first step is to create a platform that allows to analyze the data present in the three most used social media: Facebook, Twitter, Instagram. The next step is to choose an optimized platform for retrieving posts from the three social channels based on specific tracking criteria. Among the numerous real-time tracker platforms available on the market, i.e. capable of tracking hashtags, accounts, keywords and Uniform Resource Locator in real time (URL) contained in the messages, we find: Metricool, Hashtracking, Tweetreach, Keyhole, AgoraPulse, Hashtagify, Socialert. These

[1] https://www.svoboda.org/a/na-ukrainu-edut-rodnye-voennyh-O-perebroske-vojsk/31661739.html.

[2] https://www.ilsole24ore.com/art/da-foto-responsabile-ecco-come-L-ai-individua-criminali-gue-AEk4qpuB?refresh_ce=1.

platforms constantly analyze all the information entered into social networks according to keywords.

At this point it is advisable to proceed with the creation of a system based on criteria for tracking the posts published in the previously selected social channels (Facebook, Twitter and Instagram) which allows for the extrapolation of such contents by means of an online monitoring platform. We then proceed with the creation of a special storage system for storing the data collected in specific databases which will be indexed together with the Key Performance Indicators (KPI) provided by the online monitoring platform. We will then proceed with the creation of an Application Programming Interface (API) to extract the raw data in json format and to allow the analyzes that will subsequently be used by means of an Analytics dashboard or tool, in order to allow the visualization of the data (see Fig. 2).

To understand the evolution of the phenomenon over time, it is essential to analyze the following elements:

- **Reach**: measures the diffusion of the message that spreads in social media, in terms of the potential number of people who have viewed it. In the case of a communication campaign aimed at increasing awareness, it indicates the possible number of potential consumers who have been subjected to the message, and consequently who have become aware of the brand. It is an important measure but it must be combined with other metrics that integrate the information with the number of people who have actually interacted or with other users who spread the message (Influencers, People Involved);
- **Sentiment**: sentiment refers to the feeling/mood expressed by a subject through his message, and is measured in terms of intensity, i.e. whether it is positive, negative or neutral. The measure refers to individual results (mention), or more generally to a topic of discussion, in aggregate form. The Net Promoter Score (NPS), on the other hand, is an index derived from sentiment, and represents the difference between the number of promoters of the discussion topic, people who speak well of it, and the number of detractors. Also in this case, it is useful to consider the variation of these two metrics over time to evaluate their impact and the effect it has on the community;
- **Influencers**: they are the web and social media users who manage with their opinions to influence the respective opinions and decisions of a more or less consistent pool of other users, as they have built a certain credibility and a reputation as experts regarding one or more themes;
- **Engagement**: it is the metric that groups all the interactions that people have with the "discussion" or with other users, or on a message. In particular, it represents the sum of likes, comments and shares of a post. To obtain higher quality data, it is also advisable to measure the relationship between the total engagement of posts in a certain period, and the number of views of these posts, or the number of fans/followers in that given period, to understand how many active people within their community;
- **Number and variation of people involved**: they represent those people who take an action against the "discussion", who interact with likes, comments or shares. It is important to analyze these data, as well as their reciprocal variation over time (e.g. fans remain unchanged but people involved increase following a company campaign, or vice versa, etc.).

- **Total impressions**: the number of times a page featuring content is loaded during the campaign period. Impressions alone, however, are not significant because they usually represent a huge number, in the order of millions, which however does not correspond to an equal number of people who have actually viewed the video;
- **Multimedia contents in the texts**: repeatability of the same multimedia contents in the posts of a topic (e.g. 100 posts 25 with the same images).

On the basis of the analysis of the information elements indicated above, and according to their enhancement by an intelligence analyst, it will be possible to achieve specific vertical knowledge on a specific target.

A system for tracking and analyzing social media which adheres to this general architecture and concepts has been proposed in [8]. The system, called SocMINT, gives users the option to gather information from various social networking sites, and it offers an interactive dashboard that supports multidimensional analysis and uses a deep learning-based approach to forecast social sentiment in relation to both textual and visual data. A method is also put out for determining the overall sentiment of a photograph based on both visual and textual information, and it is implemented into the analytics system. A two-stage deep learning classifier is used in the model. The image is stage one, and the text inside the image is step two. By doing this, content from social media platforms like MEMEs can be incorporated into the study.

The visual feature extractor is trained by fine-tuning the basic model pre-trained on ImageNet and is based on the VGG16 network architecture. The textual feature extractor first finds and recognizes text in images before performing sentiment analysis, whereas the visual feature extractor applies to the entire image. The textual sentiment classifier uses transfer learning of the classifier utilizing a BERT layer as its foundation.

More specifically, the system workflow is characterized by the following steps:

- Data collection,
- Data storage,
- Social KPIs,
- Sentiment analysis,
- SocMINT.

In the data collection step, data is gathered from Instagram, Facebook and Twitter. Data collection process is performed through queries on hashtags and keywords from a list of selected social media accounts. Data is preprocessed to include various information, such as the text of the post, publication date, author, list of hashtags and user mentions found in the post as well as links to media in the post, number of likes or reactions, number of reposts and geolocalisation if available. The different posts are collected in a relational table, together with metadata and the included media.

For data storage, MongoDB is used as a Database Management System (DBMS) to collect the data from social networks. The structure of the database is composed of three main fields: user, storing the information regarding the user, rawposts, which is a raw collection of the posts from the social networks and posts, which is a collection of data that will be adopted for the visualization in SocMINT.

The social KPIs employed by the system to evaluate the performances of social media posts and users can be classified as trend and multimedia KPIs.

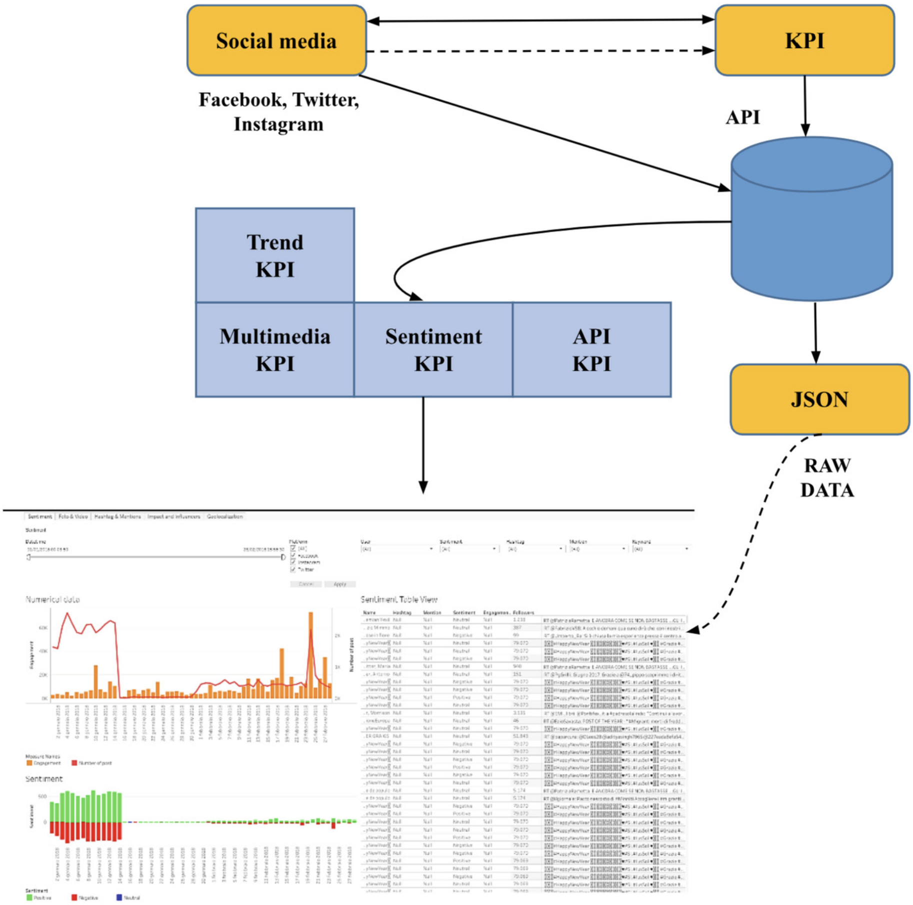

Fig. 2. SOCMINT online monitoring platform.

The trend KPIs are adopted to monitor the trends that emerge through time and characterize a phenomenon in social media. They are: (i) engagement, (ii) quantity of posts published, (iii) degree centrality, (iv) closeness centrality, and (v) entropy. Engagement is a typical metric that reflects the interest users have shown in a piece of content by engaging directly with it. The quantity of posts published is the number of published posts containing a set of keywords or hashtags. In the context of social media, a user's degree centrality can be determined by counting their friends on Facebook, followers and following on Twitter, and following on Instagram, while the degree centrality of a community group or page, both open and private, can be determined by counting subscribers and users who have subscribed, indicating interest in the contents of the page. Closeness centrality can be used to measure the influence of users within a community. This is because it offers an efficient measurement of the dissemination of information within a network, operating under the premise that the better a node's positioning within a social network, the higher its closeness centrality value. Finally, entropy evaluates the users' influence within the mobile social network and is based on the number of node links in the graph, by taking into account the number of friends (Facebook) and the number of followers (Twitter and Instagram).

About the multimedia KPIs, the combination, or more accurately the weighted average, of the individual sentiment values acquired for the text and image linked to the post will be used to evaluate the post's overall sentiment. The user will choose the associated weights to enable customization.

Sentiment analysis is performed by two sentiment classifiers (i.e. visual and textual sentiment) and by a module for overall sentiment evaluation. Sentiment from an image is captured by a VGG16 network architecture pre-trained on the ImageNet dataset and fine-tuned for understanding what can or cannot be positive. The extraction of textual features from images is performed using a two-step workflow based on the Character Region Awareness for Text Detection technique (CRAFT), and the output is used as input for a spatial attention residual network (STAR-Net). Finally, overall sentiment analysis is performed as a weighted average of the single sentiment that is obtained from the image and text analysis. Textual sentiment is computed from a pre-trained BERT layer embedded in the neural network model, which is then fine-tuned on a portion of the domain dataset in order to perform sentiment prediction. At the end, an overall sentiment score is generated as the combination of visual and textual sentiment scores.

For each aforementioned model, fine-tuning consisted in training the basic model with a training set and validating it with a validation set for avoiding the overfitting. Fine-tuning was stopped when validation error started increasing while training error was decreasing. On a total of 20,000 posts for the validation and training set, the consumed time for fine-tuning the BERT and VGG16 models was 40 and 27 min, respectively.

The SocMINT visualization and analysis dashboard is populated from a dynamic input that can be modified at any moment by the user and is made up of the following:

- A list of keywords;
- A list of hashtags;
- A list of social platforms accounts/handles.

SocMINT includes five main views: (i) sentiment, (ii) multimedia, (iii) hashtag, (iv) influencers and (v) map of location data.

3 Human Intelligence Methodology

A second relevant aspect of SOCMINT is the human intelligence methodology (VHUMINT), which is also connected to the creation of fake social profiles suitably constructed to capture the information held in different social conversations.

The key aspect underlying this phenomenon is persuasive communication, which is the set of persuasion strategies and methods useful for carrying out activities of influence and psychological persuasion, aimed at trying to influence the decisions of others. Persuasive communication in social media, so that it can produce effective and profitable results, cannot disregard the adoption of a profile analysis scheme with which one intends to interact. The profile analysis, especially in intelligence activities, represents the starting point of VHUMINT activities. For example, suppose you conduct an information search on a Facebook profile. The attention will have to focus on the analysis of the fields Education, Places where he lived, Contact and basic information, Relationship, Family members, Places visited, Sport, Movies, TV programs, Books, Application and games, Like, Events, Reviews, Group, Followers, etc. On the basis of the information taken from these fields it will be possible to proceed with the creation of a frame of knowledge which allows to outline the characteristics of the profile. One of the most common theories that emerges in the literature on virtual relationships is that of Social Penetration Theory (SPT). Conceived by psychologists Irwin Altman and Dalmas Taylor [9], SPT argues that relationships form and develop as information is gradually exchanged between two or more people. The individuals involved in the relationship exchange information according to the cost-benefit analysis of the relationship. When the perceived benefits outweigh the risk of disclosing information, a relationship will form the "Lives a Followers Friends" links on Facebook can provide us with a precise indication of the social, cultural and economic environment in which you live, including interests and tastes. The "Stories" link, enriched by stories, comments, hashtags and emoticons, allows us to learn which emotional states the user experiences or has experienced and consequently what the suggestions, considerations and welcome topics may be.

"Likes too_" present in the posts can say a lot, such as the degree of appreciation (or disgust) of a particular statement, an event, an image or a comment posted on the social network by another profile. Another important clue, capable of providing valuable information, is represented by the photos published on the profile. If they portray the person in various poses, more or less captivating, they denote the evident desire to satisfy one's ego, but they may contain subliminal messages referring, for example, to sexual aspects (photos in skimpy or very veiled clothes), cultural (the person has himself immortalized against the background of a bookstore, in front of a training institute, at the entrance to a university, etc.), professional (photo in one's office or studio, image of the company in which he works, brand of the institution or company, etc.), sportsmen (in gymnastic gear or while playing a sport). The photos of cities, landscapes, famous places, means of transport, denote the desire to convey a message referable to personal depth: a person who travels, in the collective imagination, exercises a prestigious, important and highly profitable profession. The irrepressible desire to appear important and interesting assumes, for the individual, the connotation of the primary objective to be achieved, to be accepted, esteemed, loved but above all to assume a higher level of visibility than others. This irrepressible ambition manifests itself in a non-place like the virtual world, made up mostly of people we don't know. The image society is based on a culture made up of icons generated by the world of politics, sport, fashion, entertainment and television, which permanently provide indications on behaviors, lines of thought, habits and customs to adopt to ensure a continuous and higher level of interest and importance. Emblematic is the video made by director Shaun Hington, titled "What's on your mind?" (available on YouTube), which suggests a completely different interpretation of the contents of posts published on social networks: often behind the carefully chosen images and words, there are completely different realities, made up of loneliness, boredom, failures and pain. The images published on social networks become an absolute priority which often leads to oversharing, i.e. sharing every single detail of one's private life online, if not even of every single day lived. Two psychologists from Harvard University, Diana Tamir and Jason Mitchell, have conducted research which has shown that 30–40% of communications between individuals concern topics of a personal nature, a percentage that rises to 80% within social networks. The desire to communicate one's thoughts, emotions and reflections to others is strongly connected to the activation of brain areas designated for the perception of the sense of gratification and pleasure. The research shows the existence of a form of satisfaction that arises from the impulse to talk about oneself to others, comparable to a primary need such as food and sex. Another element that emerged from the research is that relating to the role played by social media as tools capable of inducing the individual to play a role that does not correspond to reality in an attempt to hide the negative aspects of one's personality from him, enhancing the positive aspects, most of the time not even possessed. If self-centeredness and megalomania can assume a predominant role in social networks, it is also true that the need to expose oneself and show off is alive and widespread, and this produces an increase in the number of personal information entered in one's profile. Once the limits of understanding the vulnerability of personal privacy have been overcome, the user runs the risk of overcoming even those of intimacy, slipping down a tunnel that leads to the loss of a sense of reality. If such a condition can produce disastrous effects for the user, for a VHUMINT Collector, it represents the ideal scenario for conducting information acquisition activities thanks to the use of a fake profile built according to the characteristics of the identified target (see Fig. 3).

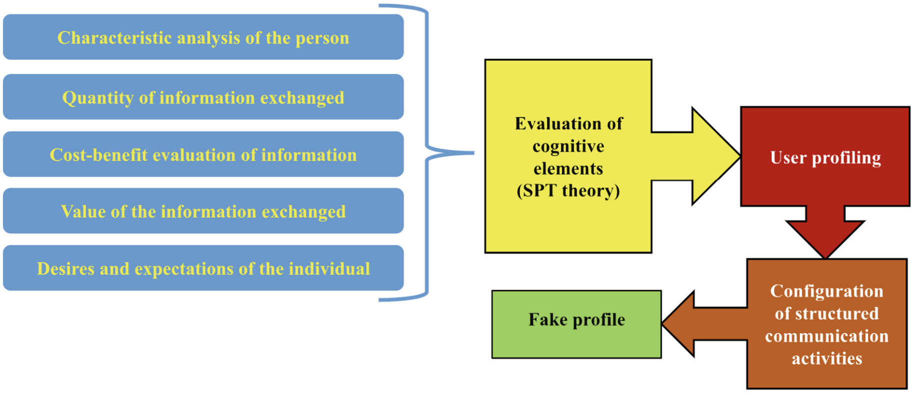
Characteristic analysis of the person
Quantity of information exchanged
Cost-benefit evaluation of information
Value of the information exchanged
Desires and expectations of the individual
Evaluation of cognitive elements (SPT theory)
User profiling
Configuration of structured communication activities
Fake profile

Fig. 3. Stages of the evaluation of the SPT profile.

4 Conclusion

Social Media Intelligence can allow you to search, extract and assemble data and information that can take on the connotation of an "intelligence product" that can be used in different sectors, such as defense, economic, industrial, social, political.

In this work, we provided an overview of SOCMINT under the social espionage perspective, which can be defined as an attempt to gain a competitive advantage by acquiring all publicly and semi-publicly available information in social media. Accordingly, we analyzed two different forms of social espionage, the first one referring to SOCMINT platforms and architectures, the second one which is connected to the creation of fake social profiles suitably constructed to capture the information held in different social conversations.

Artificial intelligence platforms, in this regard, can represent a suitable tool for the automatic processing of information, allowing to quickly produce that knowledge useful to the top decision-makers of the organization to identify the best decisions to take. Depending on the unstoppable exponential increase in data present on the Internet, and in particular on social media, the use of artificial intelligence platforms equipped with special algorithms built to conduct research and data extraction activities that are actually useful for the creation of an "intelligence product" in different sectors, such as the economic, industrial, political, social and defense sectors, represents the "highway" on which to concentrate investments for the coming years.

References

1. Teti, A.: Cyber Espionage and Cyber Counterintelligence. Cyber espionage and counter espionage. Rubbettino, 221 p. (2018). ISBN-13:978-8849852660
2. Håkansson, H., Snehota, I.: No business is an island: the network concept of business strategy. Scand. J. Manag. **5**(3), 187–200 (1989)
3. Easton, G.: Case research as a methodology for industrial networks: a realist apology. In: Naude, P., Turnbull, P.W. (eds.) Network Dynamics in International Marketing, pp. 73–87. Elsevier Science, Oxford (1998)
4. Håkansson, H., Ford, D.: How should companies interact in business networks. J. Bus. Res. **55**, 133–139 (2002)
5. Fitzpatrick, W.: Uncovering trade secrets: the legal and ethical conundrum of creative competitive intelligence. SAM Adv. Manag. J. **68**(3) (2003)
6. Miller, S.H.: Competitive Intelligence-An Overview. Competitive Intelligence Magazine. http://www.ipo.org/AM/Template.cfm?Section=Home&Template=/CM/ContentDisplay.cfm&ContentID=15904 (2005). Retrieved 14 May 2011
7. Teti, A.: Virtual Humint. The new frontier of Intelligence, 165 p. Rubbettino (2020). ISBN-13:978-8849860078
8. Mameli, M., Paolanti, M., Morbidoni, C., et al.: Social media analytics system for action inspection on social networks. Soc. Netw. Anal. Min. **12**, 33 (2022)
9. Altman, I., Taylor, D.A.: Social Penetration: The Development of Interpersonal Relationships, 459. Holt, Rinehart, & Winston, New York (1973)