LaSEI
School of Economic Intelligence & International Relations
Escuela de Inteligencia Económica y RRII
Universidad Autónoma de Madrid

01/2021

Vol. 3, n 6, pp. 51 - 58
ISSN: 2659-9805
Editor in-Chief: Javier Oubiña

JOURNAL OF ECONOMIC & BUSINESS INTELLIGENCE

Universidad Autónoma de Madrid UAM

INFERRING USER PERSONALITY FROM TWITTER

Julia Llanos Alonso[i], Álvaro Ortigosa[ii], & Rosa Mª. Carro Salas[iii]

Abstract

Personality affects how a person behaves when interacting with people and computer systems. It can determine one's needs and preferences in different contexts, which is especially useful for adaptive and recommender systems. Personality questionnaires are widely used to acquire information about user personality. However, filling in them can be tedious. Analyzing the user interactions can be another way of obtaining that information. Since personality has an impact on the way a person interacts with others, and social networks are widely used for this purpose, information about user interactions through social networks can give a clue about their personality. In this paper, we present a system able to obtain data about user interactions in Twitter and analyze them in order to infer user personality. The system has been used not only to infer personality but also to compare and evaluate different user models, classifiers and personality dimensions.

Key words: Personality inference; classifiers; social network analysis; user modeling; Twitter.

ACM Classification Keywords: H.1.2 User/Machine Systems: Human factors, Human information processing; H.5.3 Group and Organization Interfaces: Web-based interaction.

1. Introduction

The study of personality is one of the research areas that has required more efforts by the scientific community on humanities over time. Traditional techniques to get information about user personality rely on questionnaires, surveys, tests and so on [1]. However, one of the difficulties when studying personality deals with the huge variability that this trait shows in humans [2]. To obtain information about personality with a minimum statistical rigor, it is necessary to conduct studies with large numbers of users. This is not always possible: it requires large-scale recruitment capabilities and involves a high cost, which sometimes is difficult to achieve. Some of the main strategies followed to overcome those difficulties are trying to get information from different sources or using different techniques. The main challenges in this case deal with obtaining reliable information as well as avoiding user privacy violations. Knowing the user personality can be very useful in various fields. For example, if a user's personality were a risk for society in a particular context, it would be important to know it, to take the corresponding security measures. In the context of commerce, this information can be used to customize marketing strategies. In the same direction, it can determine the adaptation strategies to be followed in adaptive and recommender systems.

This work aims to get information about user personality in a rigorous and statistically reliable way. We propose to use the social network Twitter as a source of information. Some

[i] Telefónica España
[ii] Instituto de Ciencias Forenses y de la Seguridad (ICFS) Universidad Autónoma de Madrid (UAM) alvaro.ortigosa@uam.es
[iii] Departamento de Ingeniería Informática de la EPS. Universidad Autónoma de Madrid (UAM)

studies use specific information of each user actions, individually, in social networks. This does not benefit from a great amount of information available: the interconnections between the users and the interactions among them. This work attempts to make the most of that information available about users and their connections. Its main goal is to develop a system able to infer user personality with maximum reliability. With this purpose, we propose various models with different amounts and types of information. We designed and implemented techniques for gathering information from users, both to infer and to obtain user personality. We defined different prediction models based on the information gathered. We implemented a system to infer personality using different models and classifiers/clustering techniques and, finally, we evaluated the usefulness of each of those models to support personality inference.

The rest of the paper is structured as follows. Section 2 describes the state of the art. Section 3 explains the creation of user models starting from information available in Twitter. Section 4 describes the development of all the elements needed to satisfy the goals posed. Section 5 and 6 present the evaluation of this work and the results obtained, respectively. Finally, section 7 discuss the conclusions and future work.

2. State of the art

The need for socializing is an intrinsic characteristic of the human being. With the creation of the Internet, the possibilities increased exponentially and, some years later, social networks emerged. There are studies that try to throw light on the fact that social networks are the main and, sometimes, the only way for millions of users to fulfill their social needs [3]. These networks are very useful from different viewpoints. For example, some users overcome barriers that cannot overcome in person [4]. Social networks are a data source for information retrieval research [5]. Given all the information available in social networks, analyzing them is one strategy to obtain user data for different purposes, such as recommending tweets and news [6].

One of the user features that better reflects on social networks is personality [7]. User interactions or contacts are rich sources of information [8]. There are some approaches focused on studying personality in general in social networks [9], or in specific ones such as Facebook [10], Twitter [7], LinkedIn [11] or others. One useful social network to study personality is Twitter, because of aspects such as its huge number of users, the existence of both unidirectional and bidirectional relationships between them, the high frequency on the establishment of new relationships, or the big amount of public data available.

Most existing work related to personality inference in Twitter base on the extraction of user features. In [7], eight user features are used for the analysis (number of followers, number of friends, number of references, number of replies, number of hashtags, number of links, words per tweet and network density). In [12], three parameters are used (number of followers, number of friends and number of references). Finally, other works focus on simple natural language processing [13].

3. Building user models from Twitter

User models are essential for personality prediction systems. In this work, the user model contains all the data that will be useful for personality prediction. Building the user model is a key issue for the system functioning. The best-known personality theories are classified in [14]: factorial, cognitive and biological. After studying them, we have selected Eysenck's [15]. It is one of the best known and most tested one, and describes three dimensions: Extroversion/Introversion (E/I), Neuroticism (N) and Psychoticism (P). The questionnaire associated to this model [1] has been widely validated and standardized. It contains 83 questions to answer with "Yes" or "No". Some of them do not relate to any personality dimension. They are set out to get the user sincerity.

In this work, user models are constructed starting from that information available in Twitter. Each model is constructed selecting some of these attributes or processing data to obtain new ones. With the aim to decide which attributes compose each user model, we have studied their meaning as well as how each of them are expected to affect to each personality trait. It is worth mentioning that two attributes that Twitter users tend to omit are age and gender. Some studies have tried to infer them [16] [17]. However, it is complicated and the results obtained are unreliable. Because of this, we have incorporated these attributes only in one user model. In this case, it is necessary for the user to provide this information (regarding age, knowing whether he is younger than 30 is enough). Five user models have been built, each of them with different attributes, with the aim to test which ones are able to infer personality better. These models are:

Model 1 (Individual- Simple)

- Goals: To obtain results that can be compared with those from the state of the art.
- Attributes: Only those directly available from Twitter without any further processing, such as

user age, gender, location, number of friends (followed users), number of followers, number of years in Twitter, number of tweets, mean and standard deviation of tweet length, mean and standard deviation of the number of images published, mean and standard deviation of the number of links, and mean and standard deviation of the number of hashtags used.

Model 2 (Individual - Advanced):

- Goals: To enrich model 1 by including both user profile and tweet processing. To observe the effect of including these parameters and to check whether it leads to better results, so that the higher computational cost of this model is justified.
- Attributes: Those from model 1 plus tweets frequency (both daily and monthly), number of different locations (when available), and mean and standard deviation of:
  - number of positive and negative words
  - number of exclamation marks
  - number of emoticons
  - number of words with written accent
  - number of digits
  - word length
  - number of with spaces
  - number of upper case chars
  - lexical richness.

It is important to remark that these features were selected because previous studies [13] [14] show that they can be affected by personality traits.

Model 3 (Group - Simple):

- Goals: To include information not only about the user but also about his interactions with followed users and followers. However, this model only includes information retrieved through the own user profile. That is, from the Twitter API point of view, only data regarding the analyzed user was queried.
- Attributes: Those from model 2 plus mean and standard deviation of:
  - number of mentions made by the subject
  - number of retweets made by the subject
  - number of received retweets
  - number of replies made by the subject
  - number of tweets favorited by the subject
  - number of subject tweets favorited by other users.
  - subject time to retweet (measured in
  - minutes)

Model 4 (Group - Advanced):

- Goals: To enrich model 3 by including information requiring either further processing or querying the Twitter API for other users' data.
- Attributes: those from model 3 plus mean and standard deviation of:
  - number of received mentions
  - number of received replies
  - subject time to reply
  - time the subject is replied
  - number of friends of subject's friends
  - number of followers of subject's friends
  - number of friends of subject's followers
  - number of followers of subject's followers.

Features considered in models 1 to 4 imply a progressive increment on the amount of data that need to be retrieved and processed. Of course, the assumption is that these models are going to be progressively more accurate. The goal is to verify if the additional accuracy is worthy of the downloading and processing overheads.

Model 5 (Group – Complete except for gender and age):

- Goal: To study the influence of gender and age on personality inference. The reasoning is that these two features normally are not present in the user profile.
- Attributes: model 4 minus gender and age.

Model 6 (Group – Others' personality):

- Goal: To study the effect of knowing the friends and followers' personality on the prediction of the subject personality. In this case, both the interactions among users and their personality are analyzed.

Attributes: model 4 plus means values of extraversion, neuroticism and psychoticism for followers and friends.

4. The System

The goal of the system is to be able to infer user personality. In order to do so, the system classifies the user according to each trait. Therefore, besides the attributes that compose the user models, some extra information is needed in order to train the classifiers, as well as to evaluate the accuracy of the different models and classifiers. In particular, knowing the user personality, obtained by means of the Eysenck questionnaire, is essential. Next, we describe, firstly, the application developed to support the realization of the Eysenck questionnaire, secondly, the way of retrieving information from Twitter and, thirdly, how the inference system works.

4.1. Web application: Personality Questionnaire

We have developed a Web application that supports the realization of Eysenck questionnaire and the access to the results obtained (available at: https://www.icfs.uam.es/personalidad/indexPersonality.php). Firstly, the user has to provide information about his age, gender, occupation, mother tongue, educational level (these data, apart from age and gender, are for future statistics). Next, the 83 questions of the questionnaire itself are presented. Once the user completes the questionnaire, he must write the name of his user in Twitter and select the preferred way to receive the results through the API of Twitter: in a private message, sent by @personalityCNEC, or publicly, with a specific mention to the user in a tweet. The answers given by the user, as well as the calculated values for his personality traits, are stored in a database. This application also shows the descriptions of the personality tracks and explanations about the meaning of high/low values in each of them. The web application has been implemented in PHP, using the TwitterOAuth library and PostgresQL.

4.2. Twitter User Data Retrieval

In order to retrieve the information available about each user from Twitter, we have used the API REST, which allows developers to connect their applications with Twitter. REST (Representational State Transfer) supports the communication between two systems, through HTTP requests, to transfer data or to ask for the execution of operations within these data in diverse formats (e.g., JSON, XML, etc.). The queries to retrieve information about the users are coded in Python. The answer received includes the user profile and his tweets. Both the number of tweets to retrieve per query (200) and the maximum number of queries per program are limited.

Once the information about users, friends and followers is retrieved, some calculations are made to generate new attributes for some of the models. For example, for word analysis, tweets are processed: letters are low-cased and accent marks are removed; stopwords are removed; emoticons are identified; then, positive and negative words (and emoticons) are identified, by comparing each word with those in the corresponding dictionaries. To overcome the need of identifying all possible variations of a word, dictionaries contain the lexical roots of the words, and letters from tweets are removed one by one during the comparisons.

4.3. Inference system

Once all the data are collected and calculations are made, personality can be inferred. Firstly, the data obtained and stored previously are prepared. Then, the classifiers are trained. Afterwards, the system is able to receive new data related to a user and classify them, obtaining the values inferred for each personality trait of that user.

Training a classifier requires multiple instances of the user model chosen to learn from these data. Preparing data means generating a set of data for training and testing, containing, on one hand, the attributes that will be used for the inference, and, on the other hand, the labels of the classes, which define each user's personality.

As it was mentioned above, six different user models have been defined. Therefore, for each model, the corresponding attributes are extracted, which define the user feature vector. Classes are defined as the real values of each personality trait for the user. In this work, three classes have been defined for each trait (high, medium or low). Therefore, if the three personality dimensions are $d = \{D1, D2, D3\}$, for K output values, each dimension is defined as Dn = $\{q_1^n, \dots, q_i^n, \dots, q_k^n\}$.

Since, in this work, each personality trait has its own classifier, the train-test dataset for each individual is:

Feature vector	q_i^1
Feature vector	q_j^2
Feature vector	q_m^3
$i, j, m \in \{1 \dots K\}$	

The user model 6 (Group- Other's Personality) needs some additional preparation, since other user's personality traits must be included in the feature vector. All these data were divided in two sets: the first one was used for training the classifiers. The second one was used to estimate the classification errors. We used Weka for these tasks.

5. Evaluation

The total number of questionnaires filled in were 1.320. The total number of valid questionnaires were 922. Some invalid tests were discarded: those filled in to test the system and intrinsic errors, and those who showed defects that made their processing not possible. The dataset used to generate models and predictions focus on those users for which it was possible to collect not only their own information, but also that from all their friends and followers (which, sometimes, were thousands). The total number of main users was 47. The total number of users plus their friends plus their followers were 10.175. And the total number of tweets collected were 9.644.195.

Let us notice that the computational cost and the authorization requirements have been pretty high. The simpler models could have been analyzed with data from all the individuals that filled in the questionnaire (more than 900). However, in order to compare all the models proposed, 47 users were used, since some models need those "main" users with their friends and followers, and, therefore, the latter cannot be considered as "main" users. Before building classifiers, the system functioning has been tested with 50 individuals. Variability has also been calculated, both intrinsic and related to the passing of time. Afterwards, the different models were considered and the corresponding classifiers were trained using half of the users' data. Next, the personality of the other half of the users was inferred. The selection of datasets for training and testing was done with cross-validation methods. Then, we calculated the errors by comparing the results obtained by classifiers from those got from questionnaires. Finally, we compared the results obtained from different models. The results are presented next.

5. Results

The results obtained for the personality traits in each model are presented next. For each trait, we present the results obtained when applying classifiers and those got when applying clustering techniques. Table 1 shows the results for extraversion when using classifiers and table 2 shows the results when using clustering techniques. The best results have been obtained when using Sequential Minimal Optimization (SMO). However, clustering algorithms resulted better for smaller models such as 1 and 2. The worse results in general were those from Naive Bayes.

Table 1. Error percentages when inferring extraversion through different classifiers.

Classifiers	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Naive Bayes	67	76	69	69	74	69
Multilayer Perceptron	61	59	50	56	**52**	
SVM	59	61	63	63	65	63
SMO	61	52	**45**	**54**	65	**39**
J48	54	61	54	65	63	61
Random Forest	63	63	63	63	63	63

Table 2. Error percentages when inferring extraversion through clustering techniques.

Clusters	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
EM	56	59	54	56	54	61
Filtered Clustered	**48**	50	52	56	61	56
Simple Kmeans	**48**	**50**	52	56	61	56

Tables 3 and 4 show the error percentages obtained for neuroticism. In this case, the best results for complex models were those from Expectation Maximization algorithm (EM), while SVM and Random Forest worked a little bit better for simpler ones. Once again, the worst results were those from Naive Bayes classifier.

Table 3. Error percentages when inferring neuroticism through different classifiers.

Classifiers	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Naive Bayes	69	83	83	65	76	69
Multilayer Perceptron	61	72		72		69
SVM	**56**	**56**	59	56	59	56
SMO	63	67	67	69	59	63
J48	56	61	56	63	56	69
Random Forest	**56**	**56**	56	56	56	56

Table 4. Error percentages when inferring neuroticism through clustering techniques.

Clusters	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
EM	59	59	54	**48**	**50**	**43**
Filtered Clustered	61	59	**52**	52	61	50
Simple Kmeans	61	59	**52**	52	61	50

Tables 5 and 6 show the results obtained for psychoticism. In this case, classifiers work better for all the models in general. The ones obtaining the best results for every model are SVM and Random Forest. Once again, Naive Bayes gets the worst results.

Table 5. Error percentages when inferring psychoticism through different classifiers.

Classifiers	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Naive Bayes	61	59	65	65	65	56
Multilayer Perceptron	43	43	50	43		
SVM	**41**	**41**	**41**	**41**	**41**	**41**
SMO	51	43	52	**41**	**41**	43
J48	50	**41**	45	**41**	**41**	61
Random Forest	**41**	**41**	**41**	**41**	**41**	**41**

Table 6. Error percentages when inferring psychoticism through clustering techniques.

Clusters	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
EM	54	61	63	56	59	63
Filtered Clustered	50	56	52	59	56	56
Simple Kmeans	50	56	52	59	56	56

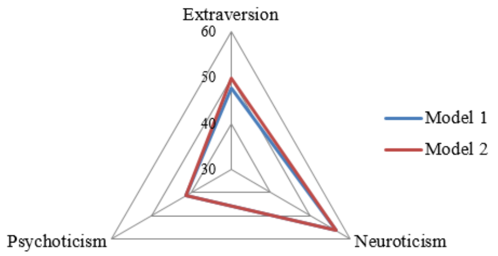
These results seem to show that clustering techniques work well for smaller models, while classifiers work better for bigger and more complex models. On one hand, those results have been obtained when working with data from 47 users (plus their friends and followers, making more than 10000 users). A dataset of 922 users could have been used for models 1-5, but we decided to use 47 because of the reasons explained above, related to model 6 and the comparison among all the models. We plan to use all those data in further research. On the other hand, intrinsic errors got from the validation process could be subtracted from the results shown, making error percentages lower.

5.2. Comparative evaluation

One of the main goals of this work was to generate prediction models richer than the ones described on the state of the art leading to better results. Let us compare the different models generated and the results obtained. As it can be seen in the previous tables, when comparing models 1 and 2, the results obtained through different classifiers and clustering techniques are similar. Therefore, it does not seem to be useful to include attributes that, although being simple, require certain processing time, such as word polarity analysis.

Figure 1. Minimum error percentage of a) models 1 and 2 and b) models 1 and 3, for each trait. The smaller area, the better.

a)



b)

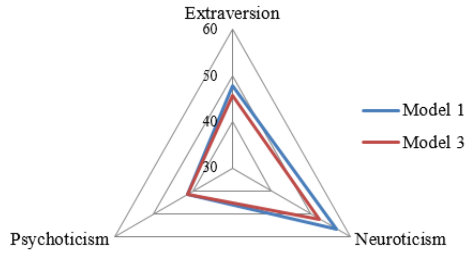
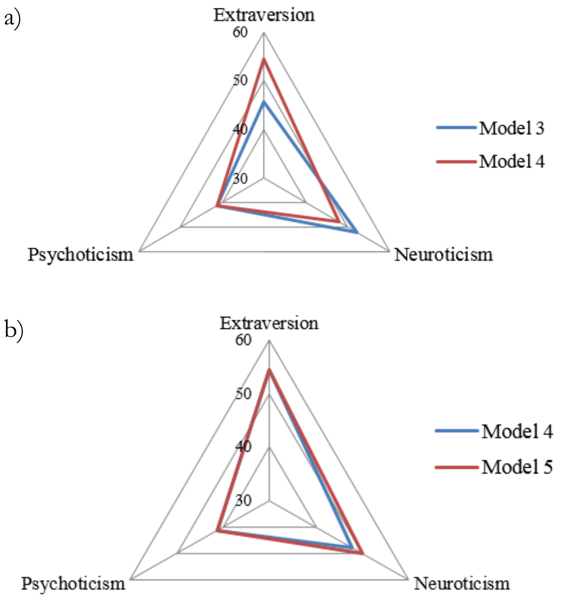


Figure 1a) shows the minimum error percentage of each model in each dimension. Globally, model 2 does not seem to lead to better results than model 1. When comparing models 1 and 3, it can be seen that incorporating the user interactions with his environment in the model leads to better results. Figure 1b shows the global results of those models. The improvement can be easily seen (it is smaller in psychoticism, because the error was low even in the simple model).

Now we want to check whether complex (calculated) attributes could be useful in group models, even if they were not relevant in the individual model (comparison between models 1 and 2). For example, whether considering information about the frequency of friends' and followers' tweeting is more useful than only taking into account simple data about their interactions. Up to our knowledge, this is the first model in which information about related users is incorporated for personality inference, making this strategy novel in this sense.

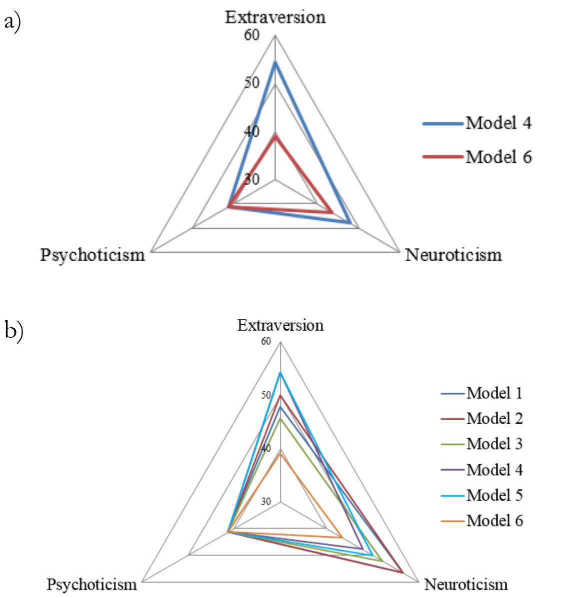
When comparing models 3 and 4, it can be seen that those attributes reduce the error percentage when inferring neuroticism. However, they do not lead to global better results (see figure 2a).

Figure 2. Minimum error percentage of a) models 3 and 4 and b) models 4 and 5, for each trait



Given that the users normally omit information about their age or gender, and the difficulty of inferring these attributes with accuracy, let us check whether omitting them has a negative impact on the results obtained. When comparing models 4 and 5 we can see that the results for all the traits does not vary significantly. They only show a small improvement for neuroticism (figure 2b).

Figure 3. Minimum error percentage of a) models 4 and 6 and b) models 1-6, for each trait.



Now let us check whether the personality of friends and followers of a given user can have an impact on the inference of that user personality (model 6). As it can be seen, the use of this information leads to better results in two of the three traits, not changing the result in the third. Therefore, it can be concluded that including this information in the models is useful (see global impact in figure 3a).

Finally, it should be checked whether the results obtained for model 6 improves the results obtained in more traditional models and those proposed in this work. Figure 3b shows a comparison between the 6 models simultaneously. As it can be seen, including information related to the users does lead to better results.

6. Conclusions and future work

We have reached our goal of developing a system able to obtain large datasets from Twitter, to process them and to use them both for inferring user personality and for comparing different inference models. The results obtained when inferring personality can be improved by training and testing the classifiers with bigger datasets. Due to our goal of comparing the six models proposed, we have limited ourselves to use data from 47 users along with those from 10175 friends and followers. However, data of the 922 that filled in the questionnaire could be used in models 1-5, as explained above, leading to more statistically significant results. Intrinsic errors could also be considered to decrease the error percentages obtained. We plan to do both tasks in the immediate future.

The comparison between the models proposed has thrown light about the usefulness of considering information about the user's environment (friends and followers) for personality inference. Moreover, it has shown the importance of analyzing which attributes should be incorporated in user predictive models before generating them. For example, we have observed that it is possible to include, wrongly, attributes that involve high computational costs, such as age or gender, which do not improve the results obtained significantly. We have also found attributes that are not relevant for personality prediction. All this work on attribute analysis is expected to have an impact in the creation of better user models and, therefore, more accurate personality inference systems.

Future work comprises training and testing classifiers with all the data already available from more than 900 users, and considering the possibility of combining classifiers with the aim of getting better results. Other approaches, such as analyzing not only positive/neutral/negative words in tweets, but also their meaning in the context of personality, can be explored. Analyzing the tweets syntactically could also improve the results obtained. This has led to good results when inferring emotional states in Facebook [18]. However, we do not know whether the results would be similar in Twitter, because of the different writing style: mainly, shorter sentences.

7. Acknowledgments

This work has been funded by the BLIND, project BLIND (BLIND) and by BLIND, project BLIND (BLIND). We also thank all the people that participated in this research fulfilling the personality test.

8. References

1. H. J. Eysenck and S. B. G. Eysenck. 1994. *Manual for the Eysenck Personality Questionnaire*: (EPQ-R Adult). Educational Industrial Testing Service.
2. B. M. Baird, K. Le, and R. E. Lucas. 2006. On the nature of intraindividual personality variability: reliability, validity, and associations with well-being., *Journal of personality and social psychology*, vol. 90, 3, p. 512
3. K.-Y. Lin and H.-P. Lu. 2011. Why people use social networking sites: An empirical study integrating network externalities and motivation theory, *Computers in Human Behavior*, vol. 27, 3, pp. 1152-1161.
4. C. Kadushin. 2012. *Understanding social networks: Theories, concepts, and Findings*. Oxford University Press.
5. A. Guille, H. Hacid, C. Favre, and D. A. Zighed. 2013. Information diffusion in online social networks: A survey, *ACM SIGMOD Record*, vol. 42, 2, pp. 17-28.
6. W.-J. Lee, K.-J. Oh, C.-G. Lim, and H.-J. Choi. 2014. User profile extraction from twitter for personalized news recommendation, in *Procs. of 16th International Conference on Advanced Communication Technology (ICACT)*, pp. 779-783, IEEE.
7. J. Golbeck, C. Robles, M. Edmondson, and K. Turner. 2011. Predicting personality from twitter, in Privacy, Security, *IEEE International Conference on Risk and Trust*, pp. 149-156, IEEE.
8. J. Staiano, B. Lepri, N. Aharony et al. 2012. Friends don't lie: inferring personality traits from social network structure, in *Procs of the ACM Conference on ubiquitous computing*, pp. 321-330, ACM.
9. T. Correa, A. W. Hinsley, and H. G. De Zuniga. 2010. Who interacts on the web?: The intersection of users' personality and social media use, *Computers in Human Behavior*, 26, 2, pp. 247-253.
10. A. Ortigosa, R. M. Carro, and J. I. Quiroga. 2014. Predicting user personality by mining social interactions in Facebook, *J. of Computer and System Sciences*, vol. 80, 1, pp. 57-71.
11. A. Verschuren and K. Ranganath. 2012. *Relying on linkedin profiles for personality impressions*. PhD thesis, Master's thesis, Tilburg University, The Netherlands.
12. D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft. 2011. Our twitter profiles, our selves: Predicting personality with twitter, in *IEEE PASSAT*, pp. 180-185, IEEE.
13. C. Sumner, A. Byers, R. Boochever, and G. J. Park. 2012. Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets, in *Procs. of 11th International Conference on Machine Learning and Applications (ICMLA)*, vol. 2, pp. 386-393, IEEE.
14. L. A. Pervin and O. P. John. 1999. *Handbook of personality: Theory and research*. Elsevier.
15. H. J. Eysenck et al. 1968. *Eysenck personality inventory*. Educational and Industrial Testing Service. San Diego.
16. J. D. Burger, J. Henderson, G. Kim, and G. Zarrella. 2011. Discriminating gender on twitter, in *Procs of the Conference on Empirical Methods in NLP*, pp. 1301-1309, ACL.
17. D. Nguyen, R. Gravel, D. Trieschnigg, and T. Meder. 2013. "How old do you think i am?"; a study of language and age in twitter, in *Procs of the 7th International AAAI Conference on Weblogs and Social Media*, pp. 786-796, AAAI Press.
18. A. Ortigosa, J.M. Martin, R.M. Carro. 2014. Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior* 31, 527-541.