# The spider-man behavior protocol: exploring both public and dark social networks for fake identity detection in terrorism informatics

Matteo Cristani, Elisa Burato, Katia Santacá, and Claudio Tomazzoli

University of Verona
{matteo.cristani, elisa.burato, katia.santaca,
claudio.tomazzoli}@univr.it

**Abstract.** Hiding true personality behind a façade is one of the basic tricks adopted by humans who live double lives for illegal purposes. In particular terrorists have historically adopted the protocol of a façade behaviour coupled with a second life consisting mainly in illegal activities and their planning.
Nowadays a few cases of behaviours that hide a dangerous activity, possibly illegal, behind an apparentely neutral and mean public person, can be replicated, and sometimes just provided, by a social network profile. Recognizing that a social network profile is fake, in some extreme cases, a bot, and determining the contour relationships that limit such a condition is one of the most important weapons for terrorism fight.
In this paper we show that what we name the *Spider-man protocol*, a set of behaviour rules that bring to hiding a personality behind a façade, has several weaknesses, and it is prone to a set of attacks that permit to detect these behaviours. We provide the description of an experimental architecture that is used for determining violations of the protocol, and therefore breaches in the secrecy of the individual protection settled by the terrorists.

## 1 Introduction

In the recent past, it has been found that the web is also being used as a tool by radical or extremist groups and users to practice several kinds of mischievous acts with concealed agendas and promote ideologies in a sophisticated manner, so that several studies have been performed on how to understand and identify tension or deviant behaviors before these can lead to acts of terrorism. These investigations are paired by those studies, especially in the information security research area, that aim at determining cases of phishing, where people are showing off themselves as individuals different then they are, to obtain illegal profits.

To the best of our knowledge, however, only a few investigations have been carried out that combine these two aspects. It is clear that, when someone passes the border and becomes a terrorist, there is an observable phase in which part

of her life is still public though partly hidden, whilst after this phase that person becomes invisible. A few behaviours can be classified that correspond to *become clandestine* for illegal purposes, and, on the other hand, there are a few behaviours that can make such condition disclosed.

In this paper we study the ways in which the aforementioned transition happens (Section 4.1), how you can provide the recognition of a breach in such a protocol (Section 4.2) and present an architecture to deal with such a recognition need (Section 5). Before to do so, we need to model the behaviours (Section 2) and introduce a method, that is the extension of an existing approach [3] to more general cases (Section 3). At the end of the above presented studies, we review the recent literature (Section 6) and finally introduce some further perspective (Section 7).

## 2 How do terrorists behave on the web?

The majority of radical and extremist groups do not appear as regular individuals on the public web. They typically hide themselves under the level of publicly mapped web sites, the so-called Deep Web. This area of the web, often also known as *invisible web* is essentially identical to the visible part, aside from the lack of association of the web addresses to the web spiders of Google and other search engines. Within the Deep Web umbrella, many of those individuals interact in a social network that is totally hidden to the public web, the so-called darknets. This area of the web is known as *Dark Web*.

There are many hidden social networks, including, in particular, the well-known *AnonPlus* secret network, or the less known but very important *Dark-NetMarket*, used to interact in the Dark Web by criminals, including drug marketers, pedophiles, terrorists. The majority of these web sites need specific tools to be used, as, for instance *Tor*, *Freenet* or *I2P*, and employ specific P2P protocol methods, including the files used for the specific P2P purpose, the *.onion* ones. The public part of the web is also referred to, in particular by the users of the Darknets, as *the Cleranet*.

The notion of a terrorist used in the current literature is that he is an individual who is acting in a public environment and secretly fighting for a social, political, religious, ethnic, or national cause. This definition implies that a terrorist can be in one of the following general conditions:

- *Fully clandestine*, the condition of terrorists acting completely on the secrecy, hidden in a place where they cannot be found. This is the case, for instance, of Al-Qaeda in certain areas, like Europe and the United States.
- *Rebel*, when a fighting individual lives separately from the counterpart, in a publicly known area, but protected by an openly fighting group. ISIS is acting in this way.
- *Double living*, when they act publicly as apparently harmless people, whilst living a second life of active fighters for some causes. This is the way in which Al-Qaeda members act in the same areas where fully clandestine members also exist.

## 3 Detecting terrorists: social media and dark web analysis

The basis of the web analysis we provide is a twofold approach: we aim at tracing individuals who act under the umbrella of the Dark Web in double living style as defined in Section 2. We trace individuals in the Darknets, and individuals in the Clearnet, and use a combination of Social Network and Sentiment Analysis for coupling profiles on the two sides.

The approach is based on the idea that when it is possible to establish a clear correspondence between an individual living a double lives style, it is also possible to mark that individual as a potential suspect, and therefore enshorten significanly investigative efforts. The potential is expressed in the duality of Darknet expression of ideas whose admissibility in public domain is deputable, especially when those ideas have a political origin, in a very general sense, including in this also religious, class, national and ethinc principles. The concept is that when it is possible to decontour two individuals that are likely to coincide in the reality, and one of these individuals have a specific interest in political issues, there is a suspect of terrorism (potentially).

To determine an individual to correspond in the Darknets and in the Clearnet, we use the *homophily* principle, namely we consider two invidiuals to be as close as their interests are in common. On the other hand, we make use of the so-called Social Network Analysis, considering two individuals to be as close as their reference networks overlap.

The difficulty in comparing individuals belonging to the two distinct sides of the web, is that they try to hide their correspondence, namely they try to make almost impossible to compare them. The behaviour of individuals that aim at avoinding any overlapping between their harmless public counterpart and their secret dark counterpart is here referred to as the *Spider-man protocol*. Clearly, if an individual is rigourous in keeping the two sides apart, and prevents any leak of information the protocol is respected, and no one can ever discover this secret. In Section 4, we analyse two situations in which it is possible to provide an attack to terrorism privacy, that can be used for useful purposes. The major weakness phase is the initial one, when an individual *becomes* a terrorist. Minor cases regard the preparation of a terroristic attack, and the phase in which an individual plays with the idea of exiting an organization.

### 3.1 Social network analysis: the social network measures of terrorists

A connection network of an individual $i$ contains people that etiher have a personal relationship with $i$, or have a certain group of interests in common. When it is possible to detect the existence of interests in common (homophily), we can establish that $u$ shares some interest with $J$.

Clearly, to share interest does not imply to share viewpoints, and thus an extremist can have a high homophily with a moderate person, being both interested in politics, and maybe being both on similar position, but still not sharing

the model of acting, as in particular, being different in acceptin or not acts of violence as means for making own ideas succed.

If an individual $i$ is a terrorist and and an individual $j$ is homophilic and connected to $i$ it is plausible that also $j$ should be suspect of terrorism. Therefore, once we know that an individual is connected to a potential terrorist, we attempt at determining connections that can be referential for other individuals.

### 3.2 Sentiment analysis: words of terrorists

Every terrorist organization employs a specific war lexicon, a sort of glossary of the fighter. The analysis of the posts of people close to terrorists, as well as many communications from self-declared terrorists, shows that there is combination of *extremism* and specificity of the referential ideology. Communist terrorist movements mixed up, for instance, words of war like fight, battle, kill, and many others with words of communism as working class, revolution, proletariat dictature, and others.

The common style of terrorist communication is also the usage of secret words, the so called *code language*. A famous example of this method of communication is the use of the term *pack* by tupamaros terrorists in South Amrrica in the Seventies, to refer a potential victim of a terrorist attack.

## 4 Weak passages of terrorism web behaviour

There are three phases of the terrorist activities in which the Spider-man protocol is weaker in resisting to attacks:

- In the phase in which an internaut becomes a terrorist, or in formal terms, enter a double lives behaviour;
- In the near temporal proximity of a terrorist attack, epecially during the preparation days;
- The phase in which an individual is planning to exit the terrorist organization he belongs to.

Majorly, during these phases, it is relatively easier that the terrorist makes errors, namely he breaches the protocol, by revealing directly or indirectly his identity.

### 4.1 The radicalisation phase

The radicalization phase is the period of time in which a person starts to move his political ideas close to those of an active terrorist group, or more generically, to the ideas of a political area where violence is considered an option.

From an use of the language viewpoint, it is relatively simpler to determine such a change of behaviour in those contour conditions where radicals exist and are contiguous to extremists and moderates in a general large organization. For

instance this happens for islamic terrorists, and to a more restricted extent, due to the reduction of size of the general movement, for revolutionary communist groups.

Analogously, the social network analysis of these groups reveals that the number of contacts of a newbie radical increase suddenly, during the radicalisation phase. This is due to entering the organization, and is also due to the attraction to other potential newbies generated by the appearance of the newbie in the panorama of radicals and extremists. After a phase like that, the radical pass to the double lives. When this happens, again relatively suddenly, the darknet side of them appear.

From a pure observational viewpoint, this is the phase in which apparent continuity of the Clearnet user is not anymore present: they need to be partly in the Darknets, and this absences are less justified than those of others, because the Clearnet radicals, not the terrorists, obviously, miss the presence of the newbie. In this phase, the number of posts, comments, sharing and other behaviours decrease. Simultaneously a newbie with some omophily of the Clearnet user appears. The radical who is moving to a terrorist group passes a phase in which his *above board* personality needs to be guarded, and therefore the Clearnet user appear often to become less aggressive, and less interested in establishing connections with other radicals. Recognizing these behaviour treats is a viable method to identify a potential terrorist in his initial phase.

### 4.2 Breaching protocols: when terrorists leave permanent traces on the web

During the preparation of a terrorist attack the members of an organization intensiify their darknet communications. Provided that you have connected a Clearnet personality to a hidden Darknet one, the hidden persobality can bring you the information (in this case, regarding an attack in the near future), and the Clearnet one can bring you to the terrorist actual life.

On the other hand, when a member of a terrorist organization is about to leave the organization itself, he tends to dissimulate his desire at most, being this passage much more dangerous in terms of personal freedom than it can be the opposite one, when someone enters an organization. However, a few errors are well-known as providing a view of this cases. In particular it is well known from military literature that moderate dissimulation that is a typical façade of terrorist with double lives on their Clearnet personality, decreases in those phases.

The ability of recognizing the aforementioned treats completely relies upon the combination of sentiment analysis and social network analysis. A flexible architecture for providing such a method is prsented in next section.

## 5 An architecture for the detection of potential terrorists

In this section we introduce an architecture for the detection of terrorists and potential ones called *DetectTerror*.

DetectTerror aim is to detect fake identity analyzing data coming from both public and dark social networks, as summarized in Figure 1.

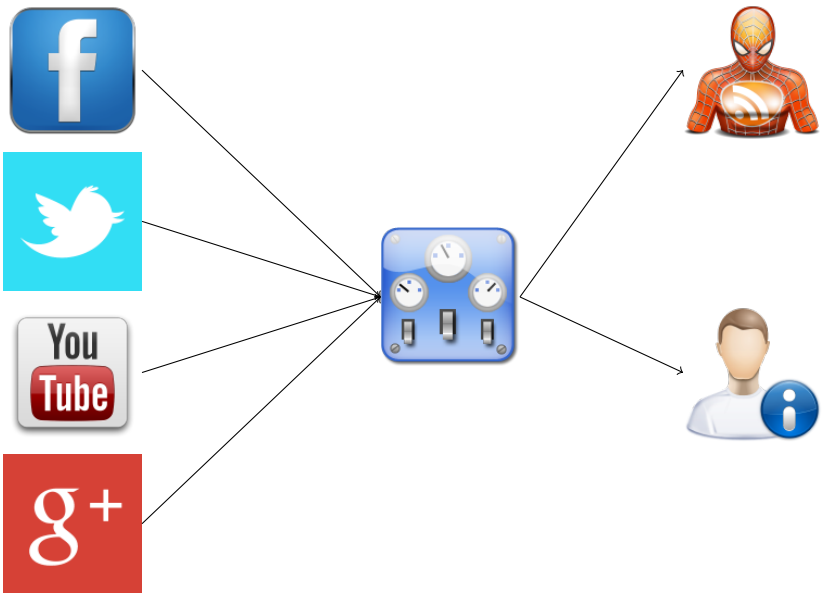


**Fig. 1.** The operative concept of DetectTerror architecture

DetectTerror is made of several modules, each one with a single responsibility; the logic model of DetectTerror is reported in FigureDetectTerror 2.

Every module is related to at leas one other, while all refer to one named *Orchestrator*:

**Crawler:** this component aims at the retrieval of raw informations form a specific source (i.e. Twitter, Facebook) following information structure to gather the correct piece of data. To add a new source to DetectTerror the only implementation regards this module and its related *Normalizer*

**Normalizer:** this component can analyze data and format them so that they are all in the same format and with a structure which makes them ready for the analysis

**Analyzer:** this module takes as input normalized data and gives as output a representation suitable to be later exposed to the *Reasoner*, a kind of digital identity fingerprint

**Reasoner:** this is the actual core of DetectTerror, the one which aim is to discover the relations between digital identities fingerprints to exploit where connections are

**Orchestrator:** this module is the "main app" of DetectTerror actually coordinating all others

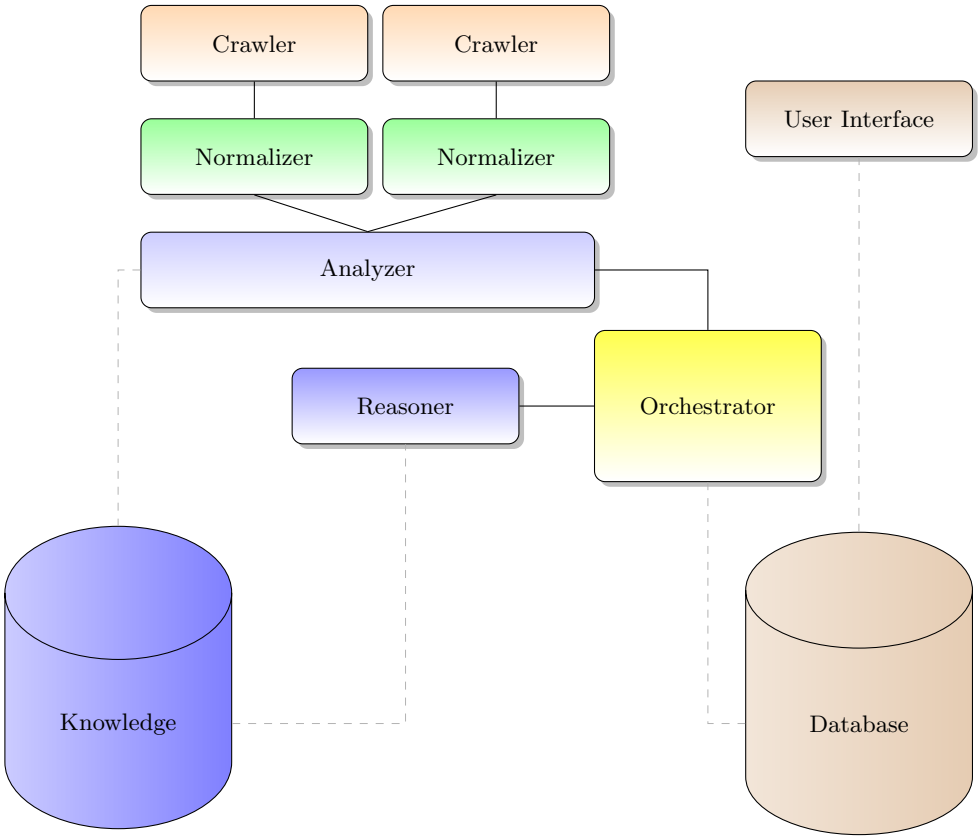


**Fig. 2.** Logic model of DetectTerror architecture

**Knowledge base:** where the knowledge base is stored; the *Reasoner* will access it for reasoning and the *Orchestrator* will increment it after evaluation of the results of the reasoner

**Database:** where all application data are stored, including partially evaluated retrieved data, configurations needed to effectively access information on social media, rules for data normalization, etc.

**User Interface:** tho module provides visualization of all data, allow user to modify parameters

## 6 Related Work

There are numerous natural language processing applications for which subjectivity analysis is relevant, including information extraction and text categorization. According to Wiebe [36], the subjectivity of a text is defined as the set of elements describing the private state of the writer (emotions, opinions, judgments, etc.).

The term Sentiment Analysis has been introduced in 2001, in order to describe the process aimed at automatically evaluating the polarity expressed by a set of given documents [9]. The term Opinion Mining has been introduced in 2003 in order to describe the activity aimed at *processing a set of search results for a given item, generating a list of product attributes and aggregating opinions about each of them* [10]. While OM is mainly focused on recognizing opinions expressed in a given text with respect a specific attributes, SA is focused on classifying a given document according with the polarity [23].

There are several recent studies about sentiment analysis. A common approach for SA is to select a machine-learning algorithm and a method of extracting features from texts and then train the classifier with a human- coded corpus. The main features in representing documents are: bag-of-words as in [1, 17, 16, 33, 35] or tree sentence parsing as in [30, 19].

In [20], the authors show how the SA depends upon the adjectival and adverbial modification of nouns and verbs. Adjectives and adverbs are largely studied as word-sense modificators in the NLP community [28, 14, 7, 12, 21].

In [25] the authors show how to detect authorship by a similarity measure among documents represented by vector space model to identify fake content and fake users. The problem of authorship attribution is to identify the author of a new document having a corpora of documents of known authors [27, 32]. On the other hand, the authors of [26] present a web service that tracks the diffusion of a set of keywords to detect atroturfing and fake content by means of social network analysis procedures.

The authors of [34] present a survey of the issues in social interaction and the recognition of user behaviour in social channels. The analysis of social behavior and patterns of users is the main part in the identification of user groups, as in [22], and in [13].

Some scientist, thanks to the release in 2010 from the famous social network *Twitter* of remote stream APIs that enabled performing of real-time analytics, concentrated on extracting meaning from *tweets*[11].

The authors in [6] explored the frequency of retweets surrounding an event and the duration between the first and the last of these retweet to extract information on how people behave when confronted with both positive and negative events.

Using the aforementioned *Twitter* API, in [5] data have been used in the study of the spread of online hate speech, or *cyber hate*, and forecast the likely spread of cyber hate; a classifier was used based on Bag Of Word model and the presence of key terms.In [29] word are tagged using TreeTagger (Schmid, 1994) and interpreted the difference of tag distributions between sets of text (positive, negative, neutral or subjective, objective), while in [15] the authors make use of ontologies to enhance sentiment analysis and attach a sentiment grade for each distinct notion in *Twitter* posts.

Always analyzing *Twitter* data, in [4] there is an attempt at understanding tension at an early stage and evidence is given that a combination of conversation

analysis methods and text mining outperforms machine learning approaches at such task.

In [37] the whole chapter is related to topics of sentiment analysis based on visual and textual content, where information is extracted from meaning of words or images.

In [31] the authors search for Negativity, Fear, and Anger showing that fear and anger are distinct measures that capture different sentiments, and they achieve these results using dictionary-based sentiment analysis.

Mining opinions and sentiment from social networking sites is the aim in [18] where the tool used is a bag of words feature set enhanced by a statistical technique named *Delta TFIDF* to efficiently weight word scores before classification.

To exploit certain types of information from reports on terrorist incidents, the authors in [8] perform syntactic and semantic analysis and uses lexicons of various categories of terms.

In [24] the focus is the problem of real-time sub-events identification in social media data (i.e., Twitter, Flickr and YouTube) during emergencies, and the method used involved tracking the relevant vocabulary to capture the evolution of sub-events over time.

There is also a study [3] in which Social Network Analysis is combined with Sentiment Analysis to explore the potential for the possibility of individuals being radicalised via the Internet; key terms and their frequency are used in this analysis.

As a matter of fact, it is not only what is said that counts, but also who is speaking. There are people more likely to be listened to (or *followed*) than others and it can be of relevance to identify radically influential users in web forums, which the subject of other studies[2].

## 7 Conclusions

This paper describes an architecture that can be used for detecting terrorists when they use Darknets and the Clearnet in a substantially different and anyhow permeable way, breaching what we call the *Spider-man behaviour protocol*.

There are three different ways in which this research has to be taken further. First of all, we shall implement the technology in practice and experiment it with real-life cases, in order to provide a direct and verifiable example of what suggested in this paper. Secondly we need to refine both social and sentiment techniques in order to detect terrorists at different developing stages: early stage, namely when they enter the organization and pass to a clandestine (possibly partly) life, phase before exiting the organization (that can be used to prevent attacks). Finally it is of strong interest to provide a ranking, possibly regarding belonging to an organisation as well as a form of measure for the probability of an individual to enter an organisation.

## References


1. Marilisa Amoia and Claire Gardent. Adjective based inference. In *Proceedings of the Workshop KRAQ'06 on Knowledge and Reasoning for Language Processing*, KRAQ '06, pages 20–27, Stroudsburg, PA, USA, 2006. Association for Computational Linguistics.
2. T. Anwar and M. Abulaish. Ranking radically influential web forum users. *Information Forensics and Security, IEEE Transactions on*, 10(6):1289–1298, June 2015.
3. A. Bermingham, M. Conway, L. McInerney, N. O'Hare, and A.F. Smeaton. Combining social network analysis and sentiment analysis to explore the potential for online radicalisation. In *Social Network Analysis and Mining, 2009. ASONAM '09. International Conference on Advances in*, pages 231–236, July 2009.
4. Pete Burnap, Omer F. Rana, Nick Avis, Matthew Williams, William Housley, Adam Edwards, Jeffrey Morgan, and Luke Sloan. Detecting tension in online communities with computational twitter analysis. *Technological Forecasting and Social Change*, 95(0):96 – 108, 2015.
5. Pete Burnap and Matthew L. Williams. Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making. *Policy and Internet*, 2015.
6. Pete Burnap, Matthew L. Williams, Luke Sloan, Omer Rana, and alt. Tweeting the terror: modelling the social media reaction to the woolwich terrorist attack. *Soc. Netw. Anal. Min.*, 2014.
7. Stergios Chatzikyriakidis and Zhaohui Luo. Adjectives in a modern type-theoretical setting. In Glyn Morrill and Mark-Jan Nederhof, editors, *Formal Grammar*, volume 8036 of *Lecture Notes in Computer Science*, pages 159–174. Springer Berlin Heidelberg, 2013.
8. Sumali J. Conlon, Alan S. Abrahams, and Lakisha L. Simmons. Terrorism information extraction from online reports. *Journal of Computer Information Systems*, 2015.
9. Sanjiv Das and Mike Chen. Yahoo! for amazon: Sentiment parsing from small talk on the web, 2001.
10. K. Dave, S. Lawrence, and D.M. Pennock. Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th International World Wide Web Conference (WWW)*, pages 519–528, 2003.
11. Amir Hossein and Akhavan Rahnama. *Real-time Sentiment Analysis of Twitter Public Stream.* University of Jyvskyl, 2015.
12. Bjørn Jespersen and Giuseppe Primiero. Alleged assassins: Realist and constructivist semantics for modal modification. In *Logic, Language, and Computation - 9th International Tbilisi Symposium on Logic, Language, and Computation, TbiLLC 2011, Kutaisi, Georgia, September 26-30, 2011, Revised Selected Papers*, pages 94–114, 2011.
13. Nitin Jindal, Bing Liu, and Ee-Peng Lim. Finding unusual review patterns using unexpected rules. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, pages 1549–1552, New York, NY, USA, 2010. ACM.
14. Christopher Kennedy. Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, 30(1):1–45, February 2007.

15. Efstratios Kontopoulos, Christos Berberidisand Theologos Dergiades, and Nick Bassiliades. Ontology-based sentiment analysis of twitter posts. *Expert Systems with Applications*, 2013.
16. Zhaohui Luo. Formal semantics in modern type theories with coercive subtyping. *Linguistics and Philosophy*, 35(6):491–513, 2012.
17. Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 142–150, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.
18. Justin Martineau and Tim Finin. Delta tfidf: An improved feature space for sentiment analysis. In *Proceedings of the Third International ICWSM Conference*, 2009.
19. Jeff Mitchell and Mirella Lapata. Composition in distributional models of semantics. *Cognitive Science*, 34(8):1388–1439, 2010.
20. A. Moreo, M. Romero, J.L. Castro, and J.M. Zurita. Lexicon-based comments-oriented news sentiment analyzer system. *Expert Systems with Applications*, 39(10):9166 – 9180, 2012.
21. Marcin Morzycki. Modification, 2013. Book manuscript. In preparation for the Cambridge University Press series *Key Topics in Semantics and Pragmatics*.
22. Arjun Mukherjee, Bing Liu, Junhui Wang, Natalie Glance, and Nitin Jindal. Detecting group review spam. In *Proceedings of the 20th International Conference Companion on World Wide Web*, WWW '11, pages 93–94, New York, NY, USA, 2011. ACM.
23. Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, January 2008.
24. Daniela Pohl, Abdelhamid Bouchachia, and Hermann Hellwagner. Online indexing and clustering of social media data for emergency management. *Neurocomputing*, 2015.
25. Tieyun Qian and Bing Liu. Identifying multiple userids of the same author. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1124–1135, 2013.
26. Jacob Ratkiewicz, Michael Conover, Mark Meiss, Bruno Gonçalves, Snehal Patil, Alessandro Flammini, and Filippo Menczer. Truthy: Mapping the spread of astroturf in microblog streams. In *Proceedings of the 20th International Conference Companion on World Wide Web*, WWW '11, pages 249–252, New York, NY, USA, 2011. ACM.
27. Michal Rosen-Zvi, Thomas Griffiths, Mark Steyvers, and Padhraic Smyth. The author-topic model for authors and documents. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, UAI '04, pages 487–494, Arlington, Virginia, United States, 2004. AUAI Press.
28. Susan Rothstein. Fine-grained structure in the eventuality domain: The semantics of predicative adjective phrases and be. *Natural Language Semantics*, 7(4):347–420, 1999.
29. Suprajha S, Yogitha C, Architha J Sanghvi, and Dr. H S Guruprasad. A study on sentiment analysis using tweeter data. *International Journal for Innovative Research in Science and Technology*, 1(9), 2015.

30. Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642, Seattle, Washington, USA, October 2013. Association for Computational Linguistics.
31. Stuart Soroka, LorI Young, and MeItal Balmas. Bad news or mad news? sentiment scoring of negativity, fear, and anger in news content. *Annals of the American Academy of Political and Social Science*, 2015.
32. Efstathios Stamatatos. A survey of modern authorship attribution methods. *J. Am. Soc. Inf. Sci. Technol.*, 60(3):538–556, March 2009.
33. Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173, 2012.
34. David C. Uthus and David W. Aha. Multiparticipant chat analysis: A survey. *Artificial Intelligence*, 199?200(0):106 – 121, 2013.
35. Anil Saroliya Vijay Dixit. A semantic vector space model approach for sentiment analysis. *International Journal of Advanced Research in Computer and Communication Engineering*, 2, 2013.
36. Janyce Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. Learning subjective language. *Comput. Linguist.*, 30(3):277–308, September 2004.
37. Jianbo Yuan, Quanzeng You, and Jiebo Luo. Sentiment analysis using social multimedia. In Aaron K. Baughman, Jiang Gao, Jia-Yu Pan, and Valery A. Petrushin, editors, *Multimedia Data Mining and Analytics*, pages 31–59. Springer International Publishing, 2015.