SURVEY OF SENTIMENT ANALYSIS FOR GENERAL SOCIAL DATA

[1]Syed Zabiulla SK, [2]Md.Ziaulla
[1]Asst. Professor Dept. of ISE , KNSIT ,Bangalore -560064
[2]KNS Institute of Technology, Bangalore
Email: [1]syed.zabihulla@gmail.com.

Abstract– In this era of automation, machines are constantly being leveraged to interpret exactly what people are saying on social media. Humanity these days has become preoccupied with the concept of what and how people think, and subsequent decisions are largely based on massive drift on common platforms. In this article, we'll take a multi-faceted look at how sentiment analysis has evolved into the limelight with the sudden explosion of internet data overload. This article also describes the process of collecting data from social media over the years and similarity detection based on similar choices made by users in social networks. This article also described techniques for sharing user her data. Data in various formats were also analyzed and presented as part of the research in this article. Separately, the methods used to assess emotions have been studied, categorized, compared, and their limitations revealed, with the hope that this will provide space for better research in the future.

Terms- Massive ,Sentiment Analysis ,Emotions

I. Introduction

As people, we always tend to be attracted to like-minded people. Research shows that we feel comfortable connecting with people who share our beliefs, who we trust, and who can help us reach our goals. Etymologically, people tend to associate with like-minded communities. Several clusters form communities. Modularity is one of the most important mechanisms considered in determining the number of communities [1]. A detailed analysis of cluster characteristics can help identify specific character sets or groups of like-minded people in individual clusters.In other words, the existence of common ties between groups of people ensures that similar principles and purposes exist among them.

Specifically, there are two types of them on social media. Social networks and online communities.

Social networks are formed by people who are linked by previous personal relationships and prefer to join new associations to keep them social and expand their personal contacts. Connect people with direct connections. Compared to the former, communities include people from multiple walks of life with little or no connection between them. The primary bond between individuals within a community lies in their fondness for intimate interest. Apparently people stay in communities for a variety of reasons. It could be a preference for something specific. Or maybe you feel that you need to connect with that community or that committing to it will help you achieve something. While social networks contain clearly organized arrangements, communities contain overlapping and nested arrangements between them. Social media is a way of sharing data with a large and wide audience. It can be treated as a medium for disseminating information through interfaces. The combination of social media and social networks allows individuals to contribute their content to the wider community and reach more people to share or promote [2].

Sentiment analysis is the process of classifying opinions expressed about a particular object. With the advent of multiple technological tools, recognizing the public's perception of a company, product, or shared likes and dislikes issue has become an important measure. Tracking the sentiment behind your social media posts can help you identify the context in which you want users to react and move forward. This article presents the evolution of social network analysis (SNA) from more than 100 articles and introduces the research that has been done in social networks and related fields. This article is organized

as follows: Section II discusses the emergence of social networks in the research environment. Section III explains the motivation for this article. Section IV contains detailed methods for finding clusters and communities from over 40 articles. Section V reviews 45 articles in related areas, and Section VI describes the various efforts that have been made to social media data. Section VIII describes various techniques used to identify accurate sentiment from social media data. Section VIII concludes the work by presenting potential future research in this area. Crawlers, tools for evaluating semantics, engines that enable language preprocessing, and classifiers are key components of sentiment analysis systems [3].

II. The beginning of social networks and their analysis

SNA refers to the interconnectedness of people in social networks as cliques. A clique can be defined as a structure in which each member of a group of people is directly and coherently connected to each other. Bron and Kerbosch [4] proposed that there is a way to find the maximum total number of subgraphs in an undirected graph. Specifically, it uses a backtracking algorithm that uses branches and joins to truncate branches that do not lead to clique formation. Work related to developing algorithms for clustering correlated data from various social network applications began as early as 1975 [5]. There, the convergence of the correlation product-moment matrix was used in his CONCOR algorithm for hierarchical clustering. It's also worth considering that with the recent social media data deluge, it must have been a daunting task for search engines to index the billions of data stored on websites. Therefore, Google's framework is detailed [6] and includes scalability and robustness features to provide perfect search results. Providing error-free web page results on the Internet has been a concern since the inception of social networks and their analytic contexts, so researchers have sought to find relevant information when searching for information on a wide range of topics on the Internet. We proposed various techniques to track data efficiently [7]. Certain factors had to be considered, such as the exponentially increasing amount of data and the associated high quality of results. For a broad theme, page quality was also important. Because they are the most relevant to the user, and in the end, he found the hub his page closely linked to a series of correlated and authoritative pages that could give an overview of social relevance.

The contribution of this article lies in the fact that the development of digital data, which is so necessary for understanding public opinion, has been studied from the beginning. We were shown the path from website data to the final sentiment analysis. Different areas where data was collected to analyze sentiment were provided in one place. Collecting and ultimately evaluating data from social media offers many options for solving real-world problems.

III. Motivation

With the vast amount of data accumulating on the internet, it's time to put social media and the issues surrounding that data to the forefront. What others think everyone should follow is a trend that has been noticed by the masses. We describe it in chronological order and finally detail the prediction of sentiment within this data. To the best of our knowledge, this article is the first attempt to gather this wealth of information from over 200 articles and cite their contributions.

field. It is a well-known fact that qualitative research involves how people observe the reality unfolding around them. The problem is therefore learned in its natural environment. This article provides qualitative insights into the use of social media data and how it can be analyzed to help the world understand the emotions and behavioral patterns of the people around them. The authors strongly believe that this unique paper will help experienced researchers access a range of meaningful research in the same place, and will allow them to appreciate the amount of research already done in this field. The future coverage of each paper listed in the table serves as an easy reference for considering ideas for future work. serves as a basis for starting. As we become more dependent on technology, this article will serve as a benchmark for other researchers who wish to continue addressing the aforementioned challenges facing social media data.

Since almost all conversations take place online rather than offline, it's important to get a good understanding of social media and its components to analyze sentiment. We find that since 2004 [207] has focused more on sentiment analysis. As such, the author considers his post-2008 social media posts and accompanying material in this review. A review of previous research on sentiment is not sufficient to adequately capture the methods used to assess the semantics of sentiment. It was therefore a strategic decision to provide a detailed description of social media and its components, and finally jump to the main topic of trend sentiment analysis. In short, the main purpose of this paper is to present the research that has been done in this particular field and to address its limitations that define the broad scope of research for scientists.

IV. DATA COLLECTION

Collecting data from social networks requires great care to accurately predict the ideology of the users behind posting this data on social media. Popular social media sites such as Facebook, LinkedIn, and Twitter provide a medium for communities to post, share, like, and comment with other friends within the same network. Meanwhile, sites such as Youtube, Flickr, and Digg are also preparing various features to improve your connection with your social friends. Data collected from social media is associated with many factors, may be noisy or not, homogeneous or heterogeneous, and diverse.

The main techniques used to extract data from social media include:

1) collection of new data; 2) reuse of previously available data. 3) reuse of data that does not belong to a specific individual; 4) Data Collection. 5) Data from the internet (social media, text, photo sets).

Collected data is processed first. Data processing involves multiple actions, such as checking authenticity, understanding the overview, as well as modifying and adapting it to a suitable form for further use. after this step, The processed data is analyzed to derive the actual underlying emotional results of the user behind this data. Technically, he has three main methods widely used to collect data [8].

1) Network Traffic Analysis - This is a method of collecting packet streams from your network connection, which can further help you track your browsing information on the network. For security reasons, this method is rarely used, especially in private groups.

2) Ad Hoc Applications - This is a set of Air Position Indicators (APIs) that can provide information about the account holder and further track user activity on certain her websites. 3) Crawl - The most common way to collect data from social media. Public information is provided when a query requests specific types of data. Crawling is also useful for retrieving data through APIs available on some social media sites.

V. SOCIAL NETWORKS

Social networking can be defined as using social media to connect with known and sometimes unknown acquaintances. Sometimes we connect with friends, relatives, contemporaries and colleagues, sometimes we connect with users for business purposes. The full story that led to the progress of SNA can be found in [9]. Early work in this area dates back to [10]. There, movie data was collected to build models from relational statistics. Considering the vast amount of data, we considered a language for extracting relevant information and an algorithm for constructing a relational theory of probability to create a relational data classifier. In [11], a small sample of emails showed that graph-based algorithms were more efficient in identifying who knew what in an organization compared to content-driven algorithms. rice field. Relational dependency networks were proposed in [12], highlighting the fact that models that can find dependencies lead to better classification of data.

VI. Clusters and Communities in Social Networks

Clustering is one of the most popular methods for distinguishing and evaluating community structure in social networks, but there are differences between both terms. Different types of traits are taken into account when dealing with clusters, but the community generally sticks to one type of attribute while it is being recognized. Both clusters and communities have differences in how they consider connectivity. Cluster detection can easily be performed on congested links, while the latter detection assumes that there are few links in the network. A. Clustering and its applications

Clustering can be defined as a technique that identifies regular arrangements within groups of entities [13]. Probing properties of various types of traditional and relatively new spectral clustering methods help successfully solve the problem of finding the similarity in controlling a data set [14]. On social media, specific clustering mechanisms such as hierarchical clustering [15] have been applied to a large extent to folksonomy, helping to understand the intended customer preferences as well as specificity of resources [16]. Studies show that social graphs have also been used to represent important clusters of user connections [17], where factors such as frequency, intimacy, and contact times are included. recently helped define a significant online link. Social cohesion has been an issue of concern even before the advent of the computer age, and therefore, the detection and scrutiny of components is necessary for a proper understanding of social networks. 18]. Getting users to share opinions on different issues on social media and better differentiate these topics is not an easy task, which requires clustering of visits. Webview for intelligent detection of security threats. In this regard, the scalable distance-based algorithm [19] shows high accuracy in discovering core problems while eliminating noise. The noisy association detection and its reduction effect in clustering can also be eliminated by using arbitrarily matched regions combining related data and social index with cluster information [20]. Clustering has also been applied to combine both socially and geographically to find the proximity of visits to related clusters [21]. This method outperforms traditional spatial clustering [22] by clustering a lot of spaces in a short period of time. With the recent explosion of social data, the problem of finding the set of repeating elements in big data has been solved by the MapReduce model [23], which implements the k-means clustering algorithm [23]. 24] to preprocess the Data and extract frequent data sets using a priori algorithm [25] and Eclat [26] [27]. Experience proves that this method gives excellent results on big data at very high speed.

B. Detecting community on social networks

To manage the constant increase in indexing of web pages and to maintain stability of accuracy and recall, it is imagined that defining a consistent community and linking it to relevant links should solve the problem [28]. The purpose of the community is to identify associations of individuals who are strongly involved in social networks. Traditional approaches such as Kernighan-Lin are problem-based. Therefore, many algorithms have been proposed to identify community patterns in networks, some of which may differ from traditional community detection methods but ultimately yield similar qualitative results and also involves redundancy of vertices [29], while the other uses ethereal methods. techniques, testing clusters and perception of modularity [30]. Some techniques have also used traditional methods that impose certain shortcuts, resulting in the execution of algorithms in linear time [31]. In [32], Lennard-Jones clusters were investigated to determine the potential energy landscape (PEL) using the network topology and the community structure that was discovered. Considering the measure of the centrality of an edge in the network [33], algorithms have also been devised to find the most important edge by truncating the less important edges. algorithms are also devised to discover the most significant edge by trimming the less important edges in turn to eventually form disconnected clusters. Other variations of the general but fast division algorithm were invented in the same year [34], [35], one recounts the number in the middle after removing each edge while the other refines the algorithm. Computationally expensive GN computations require non-topological data to clarify branch details. that is structurally important. Initial research focused on the diagrams from which the complete configuration was determined. In addition, Clauset [36] proposed an agglomeration algorithmic approach that works on dynamic and over-heavy graphs. This approach considers each vertex one by one and also demonstrates that a simple application of this algorithm would be suitable for implementing robotic programs that help unearth neighboring communities on the Internet. Among the algorithms that have been deployed on large networks, the research of [37] has shown methods to discover the extremely overlapping, nested and constrained association of nodes in binary networks. A strict version of the web community detection mentioned in [38], incorporating the construction of a Gomory–Hu tree [39] yields efficient computational results. In addition to the diverse approaches based on the works of GN, a technique mentioned in [40], this technique maps the dilemma of community identification in finding the status quo. Baseline situation of Potts vortex glasses of unrelated scope through consolidated data from ansatz with the cooperation of present and absent collaborators. Although the search for communities within networks has been appreciated since its inception, there is no prescriptive definition of the same, taking this into account, the researchers of [41] uses reference techniques and represents the community detection as a problem of inference or maximum probability. A notable example of community detection using matrix eigenvectors has been described in [42], where the modulus function is reconstituted as a matrix resulting in a work representation. optimization as a spectral dilemma. In another case [43], the concept of order from a single vertex was extended to subgraphs reducing the overall complexity and the same was applied to create a tool called ModuleNetwork (MoNet) [43], to find community training courses with . The random walks that have been used to compute similarity in structures in vertex space, if implemented in hierarchical algorithms, will yield efficient results [44]. Examples of methods that take into account the smallest loops passing through a particular node have also been cited in [45], which is claimed to be the modification of the neighborhood skill introduced by Latora and Marchiori [45]. forty six]. Raghavan et al. [47] probed a mechanism that considers only the structure of the network and implemented a simple label generation algorithm where each node is given a unique label and at each step the individual nodes obtain a label that most Neighboring nodes currently own. Other notable works that should be mentioned in the field of community detection are techniques where heuristics have been used to optimize modularity [48], genetics-based methods are used. to discover communities by optimizing the simple yet successful exercise function like mentioned in [49]. Although the concept [47] has been used to detect real-time communities in

large networks [50] and are used to include information about multiple communities when there are parameters for managing a bilateral graph [51]; Optical data mining methods were used to identify overlapping communities to facilitate selection of appropriate constraints after viewing preliminary data visualization [52]. Bridging the gap between people and social media was taken into account to choose a group structure, using interconnected data between users and tags applied on social media. to understand a group's preference for an individual. [53]. Other uses of community-related services extend to maintaining interpretative meaning among question-answer duets data sets, where a structure inspired deep trust through Thinly worded features were expected in the social community, as illustrated in [54]. A comprehensive review of community detection methods, especially in social networks, can be found in [55]–[58]. In addition to these, edges are considered to yield optimal community detection results [59], several data sources are combined to improve the performance of distinct communities [60] and predictions designed to determine how information about trends in the community will be spread in a contagious manner. road. [61]. VII. Centrality Factors for Measuring Node Influence in Social Networking

Another important aspect of networks is the centrality factor. Highly central networks are dominated by entities that control transit flows in the network. In [62], the method of computing the centrality measure was revisited and two deviations were proposed. On the one hand, we found the centrality measure through traffic flow, and on the other through network expansion. Finding the right flow also gives results about node importance and contribution. The observation of centrality and its reducing effect in the presence of error is mentioned in [63]. [64] describes the design of an effective algorithm that generates top-level intimate centrality nodes in a network in less time. Around et al. [65] shows through parameterized measures how terrorism is spread through the use of social networks by connecting with like-minded clusters, so their use should be monitored. Social networks are made up of millions of links and nodes, making it difficult to organize and grasp the big picture. To maintain a balance between systematic and flexible social network discovery, a system called

SocialAction was provided in [66] to efficiently account for multiple geometric and optical network analysis parameters. The parameter ranking methodology in this paper not only allows us to obtain summaries, skew nodes, and identify outliers, but also integrates nodes to reduce complexity, find consistent subsets, and pay attention to communities of interest. can also help pay for We used social networks to predict mental fitness, as shown in [67]. Here, networks outside family and friends were detected. This helped predict symptoms of depression in these isolated individuals. Over time, the growing importance of large networks has created major problems in processing the real-time data generated by these networks in order to understand the fundamental nuances of these groups. . Backstrom, etc. [68] developed a method to predict how large groups within a network will exhibit growth in terms of membership and entity. We also examined networks at the individual level to determine membership in specific overlapping communities. Wikipedia [69], one of her largest sources, doubts the credibility of the material provided. To illustrate this situation, Sicily et al. [70] suggests that reliability issues increase as the network grows, so this issue should be addressed first. It's graphed with references showing people who contributed to the wiki and links to external sources, with specific metrics to exfiltrate data from the network. At the commercial level, social networks have started to play a significant role in the decision-making process of entrepreneurs. In this regard, Aldrich and Kim [71] defined his three models of networks - random, small-world and truncated scale-free. Two association formation models were found. One is logically formed and the other is based on personal relationships.Many variations of the system have been proposed, but rather than relying on one particular type, it is proposed to encapsulate an entire series of entrepreneurs with the goal of growing the company. A businessman promoting his brand used social networks to reach the maximum number of customers through social networks. As more data corresponding to specific consumer networks become available, it may be useful to identify controllers using the concept of centrality. However, determining which parameters to consider is still a problem. In [72], real-time network data were considered to evaluate various centrality parameters for message delivery within the network. SenderRank [72] is a new centrality metric introduced in the paper, which outperforms many existing metrics, but the message category and its content, as well as the type of network, influence the prediction of favorable customers. increase. As social networks appear to shape the entirety of an individual's habits, it is necessary to represent a person's mental state as well as geographic preferences on social networks that are ultimately intended to manipulate one's travel habits. I have. However, challenges in data collection for this study showed that many factors, such as economic stability, mentality, and future plans, influence projections [73]. On the academic side, retrieval of a researcher's information profile through the architecture of the ArnetMiner structure was considered in [74]. Here, we used a unified tagging approach to retrieve researchers' profiles from the Internet without human intervention. The tool also helps integrate existing information about publications in online libraries into networks, forming an entire scholarly network and providing investigative facilities for researchers. The structure of the tool also provides a probabilistic solution for dealing with redundancy issues caused by similar names. With the advent of mobile phones in our daily lives, Eagle et al. [75] taught us how to collect information from mobile phones instead of self-reporting. This process resulted in a vivid depiction

We have also examined the real dynamics between individuals, allowing us to study the progression of these associations over time. However, privacy and security should be considered the most important criteria when using data in these species. In this context, security breaches can also be a serious problem for users of private her profiles. Therefore, users should be careful connecting to and participating in groups, as they may offer to reveal personal information that should be kept confidential [76]. [77] Announced social network middleware MobiClique. This avoids the need for intermediate machines to connect and exchange messages when other devices meet. This middleware can be used to easily distribute messages within another network. A new idea of ModulLand was presented in [78] to develop a better analysis of social networks. Here, the connected parts within the community with the chosen centrality threshold were called modules. ModuLand's approach includes his four steps: setting the ability to control nodes or links, creating a community his landscape, formatting mounds within the community, making higher-level decisions, and breaking down the hierarchy of stronger networks. includes one step.

VIII. SPAM DETECTION IN SOCIAL NETWORKS

Based on some assumptions, experiments were performed and the results were presented in [79]. There, dissemination of fitness behaviors through social networks was implemented. The results showed that the larger the range and number of clusters in the network, the greater the effect on expanding health behaviors. We observed that clustered networks performed better and took less time than random networks. Because users can post arbitrary content or use social platforms to distribute unwanted content using social networks, Gao et al. [80] proposed a method to measure and distinguish spam advertisements run through pseudo-accounts on online networks. Using recordings of messages from the “Facebook” wall, we observed that approximately 97% of accounts were created with the sole purpose of sending spam to the network, and spam messages were typically activated early in the morning. Therefore, it is more or less established that social networks are targets for the spread of spam and malware, and there is a need to develop methods to detect online social spam. [81] proposed another framework that available social networking sites can adapt to limit spam. This proposed framework had many advantages, such as identifying spam on the network and disseminating information about it throughout the network. We observed that this model performs well in terms of accuracy on large datasets. It can help you detect spam on your website efficiently when compared to the huge burst of data from social websites. It was also hoped that this method could be used to prevent spam early on new social networking sites. Multiple classifiers were used and the results were passed to AND, OR, majority voting and Bayesian strategies for spam detection. All entities have their pluses and minuses, and social networks are no exception. Algorithms have also been proposed for the same reason, but [82] combines concepts from graph theory and machine learning and can be applied to small networks. The special thing about this method is that we only need to build a graph topology to detect spam. Social networks like “Facebook” are used for malicious activities on the one hand and for educational purposes on the other [83]. As the teaching paradigm began to shift from traditional methods to more technology-rich methods, as early as 2010 73% of his teachers had Facebook accounts, compared with 93% of his students. had an account. College attendees also checked Facebook and email regularly, while faculty members checked email more than social networking sites. However, it turned out that college students and faculty alike did not see Facebook as an important instruction-sharing vehicle and stuck to the name social rather than educational. In contrast, another study [84] used his algorithm a priori with association rules to understand Facebook's involvement in student-to-student and student-teacher connections. Considering several aspects such as how often students check their social media accounts and how much time they spend on Facebook, students consider Facebook to be the best medium for accessing rich information access I understand.

IX. Impact Analysis in Social Networks

In addition to online social networks, mobile social networks are gradually penetrating the digital world. Therefore, to disseminate the information and persuade people to adapt it, we need to identify key people. A greedy approach is adopted by the community to mine the most important nodes in mobile social networks [85]. First, a community was selected and the community to be used as a medium for disseminating information was determined. This was followed by a dynamic programming methodology describing dominant nodes. This method shows better results than the traditional greedy approach with minimal error. To facilitate procedures between public systems, a framework was developed in [86] to find the maximum number of user profiles used by a person. To accomplish this task, parameters are considered to determine whether two profiles belong to the same person, weights are assigned to them without physical and human intervention, semantic comparisons are performed, Decisions were made using aggregate functions. The results were good compared to existing conventional techniques. Another interesting service for active participation in social networks is the microblogging service Twitter. Twitter seems to be doing the opposite of a normal personal network. His platform requires him to write 140 characters, so analysis of tweets shows that most posts on Twitter are based on headlines and trending his news[87]. As with other social networking sites, practical research was conducted on Twitter to look for widespread computer-generated illicit systems. Internal society group reveals illegal accounts exposed Connected and influential nodes in criminal systems tend to track more accounts. The inclusion of Mr. SPA [88] groups accounts into compartments. In other words, it's a social butterfly that randomly hops back and forth between Twitter accounts. However, it's worth considering that these accounts are extremely rare in nature. Another algorithm, Criminal Account Inference [88], was also applied to this task, collecting details in this area by analyzing social accounts and computing semantic harmonizations between them. [89] presented a case study in which data were collected from various websites such as homepages, blogs, and Twitter in relation to the 2010 Korean parliamentary elections. The aim was to ascertain whether this means of social networking was used purely for conversations with members and citizens, or whether it was used intentionally. The experiment found that politicians on Twitter were more connected to their contemporaries than citizens. In a survey conducted on 15 of his students who were educated about including social his networks in data collected from the Internet, his SNA concepts using NodeXL [90] were demonstrated by beginners in a short time. has been shown to be easy to understand. The intention was to support beginners with the highest quality teaching materials so that they could study her community online according to their own preferences[91]. Initially, network analysis and visualization tools (NAV) were used and SNA instruments were made available to novices. Another article presents a methodology system that helps individuals visually surf in the case of strong dimensional networks [92]. This approach helps to display social communities and exchanges contained in a blended view. The network structure is accumulated as a social graph and their respective interconnected subgraphs are presented visually. As discussed, identifying potential users in clusters or communities has always been an area of research and continuing in a similar direction, [93] presents a technique for identifying key people based on Message exchange is done on a certain topic. The first step is to prepare a graph showing the relationship between posts on a particular topic, followed by the next step to prepare a user graph representing influential users. Based on the attributes collected from the two graphs, the most important graph is selected. Not only from prominent users, but many online customers are also influenced by their peers using entropy pricing techniques as mentioned in [94]. To understand the nature of the links present in social networks, clustering with association sort algorithm was used to distinguish between positive and negative links [95]. This method ensures that there is a social balance between the clusters and that preprocessing is performed on the network before the association index is calculated. But against this concept, a study mentioned in [96] shows that since most of the previous work is based on supervised learning, implicit morality initiates specific behavior of members. social workers are not included. Therefore, a deep trust network-based technique based on unsupervised learning has been proposed to identify links in the network. Besides machine learning models, deep learning mechanisms have also been used to characterize individual behaviors and visualize social networks in terms of health. A study of how social networks can influence individual behavior and how to make assumptions to choose parameters that predict one's behavior has been studied in detail in [97]. Factors such as personal inspiration, implicit and precise social persuasion, and ecological measures were taken into account to predict the individual's expected movement. Methods have also been presented to efficiently fragment the ego network using a genetic algorithm-based K-means clustering structure combined with information received from the behavior of social groups in [98]. Encouraging results have been obtained studying the evolution of associations over time on social networking sites such as Facebook and Twitter. Considering that Twitter data is extremely noisy and comes with additional loaded information, a Twitter Network structure has been proposed to fully model the network using hierarchical Poisson-Dirichlet processes for formulas and an arbitrary Gaussian task for modeling social networks [99]. The role of social media sites like Twitter in emergencies is widely accepted, but it must be understood that inadequately classified data at the start of an emergency slows learning and so it takes time to predict. This problem was addressed in [100], where a Convolutional Neural Network was used to rapidly detect and classify important tweets at the time of a disaster. Social media has had a major impact on socioeconomic aspects of the world and will continue to do so in the years to come. Among the many important factors that determine the concept of a social network, some of the factors mentioned here are relevant to the surveys conducted on these areas. Identifying key users of a social network is a major concern for researchers around the world. A study of all the techniques used depending on the network construction as well as the substance available in the network was reviewed in [101]. The trend towards which social data points, known as opinion spread, is also a concern for social analysis. Cercel and Trausan-Matu [102] detail a study of judgment dissemination as well as the extent of research in this area.

With the recent influx of data on the Internet and the need to obtain meaningful insights into huge amounts of data, automated processes are needed for analysis. Data mining techniques are appropriate to solve these problems. A list of applicable processes for understanding social data based on different dimensions is available in [103].

X. SOCIAL COMMUNICATION DATA USED IN SOCIAL NETWORKS

Social media facilitates the widespread dissemination of information through virtual media. Content on social media can vary from personal data to official documents, photos and data. It was originally used as a means of communication with each other, but over the days it became fully commercial as it has the advantage of being simultaneously to the entire population. A detailed survey of contributions in the field of social media is presented in Supplementary Table I [104]–[174].

XI. SOCIAL NETWORK SERVICE ANALYSIS

As observed in Supplementary Table I, social media mainly processes comments from users, it is very important to analyze them effectively to help understand emotions and opinions. of the general public. The proximity with which people use social media platforms to present their views on every event leads to the need to explore emotions and try to find the best way to evaluate them. The preference for including public emotions through online platforms has increased in recent years.

Sentiment analysis is referred to as a method in which unintended procedures are constructed to infer the sentiment of a text. Individual data have been determined by calculation to help form the information planned for use by decision makers. Due to rapid advances in technology and unlimited access to social media, sentiment analysis is constantly gaining popularity in today's business landscape. Sentiment analysis requires the use of natural language processing management and its various responsibilities such as microtext analysis, irony detection, opposite detection, situational as well as characteristics identification.

Social media data from multiple domains is collected from multiple mediums to extract emotions from text. The presence of multiple languages in texts on social networks, informal writing due to limitations in message size, typos, grammatical and logical errors make the task of sentiment analysis on social networks become difficult. In some cases, file representations in the form of N-gram graphs have been presented to assess substance-based sentiment analysis [175]. It cleverly captures the sentiment of words by matching part of a string and doesn't guess at the primary language.

A. Sentiment Analysis Technique

There are two main sentiment extraction methods, namely the lexical-based method and the classification-based method. Using the former, [176] shows the performance of Semantic-Oriented Computing (SO-CAL). Initially, emotions-related expressions (including different parts of speech) wereused to evaluate the valence-modifying factors responsible for conveying the attitude of the text organization, and finally To be
sentiment is calculated. Two theories have been considered when calculating sentiment, first,

sentiment is independent of context and second, sentiment can be expressed through numbers. This work emphasizes the addition of a point of contradiction that reallocates the rhythm of the word in the presence of denial. Considering health as one of the major issues in our lives, a study [177] shows how analysis of small text messages collected on Twitter can help to understand people's feelings towards them. with vaccination against respiratory tract infection A (HINI). All the posts collected are related to vaccinations and provide the geolocation of the person behind the tweets. Emotion extraction accuracy is 84.29%, combining Naïve Bayes classifier [178] to identify optimistic and pessimistic tweets and Maximum Entropy classifier to identify unbiased and unbiased tweets suitable [179]. Similarly, for the disaster case, a method was presented in [180], where image analysis was used to mine for location-related tweets that influenced public emotions. The Ebola outbreak was examined in this case, and issues such as differences between several sentiment classifiers can be exposed through the model and whether optimism persists in the disaster or not was discussed in the article.

Another example shows the use of an unsupervised combined approach with language processing, dictionary-based methods, and ontological processes to generate classifiers for emotions expressed in media. social [181]. Palavras software was used to test sentiment from Portuguese texts on social networks. This process includes the collection of an entire corpus of data related to a particular topic, standardization of the text, recognition of the relevant entity, discovery of the circumstances under which the entity was extracted, selection of identification characteristics, emotional expression, peak of emotional rate, data accumulation and finally analysis of them.

B Opinion mining

While sentiment analysis is about assessing the sentiment in a text, Opinion Mining means the process of assessing people's attitude towards an entity. A detailed study of Opinion Mining on Social Media, including detailed problem definition, sentiment classification according to different dimensions, regulation of opinion naming, discovery features, relative comments mining and finally spam detection in comments are available in [182]. Perceptions of transit service in Milan on Twitter have been analyzed to provide improved route service, and perceptions can also be used to tailor service to traveler preferences [183]. The tweets were collected both at the beginning and at the bottom of the respective travel agency, and the model designed in the article analyzed the content to categorize events as well as classify attitudes towards the service. transport. One of the simple rule-based sentiment analysis methods mentioned in [184], named VADER, gives results with an accuracy of 0.96 compared to other methods. Size and value factors were taken into account to design a hand-crafted, universal-valued standard wordlist that is suitable for character-restricted microblogging texts.

C. Optical sentiment analysis

Not only in text, the analysis of human emotions is also through images. An optical sentiment analysis classifier [185] based on deep convolutional neural networks was applied on a million tagged images collected from Flicker. This method, implemented on the new Caffe deep learning framework, helps identify emotions depicted in images through the use of adjectives nominal expressions are accidentally extracted from the image. This approach has been shown to outperform conventional methods such as support vector machine (SVM) classification techniques. Another implementation of the deep convolutional neural framework was also used in [186], where however, precise deep learning feature detection was used to discover under-featured images among half a million Flicker images. These weak labels are refined using domain forwarding and transformation methods, thereby refining the neural network. In addition, large amounts of physically tagged visual sensory data through Amazon Mechanical Turk were generated in this work. From both works, one can infer that well-trained convolutional neural networks outperformed existing optical sentiment analysis methods on social networks. Although Yuan et al. [187] found that text and images are useful for detecting users' emotions on social networks. Most of the low-scoring traits were taken from the SUN database [188] and ranked to generate 102 traits with average scores, which were then used to consider emotions. On the other hand, emergent emotions are predicted using idiosyncrasies. This whole method works well to determine well-constructed positive and negative sentiment showing an accuracy of 82 after full implementation. An example of this kind of application is also found for microblogging as mentioned in [189], with a proposal for a framework, allowing a small and large scale view of the details of emotions. extracted. Social sentiment analysis has been explored in many languages such as Czech as mentioned in [174], where management machine learning techniques were used for document-level emotion detection over 10,000 articles. posted on Facebook.

To facilitate industries that strongly challenge each other, a comparative analysis of what people say on social media was modeled in [190], providing a options to design the methods needed for specific advertising strategies for a product. The tentative model is also implemented in the VOZIQ survey tool and further tested on five business concerns to generate meaningful business reports. This structure aims to uncover the most important companies relevant to a particular type of business and provides a detailed report of their performance, focusing on the essential and ultimately open characteristics. way for intelligent judgment. Dynamic Architecture for Artificial Neural Networks (DAN2) [191] solves such a problem, where product performance is tested on a Starbucks-related tweet dataset with an accuracy greater than 80 in every test case. In this case, the feature is evaluated using production-managed features resulting in the feature having seven exact dimensions. Three- and five-class emotion classification was applied to the dataset to provide effective insight into pleasant emotions that may be essential for key brand marketing strategies. As mentioned earlier, emotion detection and its analysis have been performed on most of the world's languages, such a multilingual discrimination application is expected in […] 192], where the adjective-noun duo is used to create a giant polyglot. - Optical language. The emotional system takes into account the data of 12 languages of different origins.

D. Many aspects of emotions and their analysis

Other aspects of social media, where text is visualized, have received attention in [193], which has facilitated both single and multifaceted perspectives. on sentiment analysis on social media. The Multiple Perspective Sentiment Analysis (MVSA) dataset can be viewed as a threshold that generates a positive relationship between textual and optical data. Social media data can also be used to uncover the unfinished errands of an area in Sejong city, as explained in [194], where sentiment analysis model is evaluated. using a Naïve Bayes classifier with 75% accuracy. Just the availability of a lexical source for sentiment analysis, for example, SentiWord Net, is not enough, accuracy is also important when it comes to questions regarding opinion trends. In [195], SentiMI was created to separate individual examples from the object-oriented examples in SentiWordNet, while also extracting parts of speech and evaluating general data for optimistic and sad terms mandarin. Among all the methods mentioned above, a new approach has been focused on [196], in which verbal communication processes have been considered to detect common feelings about a subject. subject. Some of the distinct and asymmetric properties of explicit and implicit expressions and the direct impact that speech patterns have on attitude power lead to the realization of the above concept. First, the agitation, amplifier, and muffler of emotional words are studied, followed by vehicles designed to express emotions without the use of emotional words. and finally, it shows how the connection between sentences determines the overall nature. of a magazine. The first work on identifying irony was reported in [197], where emotions were explored to find irony in texts. The use of custom features and pre-trained models for feature discovery has yielded very successful results. The classification was performed by first applying CNN, then SVM [198]. A global descriptive advance in the field of sentiment analysis over the past decade has been described in the previous sections. Supplementary Table II [199]–[218] provides a tabular format that includes some of the key details of sentiment analysis.

In recent times, due to the constant need to calculate accurately determine emotions, researchers are actively working on many different aspects. This individual technique is not only explored on a daily basis, but hybrid techniques are also followed to prove to be the producer of better results in the case of perfect sentiment detection. The visual sentiment analysis framework together with the merging of low and medium level features of the images proposed in [219], resulted in a 9% increase in accuracy. Supervised learning methods such as K-nearest neighbor (KNN) [220] and SVM [198] were used to extract emotions in images. The features were designed using Single Value Decomposition (SVD) [221] and Color Saturation Intensity (HSI) methods [222]. The extraction of emotions from social media to generate health-related perceptions without the influence of drugs was discovered in [223]. The most common drug with regard to handling, value, cost and regularity of supply can be determined due to the high precision achieved in the process and it has also been found that people care current eye, skin and sexual health concerns. -in the present era.

XII. CONCLUSION AND FUTURE SCOPE

It's time for humans to prioritize the ever-increasing amount of data from social networks. Since almost every complex problem in real life, from biological types to technological types, can be represented by means of social networks, its challenges must also be addressed. The discovery of rumors [224], the echo of opinions, the trend of online conversations leading to chaotic situations and public shame [225], brought change to the preconceived idea, to understand this social popularity in the form of a number of likes, shares and retweets. Features like finding the right content and the right time to post are some of the important issues that need to be addressed on social media before it fully permeates people's lives. Even detection of fake comments should be addressed at the micro level of social sites like Twitter to avoid unnecessary spam harassment [226], [227]. Health-related issues should be addressed in further research so that they have a significant impact on social media users. It would be appropriate at this time to prepare a unified language model that can understand the emotions of users when posting comments on social networks. In order for the brain to think like a human, the subject of object perception must be focused on properly understanding the look and feel of any human object and at the same time their behavior patterns must be studied from their responses to certain events [228]. Video analytics is a major area of study that could become popular in the coming years. Influential nodes responsible for sharing relevant information should be limited by exploiting the feature so that irrelevant information is not spread in a split second.

Last but not least, personalizing the presentation of content on social media and social networks should be paramount to improving the quality of web content. Effective methods for evaluating user comments in social sites for referral systems should be used [229]. Furthermore, the reduction of ambiguity in the liters of data generated daily in these networks still yields many areas of research. Importance should also be given to the combination of literature and technology, where consistency between the adaptation of the original novel and its visual counterparts has been dealt with recently [230].

The authors of this article have recently entered the field of sentiment analysis by making a small contribution to the search for a semantic approximation of a word in a sentence as well as in documents [231], [232] and presented a survey of all the works done. in the indigenous languages of India [233]. This one-of-a-kind article presents a detailed study of social networking and related terms. Work has been done regarding the cluster, community and social network described in its scope. This article is mainly intended to highlight the shortcomings of many articles that allow researchers to easily apply sentiment analysis methods after collecting data from social networks. Novelty is also mentioned for articles on sentiment analysis to help researchers brainstorm creative ideas to train machines to recognize public opinion more effectively. The articles of the 20th century are mainly considered in terms of increasing social media data and its corresponding analysis. It is hoped that more common deep learning mechanisms can be used for social networks as they automatically detect the characteristics of patterns and thus will provide more structure to the unstructured information. architecture without minimal human intervention.

REFERENCES

[1] M. Hoffman, D. Steinley, K. M. Gates, M. J. Prinstein, and M. J. Brusco, “Detecting clusters/communities in social networks,”

Multivariate Behav. Res., vol. 53, no. 1, pp. 57–73, 2018, doi: 10.1080/ 00273171.2017.1391682.

[2] J. Leskovec, “Social media analytics: Tracking, modeling and predict ing the flow of information through networks,” in Proc. 20th Int. Conf. Companion World Wide Web, Mar. 2011, pp. 277–278.

[3] F. Neri, C. Aliprandi, F. Capeci, M. Cuadros, and T. By, “Sentiment analysis on social media,” in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining, Aug. 2012, pp. 919–926.

[4] C. Bron and J. Kerbosch, “Algorithm 457: Finding all cliques of an undirected graph,” Commun. ACM, vol. 16, no. 9, pp. 575–577, 1973.

[5] R. L. Breiger, S. A. Boorman, and P. Arabie, “An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling,” J. Math. Psychol., vol. 12, no. 3, pp. 328–383, 1975.

[6] S. Brin and L. Page, “The anatomy of a large-scale hypertextual Web search engine,” Comput. Netw. ISDN Syst., vol. 30, nos. 1–7, pp. 107–117, 1998.

[7] J. M. Kleinberg, “Authoritative sources in a hyperlinked environment,” J. ACM, vol. 46, no. 5, pp. 604–632, 1999.

[8] C. Canali, M. Colajanni, and R. Lancellotti, “Data acquisition in social networks: Issues and proposals,” in Proc. Int. Workshop Services Open Sources (SOS), Jun. 2011, pp. 1–12.

[9] B. Wellman, “The development of social network analysis: A study in the sociology of science,” Contemp. Sociol., vol. 37, no. 3, p. 221, 2008.

[10] D. Jensen and J. Neville, “Data mining in social networks,” in Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers (Computer Science Department Faculty Publication Series). Amherst, MA, USA: Univ. of Massachusetts, 2003, pp. 287–302.

[11] C. S. Campbell, P. P. Maglio, A. Cozzi and B. Dom, “Expertise identification using email communications,” in Proc. 12th Int. Conf. Inf. Knowl. Manage., Nov. 2003, pp. 528–531.

[12] J. Neville and D. Jensen, “Collective classification with relational dependency networks,” in Proc. Workshop Multi-Relational Data Min ing (MRDM), 2003, p. 77.

[13] S. M. Van Dongen, “Graph clustering by flow simulation,” Ph.D. dissertation, Dept. Center Math. Comput. Sci., Univ. Utrecht, Utrecht, The Netherlands, 2000.

[14] U. Von Luxburg, “A tutorial on spectral clustering,” Statist. Comput., vol. 17, no. 4, pp. 395–416, 2007.

[15] S. C. Johnson, “Hierarchical clustering schemes,” Psychometrika, vol. 32, no. 3, pp. 241–254, 1967, doi: 10.1007/BF02289588. [16] A. Shepitsen, J. Gemmell, B. Mobasher, and R. Burke, “Personalized recommendation in social tagging systems using hierarchical cluster ing,” in Proc. ACM Conf. Recommender Syst., Oct. 2008, pp. 259–266. [17] M. Roth et al., “Suggesting friends using the implicit social graph,” in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Jul. 2010, pp. 233–242.

[18] V. E. Lee, N. Ruan, R. Jin, and C. Aggarwal, “A survey of algorithms for dense subgraph discovery,” in Managing and Mining Graph Data. Boston, MA, USA: Springer, 2010, pp. 303–336.

[19] C. C. Yang and T. D. Ng, “Analyzing and visualizing Web opinion development and social interactions with density-based clustering,”
IEEE Trans. Syst., Man, Cybern. A, Syst. Humans, vol. 41, no. 6, pp. 1144–1155, Mar. 2011.

[20] G.-J. Qi, C. C. Aggarwal, and T. S. Huang, “On clustering heteroge neous social media objects with outlier links,” in Proc. 5th ACM Int. Conf. Web Search Data Mining, Feb. 2012, pp. 553–562.

[21] J. Shi, N. Mamoulis, D. Wu, and D. W. Cheung, “Density-based place clustering in geo-social networks,” in Proc. ACM SIGMOD Int. Conf. Manage. Data, Jun. 2014, pp. 99–110.

[22] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo, “Socio-spatial properties of online location-based social networks,” in Proc. 5th Int. AAAI Conf. Weblogs Social Media, Jul. 2011, pp. 1–8.

[23] J. Tang, J. Sun, C. Wang, and Z. Yang, “Social influence analysis in large-scale networks,” in Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Jun. 2009, pp. 807–816.

[24] J. A. Hartigan and M. A. Wong, “Algorithm AS 136: A k-means clustering algorithm,” J. Roy. Stat. Soc. C, Appl. Statist., vol. 28, no. 1, pp. 100–108, 1979.

[25] G. Schwarz, “Estimating the dimension of a model,” Ann. Statist., vol. 6, no. 2, pp. 461–464, 1978.

[26] C. Borgelt, “Efficient implementations of apriori and eclat,” in Proc. IEEE ICDM Workshop Frequent Itemset Mining Implement. (FIMI), Nov. 2003, pp. 1–10.

[27] S. Gole and B. Tidke, “Frequent itemset mining for big data in social media using ClustBigFIM algorithm,” in Proc. Int. Conf. Pervasive Comput. (ICPC), Jan. 2015, pp. 1–6.

[28] G. W. Flake, S. Lawrence, and C. L. Giles, “Efficient identification of Web communities,” in Proc. KDD, Aug. 2000, pp. 150–160. [29]

M. E. J. Newman, “Fast algorithm for detecting community structure in networks,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 69, no. 6, pp. 66133–66138, 2004.

[30] L. Donetti and M. A. Munoz, “Detecting network communities: A new systematic and efficient algorithm,” J. Stat. Mech., Theory Exp., vol. 2004, no. 10, 2004, Art. no. P10012.

[31] A. Clauset, M. E. Newman, and C. Moore, “Finding community structure in very large networks,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 70, no. 6, 2004, Art. no. 066111.

[32] C. P. Massen and J. P. K. Doye, “Identifying communities within energy landscapes,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 71, no. 4, 2005, Art. no. 046101.

[33] S. Fortunato, V. Latora, and M. Marchiori, “Method to find commu nity structures based on information centrality,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 70, no. 5, 2004, Art. no. 056104.

[34] M. E. J. Newman and M. Girvan, “Finding and evaluating community structure in networks,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 69, no. 2, 2004, Art. no. 026113.

[35] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, “Defining and identifying communities in networks,” Proc. Nat. Acad. Sci. USA, vol. 101, no. 9, pp. 2658–2663, 2004.

[36] A. Clauset, “Finding local community structure in networks,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 72, no. 2, 2005, Art. no. 026132.

[37] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, “Uncovering the overlap ping community structure of complex networks in nature and society,” Nature, vol. 435, no. 7043, p. 814, 2005.

[38] H. Ino, M. Kudo, and A. Nakamura, “Partitioning of Web graphs by community topology,” in Proc. 14th Int. Conf. World Wide Web, May 2005, pp. 661–669.

[39] R. E. Gomory and T. C. Hu, “Multi-terminal network flows,” J. Soc. Ind. Appl. Math., vol. 9, no. 4, pp. 551–570, 1961.

[40] J. Reichardt and S. Bornholdt, “Statistical mechanics of community detection,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 74, no. 1, 2006, Art. no. 016110.

[41] M. B. Hastings, “Community detection as an inference problem,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 74, no. 3, 2006, Art. no. 035102.

[42] M. E. J. Newman, “Finding community structure in networks using the eigenvectors of matrices,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 74, no. 3, 2006, Art. no. 036104.

[43] F. Luo, J. Z. Wang, and E. Promislow, “Exploring local community structures in large networks,” Web Intell. Agent Syst., Int. J., vol. 6, no. 4, pp. 387–400, 2008.

[44] P. Pons and M. Latapy, “Computing communities in large networks using random walks,” in Proc. Int. Symp. Comput. Inf. Sci. Berlin, Germany: Springer, Oct. 2005, pp. 284–293.

[45] I. Vragovi´c and E. Louis, “Network community structure and loop coefficient method,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 74, no. 1, 2006, Art. no. 016105.

[46] V. Latora and M. Marchiori, “Efficient behavior of small-world net works,” Phys. Rev. Lett., vol. 87, Oct. 2001, Art. no. 198701. [47] U.

N. Raghavan, R. Albert, and S. Kumara, “Near linear time algorithm to detect community structures in large-scale networks,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 76, no. 3, 2007, Art. no. 036106.

[48] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” J. Stat. Mech., Theory Exp., vol.

2008, no. 10, 2008, Art. no. P10008. [49] C. Pizzuti, “Ga-net: A genetic algorithm for community detection in social networks,” in Proc. Int. Conf. Parallel Problem Solving Nature. Berlin, Germany: Springer, Sep. 2008, pp. 1081–1090.

[50] I. X. Y. Leung, P. Hui, P. Lio, and J. Crowcroft, “Towards real time community detection in large networks,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 79, no. 6, 2009, Art. no. 066107.

[51] S. Gregory, “Finding overlapping communities in networks by label propagation,” New J. Phys., vol. 12, no. 10, 2010, Art. no. 103018.

[52] J. Chen, O. Zaïane, and R. Goebel, “A visual data mining approach to find overlapping communities in networks,” in Proc. Int. Conf. Adv. Social Netw. Anal. Mining, Jul. 2009, pp. 338–343.

[53] X. Wang, L. Tang, H. Gao, and H. Liu, “Discovering overlapping groups in social media,” in Proc. IEEE Int. Conf. Data Mining, Dec. 2010, pp. 569–578.

[54] B. Wang, X. Wang, C. Sun, B. Liu, and L. Sun, “Modeling semantic relevance for question-answer pairs in Web social communities,” in

Proc. 48th Annu. Meeting Assoc. Comput. Linguistics, Jul. 2010, pp. 1230–1238.

[55] S. Parthasarathy, Y. Ruan, and V. Satuluri, “Community discovery in social networks: Applications, methods and emerging trends,” in

Social Network Data Analytics. Boston, MA, USA: Springer, 2011, pp. 79–113.

[55] S. Parthasarathy, Y. Ruan, and V. Satuluri, “Community discovery in social networks: Applications, methods and emerging trends,” in

Social Network Data Analytics. Boston, MA, USA: Springer, 2011, pp. 79–113.

[56] S. Papadopoulos, Y. Kompatsiaris, A. Vakali, and P. Spyridonos, “Com munity detection in social media,” Data Mining Knowl. Discovery, vol. 24, no. 3, pp. 515–554, May 2012.

[57] M. Plantié and M. Crampes, “Survey on social community detection,” in Social Media Retrieval. London, U.K.: Springer, 2013, pp. 65–85.

[58] J. Xie, S. Kelley, and B. K. Szymanski, “Overlapping community detection in networks: The state-of-the-art and comparative study,” ACM Comput. Surv., vol. 45, no. 4, p. 43, 2013.

[59] G. J. Qi, C. C. Aggarwal, and T. Huang, “Community detection with edge content in social media networks,” in Proc. IEEE 28th Int. Conf. Data Eng., Apr. 2012, pp. 534–545.

[60] J. Tang, X. Wang, and H. Liu, “Integrating social media data for community detection,” in Modeling and Mining Ubiquitous Social Media. Berlin, Germany: Springer, 2011, pp. 1–20.

[61] L. Weng, F. Menczer, and Y. Y. Ahn, “Virality prediction and com munity structure in social networks,” Sci. Rep., vol. 3, Aug. 2013, Art. no. 2522.

[62] S. P. Borgatti, “Centrality and network flow,” Social Netw., vol. 27, no. 1, pp. 55–71, 2005.

[63] S. P. Borgatti, K. M. Carley, and D. Krackhardt, “On the robustness of centrality measures under conditions of imperfect data,” Social Netw., vol. 28, no. 2, pp. 124–136, 2006.

[64] K. W. Axhausen, “Social networks, mobility biographies, and travel: Survey challenges,” Environ. Planning B, Planning Des., vol. 35, no. 6, pp. 981–996, 2008.

[65] Y. Zhou, E. Reid, J. Qin, H. Chen, and G. Lai, “US domestic extremist groups on the Web: Link and content analysis,” IEEE Intell. Syst., vol. 20, no. 5, pp. 44–51, Sep. 2005.

[66] A. Perer and B. Shneiderman, “Balancing systematic and flexible exploration of social networks,” IEEE Trans. Vis. Comput. Graphics, vol. 12, no. 5, pp. 693–700, Nov. 2006.

[67] K. L. Fiori, T. C. Antonucci, and K. S. Cortina, “Social network typologies and mental health among older adults,” J. Gerontol. B, Psychol. Sci. Social Sci., vol. 61, no. 1, pp. P25–P32, 2006.

[68] L. Backstrom, D. Huttenlocher, J. Kleinberg, and L. X. , “Group for mation in large social networks: Membership, growth, and evolution,” in Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2006, pp. 44–54.

[69] J. Voß, “Measuring wikipedia,” in Proc. ISSI 10th Int. Conf. Int. Soc. Scientometrics Informetrics, 2005.

[70] M. A. Sicilia, N. T. Korfiatis, M. Poulos, and G. Bokos, “Evalu ating authoritative sources using social networks: An insight from Wikipedia,” Online Inf. Rev., vol. 30, no. 3, pp. 252–262, 2006.

[71] H. E. Aldrich and P. H. Kim, “Small worlds, infinite possibilities? How social networks affect entrepreneurial team formation and search,”
Strategic Entrepreneurship J., vol. 1, nos. 1–2, pp. 147–165, 2007.

[72] C. Kiss and M. Bichler, “Identification of influencers—Measuring influence in customer networks,” Decis. Support Syst., vol. 46, no. 1, pp. 233–253, 2008.

[73] K. Okamoto, W. Chen, and X.-Y. Li, “Ranking of closeness centrality for large-scale social networks,” in Proc. Int. Workshop Frontiers Algorithmics. Berlin, Germany: Springer, Jun. 2008, pp. 186–195.

[74] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, “Arnetminer: Extraction and mining of academic social networks,” in Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2008, pp. 990–998.

[75] N. Eagle, A. S. Pentland, and D. Lazer, “Inferring friendship network structure by using mobile phone data,” Proc. Nat. Acad. Sci. USA, vol. 106, no. 36, pp. 15274–15278, 2009.

[76] E. Zheleva and L. Getoor, “To join or not to join: The illusion of privacy in social networks with mixed public and private user profiles,” in
Proc. 18th Int. Conf. World Wide Web, Apr. 2009, pp. 531–540.

[77] A. K. Pietiläinen, E. Oliver, J. LeBrun, G. Varghese, and C. Diot, “MobiClique: Middleware for mobile social networking,” in Proc. 2nd ACM Workshop Online Social Netw., Aug. 2009, pp. 49–54.

[78] I. A. Kovács, R. Palotai, M. S. Szalay, and P. Csermely, “Community landscapes: An integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics,” PLoS ONE, vol. 5, no. 9, 2010, Art. no. e12528.

[79] D. Centola, “The spread of behavior in an online social network experiment,” Science, vol. 329, no. 5996, pp. 1194–1197, 2010. [80] H.
Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao, “Detecting and characterizing social spam campaigns,” in Proc. 10th ACM SIGCOMM Conf. Internet Meas., Nov. 2010, pp. 35–47.

[81] D. Wang, D. Irani, and C. Pu, “A social-spam detection framework,” in Proc. 8th Annu. Collaboration, Electron. Messaging, Anti-Abuse Spam Conf., Sep. 2011, pp. 46–54.

[82] M. Fire, G. Katz, and Y. Elovici, “Strangers intrusion detection detecting spammers and fake profiles in social networks based on topology anomalies,” Hum. J., vol. 1, no. 1, pp. 26–39, 2012.

[83] M. D. Roblyer, M. McDaniel, M. Webb, J. Herman, and J. V. Witty, “Findings on Facebook in higher education: A comparison of college faculty and student uses and perceptions of social networking sites,” Internet Higher Educ., vol. 13, no. 3, pp. 134–140, 2010.

[84] A. S. Bozkır, S. G. Mazman, and E. A. Sezer, “Identification of user patterns in social networks by data mining techniques: Facebook case,” in Proc. Int. Symp. Inf. Manage. Changing World. Berlin, Germany: Springer, Sep. 2010, pp. 145–153.

[85] Y. Wang, G. Cong, G. Song, and K. Xie, “Community-based greedy algorithm for mining top-k influential nodes in mobile social networks,” in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Jul. 2010, pp. 1039–1048.

[86] E. Raad, R. Chbeir, and A. Dipanda, “User profile matching in social networks,” in Proc. 13th Int. Conf. Netw.-Based Inf. Syst., Sep. 2010, pp. 297–304.

[87] H. Kwak, C. Lee, H. Park, and S. Moon, “What is Twitter, a social network or a news media?” in Proc. 19th Int. Conf. World Wide Web, Apr. 2010, pp. 591–600.

[88] C. Yang, R. Harkreader, J. Zhang, S. Shin, and G. Gu, “Analyzing spammers’ social networks for fun and profit: A case study of cyber criminal ecosystem on Twitter,” in Proc. 21st Int. Conf. World Wide Web, Apr. 2012, pp. 71–80.

[89] C.-L. Hsu and H. W. Park, “Mapping online social networks of Korean politicians,” Government Inf. Quart., vol. 29, no. 2, pp. 169–181, 2012. [90] D. Hansen, B. Shneiderman, and M. A. Smith, Analyzing Social Media Networks With NodeXL: Insights From a Connected World. San Mateo, CA, USA: Morgan Kaufmann, 2010.

[91] D. L. Hansen et al., “Do you know the way to SNA?: A process model for analyzing and visualizing social media network data,” in Proc. Int. Conf. Social Inf., Dec. 2012, pp. 304–313.

[92] F. Zhao and A. K. H. Tung, “Large scale cohesive subgraphs discovery for social network visual analysis,” Proc. VLDB Endowment, vol. 6, no. 2, pp. 85–96, 2012.

[93] B. Sun and V. T. Ng, “Identifying influential users by their postings in social networks,” in Ubiquitous Social Media Analysis. Berlin, Germany: Springer, 2012, pp. 128–151.

[94] S. He, X. Zheng, D. Zeng, K. Cui, Z. Zhang, and C. Luo, “Identifying peer influence in online social networks using transfer entropy,” in
Proc. Pacific-Asia Workshop Intell. Security Inform. Berlin, Germany: Springer, Aug. 2013, pp. 47–61.

[95] A. Javari and M. Jalili, “Cluster-based collaborative filtering for sign prediction in social networks with positive and negative links,” ACM Trans. Intell. Syst. Technol., vol. 5, no. 2, p. 24, 2014.

[96] F. Liu, B. Liu, C. Sun, M. Liu, and X. Wang, “Deep belief network based approaches for link prediction in signed social networks,”
Entropy, vol. 17, no. 4, pp. 2140–2169, 2015.

[97] N. Phan, D. Dou, B. Piniewski, and D. Kil, “Social restricted boltzmann machine: Human behavior prediction in health social networks,” in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining, Aug. 2015, pp. 424–431.

[98] V. Agarwal and K. K. Bharadwaj, “Predicting the dynamics of social circles in ego networks using pattern analysis and GA K-means clustering,” Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery, vol. 5, no. 3, pp. 113–141, 2015.

[99] K. W. Lim, C. Chen, and W. Buntine, “Twitter-network topic model: A full Bayesian treatment for social network and text mod eling,” 2016, arXiv:1609.06791. [Online]. Available: https://arxiv.org/ abs/1609.06791

[100] D. T. Nguyen, K. A. A. Mannai, S. Joty, H. Sajjad, M. Imran, and P. Mitra, “Robust classification of crisis-related data on social networks using convolutional neural networks,” in Proc. 11th Int. AAAI Conf. Web Social Media, May 2017, pp. 1–4.

[101] R. Rabade, N. Mishra, and S. Sharma, “Survey of influential user identification techniques in online social networks,” in Recent Advances in Intelligent Informatics. Cham, Switzerland: Springer, 2014, pp. 359–370.

[102] D. C. Cercel and T.-M. Stefan, “Opinion propagation in online social networks: A survey,” in Proc. 4th Int. Conf. Web Intell., Mining, Semantics (WIMS), Jun. 2014, p. 11.

[103] M. Adedoyin-Olowe, M. M. Gaber, and F. Stahl, “A survey of data mining techniques for social media analysis,” 2013, arXiv:1312.4617. [Online]. Available: https://arxiv.org/abs/1312.4617

[104] J. Bian, Y. Liu, E. Agichtein, and H. Zha, “Finding the right facts in the crowd: Factoid question answering over social media,” in Proc. 17th Int. Conf. World Wide Web, Apr. 2008, pp. 467–476.

[105] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne, “Finding high-quality content in social media,” in Proc. Int. Conf. Web Search Data Mining, Feb. 2008, pp. 183–194.

[106] K. Denecke and W. Nejdl, “How valuable is medical social media data? Content analysis of the medical Web,” Inf. Sci., vol. 179, no. 12, pp. 1870–1880, 2009.

[107] F. Chen, P. N. Tan, and A. K. Jain, “A co-classification framework for detecting Web spam and spammers in social media Web sites,” in
Proc. 18th ACM Conf. Inf. Knowl. Manage., Nov. 2009, pp. 1807–1810.

[108] B. Markines, C. Cattuto, and F. Menczer, “Social spam detection,” in Proc. 5th Int. Workshop Adversarial Inf. Retr. Web, Apr. 2009, pp. 41–48.

[109] H. Sayyadi, M. Hurst, and A. Maykov, “Event detection and tracking in social streams,” in Proc. 3rd Int. AAAI Conf. Weblogs Social Media, Mar. 2009, pp. 1–4.

[110] H. Becker, M. Naaman, and L. Gravano, “Event identification in social media,” in Proc. WebDB, Jun. 2009, pp. 1–6.

[111] R. W. Lariscy, E. J. Avery, K. D. Sweetser, and P. Howes, “An examination of the role of online social media in journalists’ source mix,”
Public Relations Rev., vol. 35, no. 3, pp. 314–316, 2009.

[112] C. M. De, Y. R. Lin, H. Sundaram, K. S. Candan, L. Xie, and A. Kelliher, “How does the data sampling strategy impact the discovery of information diffusion in social media?” in Proc. 4th

Int. AAAI Conf. Weblogs Social Media, May 2010, pp. 1–8.

[113] H. Becker, M. Naaman, and L. Gravano, "Learning similarity metrics for event identification in social media," in Proc. 3rd ACM Int. Conf. Web Search Data Mining, Feb. 2010, pp. 291–300.

[114] O. Oh, K. H. Kwon, and H. R. Rao, "An exploration of social media in extreme events: Rumor theory and Twitter during the haiti earthquake 2010," in Proc. ICIS, vol. 231, Dec. 2010, pp. 7332–7336.

[115] M. Naaman, J. Boase, and C. H. Lai, "Is it really about me?: Message content in social awareness streams," in Proc. ACM Conf. Comput. Supported Cooperat. Work, Feb. 2010, pp. 189–192.

[116] S. Asur and B. A. Huberman, "Predicting the future with social media," in Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol., vol. 1, Aug. 2010, pp. 492–499.

[117] N. Diakopoulos, M. Naaman, and F. Kivran-Swaine, "Diamonds in the rough: Social media visual analytics for journalistic inquiry," in Proc. IEEE Symp. Visual Anal. Sci. Technol., Oct. 2010, pp. 115–122.

[118] Z. Xiang and U. Gretzel, "Role of social media in online travel information search," Tourism Manage., vol. 31, no. 2, pp. 179–188, 2010.

[119] W. C. Jacobsen and R. Forste, "The wired generation: Academic and social outcomes of electronic media use among University students," Cyberpsychol., Behav., Social Netw., vol. 14, no. 5, pp. 275–280, 2011.

[120] K. Lee, J. Caverlee, Z. Cheng, and D. Z. Sui, "Content-driven detection of campaigns in social media," in Proc. 20th ACM Int. Conf. Inf. Knowl. Manage., Oct. 2011, pp. 551–556.

[121] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman, "Influence and passivity in social media," in Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases. Berlin, Germany: Springer, Sep. 2011, pp. 18–33.

[122] N. Michaelidou, N. T. Siamagka, and G. Christodoulides, "Usage, barriers and measurement of social media marketing: An exploratory investigation of small and medium B2B brands," Ind. Marketing Manage., vol. 40, no. 7, pp. 1153–1159, 2011.

[123] K. Sedereviciute and C. Valentini, "Towards a more holistic stake holder analysis approach. Mapping known and undiscovered stake holders from social media," Int. J. Strategic Commun., vol. 5, no. 4, pp. 221–239, 2011.

[124] F. Cheong and C. Cheong, "Social media data mining: A social network analysis of tweets during the 2010–2011 Australian floods," PACIS, vol. 11, p. 46, Jul. 2011.

[125] E. Clark and K. Araki, "Text normalization in social media: Progress, problems and applications for a pre-processing system of casual English," Procedia-Social Behav. Sci., vol. 27, pp. 2–11, Jan. 2011.

[126] C. C. Yang, L. Jiang, H. Yang, and X. Tang, "Detecting signals of adverse drug reactions from health consumer contributed content in social media," in Proc. ACM SIGKDD Workshop Health Inf., Aug. 2012, pp. 1–8.

[127] J. Tang and H. Liu, "Feature selection with linked data in social media," in Proc. SIAM Int. Conf. Data Mining, Apr. 2012, pp. 118–128.

[128] L. M. Aiello, A. Barrat, R. Schifanella, C. Cattuto, B. Markines, and F. Menczer, "Friendship prediction and homophily in social media," ACM Trans. Web, vol. 6, no. 2, p. 9, 2012.

[129] B. Han, P. Cook, and T. Baldwin, "Geolocation prediction in social media data by finding location indicative words," in Proc. COLING, Dec. 2012, pp. 1045–1062.

[130] G. Ver Steeg and A. Galstyan, "Information transfer in social media," in Proc. 21st Int. Conf. World Wide Web, Apr. 2012, pp. 509–518.

[131] C. Sengstock and M. Gertz, "Latent geographic feature extraction from social media," in Proc. 20th Int. Conf. Adv. Geograph. Inf. Syst., Nov. 2012, pp. 149–158.

[132] J. Cranshaw, R. Schwartz, J. Hong, and N. Sadeh, "The livehoods project: Utilizing social media to understand the dynamics of a city," in Proc. 6th Int. AAAI Conf. Weblogs Social Media, May 2012, pp. 1–8.

[133] S. A. Moorhead, D. E. Hazlett, L. Harrison, J. K. Carroll, A. Irwin, and C. Hoving, "A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication," J. Med. Internet Res., vol. 15, no. 4, p. e85, 2013.

[134] H. A. Schwartz et al., "Personality, gender, and age in the language of social media: The open-

vocabulary approach," PLoS ONE, vol. 8, no. 9, p. e73791, 2013.

[135] M. Kandias, V. Stavrou, N. Bozovic, and D. Gritzalis, "Proactive insider threat detection through social media: The YouTube case," in Proc. 12th ACM Workshop Privacy Electron. Soc., Nov. 2013, pp. 261–266.

[136] M. Thelwall, "The Heart and soul of the Web? Sentiment strength detection in the social Web with SentiStrength," in Cyberemotions. Cham, Switzerland: Springer, 2013, pp. 119–134. [137] F. Morstatter, J. Pfeffer, H. Liu, and K. M. Carley, "Is the sample good enough? Comparing data from twitter's streaming api with Twitter's firehose," in Proc. 7th Int. AAAI Conf. Weblogs Social Media, Jun. 2013, pp. 1–9.

[138] Y. Hu, S. D. Farnham, and A. Monroy-Hernández, "Whoo. ly: Facil itating information seeking for hyperlocal communities using social media," in Proc. SIGCHI Conf. Hum. Factors Comput. Syst., Apr. 2013, pp. 3481–3490.

[139] F. Mitzlaff, M. Atzmueller, G. Stumme, and A. Hotho, "Semantics of user interaction in social media," in Complex Networks IV. Berlin, Germany: Springer, 2013, pp. 13–25.

[140] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in Proc. 7th Int. AAAI Conf. Weblogs Social Media, Jun. 2013, pp. 1–10.

[141] W. He, S. Zha, and L. Li, "Social media competitive analysis and text mining: A case study in the pizza industry," Int. J. Inf. Manage., vol. 33, no. 3, pp. 464–472, 2013.

[142] K. Casler, L. Bickel, and E. Hackett, "Separate but equal? A compar ison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing," Comput. Hum. Behav., vol. 29, pp. 2156–2160, Nov. 2013.

[143] T. Baldwin, P. Cook, M. Lui, A. MacKinlay, and L. Wang, "How noisy social media text, how diffrnt social media sources?" in Proc. 6th Int. Joint Conf. Natural Lang. Process., Oct. 2013, pp. 356–364.

[144] A. Majid, L. Chen, G. Chen, H. T. Mirza, and I. W. J. Hussain, "A context-aware personalized travel recommendation system based on geotagged social media data mining," Int. J. Geograph. Inf. Sci., vol. 27, no. 4, pp. 662–684, 2013.

[145] M. Khan, M. Dickinson, and S. Kübler, "Does size matter? Text and grammar revision for parsing social media data," in Proc. Workshop Lang. Anal. Social Media, 2013, pp. 1–10.

[146] L. Derczynski, A. Ritter, S. Clark, and K. Bontcheva, "Twitter part of-speech tagging for all: Overcoming sparse and noisy data," in Proc. Int. Conf. Recent Adv. Natural Lang. Process. (RANLP), 2013, pp. 198–206.

[147] Y. Zhang and M. Pennacchiotti, "Recommending branded products from social media," in Proc. 7th ACM Conf. Recommender Syst., Oct. 2013, pp. 77–84.

[148] X. Ma, H. Wang, H. Li, J. Liu, and H. Jiang, "Exploring sharing patterns for video recommendation on YouTube-like social media," Multimedia Syst., vol. 20, no. 6, pp. 675–691, 2014.

[149] S. Pei, L. Muchnik, J. S. Andrade, Jr., Z. Zheng, and H. A. Makse, "Searching for superspreaders of information in real-world social media," Sci. Rep., vol. 4, Jul. 2014, Art. no. 5547.

[150] M. Abdul-Mageed, M. Diab, and S. Kübler, "SAMAR: Subjectivity and sentiment analysis for Arabic social media," Comput. Speech Lang., vol. 28, no. 1, pp. 20–37, 2014.

[151] S. Mei, H. Li, J. Fan, X. Zhu, and C. R. Dyer, "Inferring air pollution by sniffing social media," in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining, Aug. 2014, pp. 534–539.

[152] J. Choi et al., "The placing task: A large-scale geo-estimation challenge for social-media videos and images," in Proc. 3rd ACM Multimedia Workshop Geotagging Appl. Multimedia, Nov. 2014, pp. 27–31.

[153] M. Sap et al., "Developing age and gender predictive lexica over social media," in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), Oct. 2014, pp. 1146–1151.

[154] I. C. L. Memon, A. Majid, M. Lv, and I. C. G. Hussain, "Travel recommendation using geo-tagged photos in social media for tourist," Wireless Pers. Commun., vol. 80, no. 4, pp. 1347–1362, 2015.

[155] D. Bamman, J. Eisenstein, and T. Schnoebelen, "Gender identity and lexical variation

in social media," J. Sociolinguistics, vol. 18, no. 2, pp. 135–160, 2014.

[156] H. Lin et al., "User-level psychological stress detection from social media using deep neural network," in Proc. 22nd ACM Int. Conf. Multimedia, Nov. 2014, pp. 507–516.

[157] R. Nivedha and N. Sairam, "A machine learning based classification for social media messages," Indian J. Sci. Technol., vol. 8, no. 16, p. 1, 2015.

[158] A. P. López-Monroy, M. Montes-y-Gómez, H. J. Escalante, L. Villaseñor-Pineda, and E. Stamatatos, "Discriminative subprofile specific representations for author profiling in social media," Knowl.- Based Syst., vol. 89, pp. 134–147, Nov. 2015.

[159] P. Barberá, "Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data," Political Anal., vol. 23, no. 1, pp. 76–91, 2015.

[160] C. Cao and J. Caverlee, "Detecting spam URLs in social media via behavioral analysis," in Proc. Eur. Conf. Inf. Retr. Cham, Switzerland: Springer, Mar. 2015, pp. 703–714.

[161] T. Ma et al., "Social network and tag sources based augmenting collaborative recommender system," IEICE Trans. Inf. Syst., vol. 98, no. 4, pp. 902–910, 2015.

[162] M. Santillana, A. T. Nguyen, M. Dredze, M. J. Paul, E. O. Nsoesie, and J. S. Brownstein, "Combining search, social media, and traditional data sources to improve influenza surveillance," PLoS Comput. Biol., vol. 11, no. 10, 2015, Art. no. e1004513.

[163] F. Gelli, T. Uricchio, M. Bertini, A. Del Bimbo, and S.-F. Chang, "Image popularity prediction in social media using sentiment and context features," in Proc. 23rd ACM Int. Conf. Multimedia, Oct. 2015, pp. 907–910.

[164] W. Chen, C. K. Yeo, C. T. Lau, and B. S. Lee, "Real-time Twitter content polluter detection based on direct features," in Proc. 2nd Int. Conf. Inf. Sci. Security (ICISS), Dec. 2015, pp. 1–4.

[165] J. Ma et al., "Detecting rumors from microblogs with recurrent neural networks," in Proc. IJCAI, Jul. 2016, pp. 3818–3824.

[166] V. Kulkarni, B. Perozzi, and S. Skiena, "Freshman or fresher? Quanti fying the geographic variation of language in online social media," in Proc. 10th Int. AAAI Conf. Web Social Media, Mar. 2016, pp. 1–6.

[167] B. Bischke, D. Borth, C. Schulze, and A. Dengel, "Contextual enrich ment of remote-sensed events with social media streams," in Proc. 24th ACM Int. Conf. Multimedia, Oct. 2016, pp. 1077–1081.

[168] L. Liu, D. Preotiuc-Pietro, Z. R. Samani, M. E. Moghaddam, and L. Ungar, "Analyzing personality through social media profile picture choice," in Proc. 10th Int. AAAI Conf. Web Social Media, Mar. 2016, pp. 1–10.

[169] J. P. Mazer et al., "Communication in the face of a school crisis: Examining the volume and content of social media mentions during active shooter incidents," Comput. Hum. Behav., vol. 53, pp. 238–248, Dec. 2015.

[170] B. Dhingra, Z. Zhou, D. Fitzpatrick, M. Muehl, and W. W. Cohen, "Tweet2vec: Character-based distributed representations for social media," 2016, arXiv:1605.03481. [Online]. Available: https://arxiv. org/abs/1605.03481

[171] L. Song, R. Y. K. Lau, R. C. W. Kwok, K. Mirkovski, and W. Dou, "Who are the spoilers in social media marketing? Incremental learning of latent semantics for social spam detection," Electron. Commerce Res., vol. 17, no. 1, pp. 51–81, 2017.

[172] F. Krebs, B. Lubascher, T. Moers, P. Schaap, and G. Spanakis, "Social emotion mining techniques for Facebook posts reaction pre diction," 2017, arXiv:1712.03249. [Online]. Available: https://arxiv. org/abs/1712.03249

[173] J. Su, A. Shukla, S. Goel, and A. Narayanan, "De-anonymizing Web browsing data with social networks," in Proc. 26th Int. Conf. World Wide Web, Apr. 2017, pp. 1261–1269.

[174] H. Aleid et al., "Framework to classify and analyze social media content," Social Netw., vol. 7, no. 2, p. 79, 2018.

[175] F. Aisopos, G. Papadakis, and T. Varvarigou, "Sentiment analysis of social media content using N-Gram graphs," in Proc. 3rd ACM SIGMM Int. Workshop Social Media, Nov. 2011, pp. 9–14.

[176] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon based methods for sentiment analysis," Comput. Linguistics, vol. 37, no. 2, pp. 267–307, 2011.

[177] I. Habernal, T. Ptácek, and J. Steinberger, "Reprint of 'supervised sentiment analysis in Czech

social media," Inf. Process. Manage., vol. 51, no. 4, pp. 532–546, 2015.

[178] K. P. Murphy, "Naive Bayes classifiers," Univ. Brit. Columbia, Vancouver, BC, Canada, Tech. Rep., 2006, p. 60, vol. 18. [179] M. Salathé and S. Khandelwal, "Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control," PLoS Comput. Biol., vol. 7, no. 10, 2011, Art. no. e1002199. [180] Y. Lu, X. Hu, F. Wang, S. Kumar, H. Liu, and R. Maciejewski, "Visualizing social media sentiment in disaster scenarios," in Proc. 24th Int. Conf. World Wide Web, May 2015, pp. 1211–1215. [181] R. Baracho, M. Bax, L.

G. F. Ferreira, and G. Ca'ires Silva, Sentiment Analysis in Social Networks. Stanford, CA, USA: Association for the Advancement of Artificial Intelligence, 2012.

[182] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in Mining Text Data. Boston, MA, USA: Springer, 2012, pp. 415–463.

[183] A. Candelieri and F. Archetti, "Detecting events and sentiment on Twitter for improving urban mobility," in Proc. ESSEM AAMAS, May 2015, pp. 106–115.

[184] C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in Proc. 8th Int. AAAI Conf. Weblogs Social Media, May 2014, pp. 1–10.

[185] T. Chen, D. Borth, T. Darrell, and S. F. Chang, "Deepsen tiBank: Visual sentiment concept classification with deep convolu tional neural networks," 2014, arXiv:1410.8586. [Online]. Available: https://arxiv.org/abs/1410.8586

[186] Q. You, J. Luo, H. Jin, and J. Yang, "Robust image sentiment analysis using progressively trained and domain transferred deep networks," in Proc. 29th AAAI Conf. Artif. Intell., Feb. 2015, pp. 1–8.

[187] J. Yuan, Q. You, and J. Luo, "Sentiment analysis using social multi media," in Multimedia Data Mining and Analytics. Cham, Switzerland: Springer, 2015, pp. 31–59.

[188] A. Hanjalic, C. Kofler, and M. Larson, "Intent and its discontents: The user at the wheel of the online video search engine," in Proc. 20th ACM Int. Conf. Multimedia, Oct. 2012, pp. 1239–1248.

[189] D. Cao, R. Ji, D. Lin, and S. Li, "A cross-media public sentiment analysis system for microblog," Multimedia Syst., vol. 22, no. 4, pp. 479–486, 2016.

[190] W. He, H. Wu, G. Yan, V. Akula, and J. Shen, "A novel social media competitive analytics framework with sentiment benchmarks," Inf. Manage., vol. 52, no. 7, pp. 801–812, 2015.

[191] D. Zimbra, M. Ghiassi, and S. Lee, "Brand-related Twitter sentiment analysis using feature engineering and the dynamic architecture for artificial neural networks," in Proc. 49th Hawaii Int. Conf. Syst. Sci. (HICSS), Jan. 2016, pp. 1930–1938.

[192] B. Jou, T. Chen, N. Pappas, M. Redi, M. Topkara, and S. F. Chang, "Visual affect around the world: A large-scale multilingual visual sen timent ontology," in Proc. 23rd ACM Int. Conf. Multimedia, Oct. 2015, pp. 159–168.

[193] T. Niu, S. Zhu, L. Pang, and A. E. Saddik, "Sentiment analysis on multi-view social data," in Proc. Int. Conf. Multimedia Modeling. Cham, Switzerland: Springer, Jan. 2016, pp. 15–27.

[194] J. S. Jang, B. I. C. H. Lee Choi, J. H. Kim, D. M. Seo, and W. S. Cho, "Understanding pending issue of society and sentiment analysis using social media," in Proc. 8th Int. Conf. Ubiquitous Future Netw. (ICUFN), Jul. 2016, pp. 981–986.

[195] F. H. Khan, U. Qamar, and S. Bashir, "SentiMI: Introducing point-wise mutual information with SentiWordNet to improve sentiment polarity detection," Appl. Soft Comput., vol. 39, pp. 140–153, Feb. 2016.

[196] O. F. Villarroel, S. Ludwig, R. K. De, D. Grewal, and M. Wetzels, "Unveiling what is written in the stars: Analyzing explicit, implicit, and discourse patterns of sentiment in social media," J. Consum. Res., vol. 43, no. 6, pp. 875–894, 2017.

[197] S. Poria, E. Cambria, D. Hazarika, and P. Vij, "A deeper look into sarcastic tweets using deep convolutional neural networks," 2016, arXiv:1610.08815. [Online]. Available: https://arxiv.org/abs/ arXiv:1610.08815

[198] L. Wang and Z. Zhang, Support Vector Machines: Theory and Appli cations. Berlin, Germany: Springer-Verlag, 2005.

[199] B. Liu, Sentiment Analysis and Opinion Mining (Synthesis Lectures on Human Language Technologies), vol. 5, no. 1. Ras al Khaimah, United Arab Emirates: Science Publishing

Corporation, RAK Free Trade Zone, 2012, pp. 1–167.

[200] D. M. E.-D. M. Hussein, “A survey on sentiment analysis challenges,” J. King Saud Univ.-Eng. Sci., vol. 34, no. 4, pp. 330–338, 2016.

[201] S. Behdenna, F. Barigou, and G. Belalem, “Document level sentiment analysis: A survey,” EAI Endorsed Trans. Context-Aware Syst. Appl., vol. 4, Mar. 2018, Art. no. 154339, doi: 10.4108/eai.14-3-2018.154339. [202] V. S. Jagtap and K. Pawar, “Analysis of different approaches to sentence-level sentiment classification,” Int. J. Sci. Eng. Technol., vol. 2, no. 3, pp. 164–170, 2013.

[203] K. Schouten and F. Frasincar, “Survey on aspect-level sentiment analysis,” IEEE Trans. Knowl. Data Eng., vol. 28, no. 3, pp. 813–830, Oct. 2015.

[204] M. K. Dalal and M. A. Zaveri, “Opinion mining from online user reviews using fuzzy linguistic hedges,” Appl. Comput. Intell. Soft Comput., vol. 2014, p. 2, Jan. 2014.

[205] P. Gonçalves, M. Araújo, F. Benevenuto, and M. Cha, “Comparing and combining sentiment analysis methods,” in Proc. 1st ACM Conf. Online Social Netw., Oct. 2013, pp. 27–38.

[206] F. Å. Nielsen, “A new ANEW: Evaluation of a word list for sentiment analysis in microblogs,” 2011, arXiv:1103.2903. [Online].
Available: https://arxiv.org/abs/1103.2903

[207] A. Hogenboom, B. Heerschop, F. Frasincar, U. Kaymak, and F. de Jong, “Multi-lingual support for lexicon-based sentiment analysis guided by semantics,” Decis. Support Syst., vol. 62, pp. 43–53, Jun. 2014.

[208] P. Ray and A. Chakrabarti, “Twitter sentiment analysis for product review using lexicon method,” in Proc. Int. Conf. Data Manage., Anal. Innov. (ICDMAI), Feb. 2017, pp. 211–216.

[209] R. Moraes, J. F. Valiati, and W. P. G. Neto, “Document-level sentiment classification: An empirical comparison between SVM and ANN,”
Expert Syst. Appl., vol. 40, no. 2, pp. 621–633, 2013.

[210] B. Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. New York, NY, USA: Elsevier, 2007.

[211] J. Smailovi´c, M. Grˇcar, N. Lavraˇc, and M. Žnidarši ˇc, “Stream-based active learning for sentiment analysis in the financial domain,” Inf. Sci., vol. 285, pp. 181–203, Nov. 2014.

[212] A. Hasan, S. Moin, A. Karim, and S. Shamshirband, “Machine learning-based sentiment analysis for Twitter accounts,” Math. Comput. Appl., vol. 23, no. 1, p. 11, 2018.

[213] P. D. Turney, “Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews,” in Proc. 40th Annu. Meeting Assoc. Comput. Linguistics, Jul. 2002, pp. 417–424.

[214] F. Bravo-Marquez, M. Mendoza, and B. Poblete, “Meta-level sentiment models for big social data analysis,” Knowl.-Based Syst., vol. 69, pp. 86–99, Oct. 2014.

[215] M. Ahmad, S. Aftab, I. Ali, and N. Hameed, “Hybrid tools and techniques for sentiment analysis: A review,” Int. J. Multidiscip. Sci. Eng., vol. 8, no. 3, pp. 1–6, 2017.

[216] K. Elshakankery and M. F. Ahmed, “HILATSA: A hybrid incremental learning approach for Arabic tweets sentiment analysis,” Egyptian Inform. J., to be published.

[217] Q. T. Ain et al., “Sentiment analysis using deep learning techniques: A review,” Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 6, p. 424, 2017.

[218] A. Kalaivani and D. Thenmozhi, “Sentiment analysis using deep learning techniques,” Int. J. Recent Technol. Eng., vol. 7, no. 6S5, pp. 1–7, 2019.

[219] A. M. El-Gazzar, T. M. Mohamed, and R. A. Sadek, “A hybrid SVD HSV visual sentiment analysis system,” in Proc. 8th Int. Conf. Intell. Comput. Inf. Syst. (ICICIS), Dec. 2017, pp. 360–365.

[220] H. Zhang, A. C. Berg, M. Maire, and J. Malik, “SVM-KNN: Discrim inative nearest neighbor classification for visual category recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., vol. 2, Jun. 2006, pp. 2126–2136.

[221] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, “Coding facial expressions with gabor wavelets,” in Proc. 3rd IEEE Int. Conf. Autom. Face Gesture Recognit., Apr. 1998, pp. 200–205.

[222] I. S. P. James, “Face image retrieval with HSV color space using clustering techniques,” SIJ

Trans. Comput. Sci. Eng. Appl., vol. 1, no. 1, pp. 1–4, 2013.

[223] K. Mahboob and F. Ali, “Sentiment analysis of pharmaceutical products evaluation based on customer review mining,” J. Comput. Sci. Syst. Biol., vol. 11, pp. 190–194, Mar. 2018, doi: 10.4172/jcsb.1000271.

[224] G. Liang, W. He, C. Xu, L. Chen, and J. Zeng, “Rumor identification in microblogging systems based on users’ behavior,” IEEE Trans. Computat. Social Syst., vol. 2, no. 3, pp. 99–108, Sep. 2015.

[225] R. Basak, S. Sural, N. Ganguly, and S. K. Ghosh, “Online public shaming on Twitter: Detection, analysis, and mitigation,” IEEE Trans. Comput. Social Syst., vol. 6, no. 2, pp. 208–220, Apr. 2019.

[226] S. Madisetty and M. S. Desarkar, “A neural network-based ensemble approach for spam detection in Twitter,” IEEE Trans. Comput. Social Syst., vol. 5, no. 4, pp. 973–984, Dec. 2018.

[227] H. Tajalizadeh and R. Boostani, “A novel stream clustering framework for spam detection in Twitter,” IEEE Trans. Comput. Social Syst., vol. 6, no. 3, pp. 525–534, Jun. 2019, doi: 10.1109/ TCSS.2019.2910818.

[228] Y. Tyshchuk and W. A. Wallace, “Modeling human behavior on social media in response to significant events,” IEEE Trans. Comput. Social Syst., vol. 5, no. 2, pp. 444–457, Jun. 2018.

[229] R. C. Chen, “User rating classification via deep belief network learning and sentiment analysis,” IEEE Trans. Comput. Social Syst., vol. 6, no. 3, pp. 535–546, Jun. 2019.

[230] T. Chowdhury, S. Muhuri, S. Chakraborty, and S. N. Chakraborty, “Analysis of adapted films and stories based on social network,” IEEE Trans. Comput. Social Syst., vol. 6, no. 5, pp. 858–869, Oct. 2019.

[231] K. Chakraborty, S. Bhattacharyya, R. Bag, and A. E. Hassanien, “Comparative sentiment analysis on a set of movie reviews using deep learning approach,” in Proc. Int. Conf. Adv. Mach. Learn. Technol. Appl. (AMLTA), Egypt, Cairo, Feb. 2018, pp. 311–318.

[232] K. Chakraborty, S. Bhattacharyya, R. Bag, and A. E. Hassanien, “Sentiment analysis on a set of movie reviews using deep learning techniques,” in Social Network Analytics—Computational Research Methods and Techniques. Amsterdam, The Netherlands: Elsevier, 2018.

[233] K. Chakraborty, R. Bag, and S. Bhattacharyya, “Relook into sentiment analysis performed on Indian languages using deep learning,” in Proc. 4th Int. Conf. Res. Comput. Intell. Commun. Netw. (ICRCICN), Nov. 2018, pp. 208–213.