In: Social Networking: Recent Trends, Emerging Issues and Future Outlook
Editor: Xin Ming Tu, Ann Marie White and Naiji Lu
pp. 41-53
ISBN 978-1-62808-534-1

Chapter 3

RELATIONSHIPS BETWEEN PERSONALITY AND INTERACTIONS IN FACEBOOK

Fabio Celli*
CLIC-CIMeC
University of Trento

Luca Polonio†
DiSCoF
University of Trento

Abstract

In this paper we address the issue of how users' personality affects the way people interact and communicate in Facebook. Due to the strict privacy policy and the lack of a public timeline in Facebook, we automatically sampled data from the timeline of one "access user". Exploiting Facebook's graph APIs, we collected a corpus of about 1100 ego-networks of Italian users (about 5200 posts) and the users that commented their posts. We considered the communicative exchanges, rather than friendships, as a network. We annotated users' personality by means of our personality recognition system, that makes use of correlations between written text and the Big5 personality traits, namely: extroversion, emotional stability, agreeableness, conscientiousness, openness. We tested the performance of the system on a small gold standard test set, containing statuses of 23 Facebook users who took the Big5 personality test. Results showed that the system has a average f-measure of .628 (computed over all the five personality traits), which is in line with the state of the art in personality recognition from text. The analysis of the network, that has a average path length of 6.635 and a diameter of 14, showed that open-minded users have the highest number of interactions (highest edge weight values) and tend to be influential (they have the highest degree centrality scores), while users with low agreeableness tend to participate in many conversations.

Keywords: Social Network Analysis, Personality Recognition, Facebook, Natural Language Processing.

*E-mail address: fabio.celli@unitn.it
†E-mail address: luca.polonio@unitn.it

1. Introduction

1.1. Overview

Written text convey a lot of information about the personality of its author (Mairesse et Al. 2007 [22], Argamon et Al. 2005 [2]), and today machine learning techniques allow us to extract personality from text automatically, with a certain degree of accuracy. Online Social Networks (OSN) are huge repositories where user-generated written texts (posts) are found associated with their authors (users), hence they are the perfect place for the automatic extraxtion of personality and the analysis of how it affects interactions among users.

Facebook in particular is a social network service launched in 2004 that allows any person, who declare themselves to be at least 13 years old, to become registered users. In 2012, Facebook counted over 900 million active users, who can create personal profiles and communicate with friends and other users through private or public messages and receive information about their friends by means of a news feed timeline. To allay concerns about privacy, Facebook enables users to choose their own privacy settings and choose who can see specific parts of their profile. Only a user's name and profile picture are required to be accessible by everyone. The rest of the information on users' pages are by default visible only to friends or to friends-of-friends. Boyd and Hargittai 2010 [5] highlight that while news media were critic toward the company's privacy policies, Facebook has continued to attract more users to its service, indicating that people cares a lot about privacy issues. Another interesting study (Bunloet et Al. 2010 [7]) over more than 7000 students in Thailand, showed that two top reasons why people use Facebook are 1) having conversation with friends and 2) reducing stress.

In recent years there has been a great effort in the analysis of OSN, and there has been a great interest toward the analysis of Facebook in particular (see for example Catanese et Al. 2011 [8]) because, despite it is very challenging to extract data from it for its privacy policy, it is one of the largest and general purpose existing social networks online.

Personality Recognition from Text (PRT henceforth) consists in the automatic classification of authors' personality traits from pieces of text they wrote. This task, that is partially connectecd to authorship attribution, requires skills and techniques from Linguistics, Psychology, Data Mining and Communication Sciences. For example PRT requires some correlations between language features and personality traits (provided by psychologists), a solid background in Data Mining for classification, a good knowledge of communication practices for the social analysis and, most important, a formalized personality schema in order to define classes.

PRT in social networks online is a really challenging task: posts are often very short and noisy, and normal tools for Natural Language Processing (NLP) often perform bad online (Maynard et Al. 2012 [23]). In addition the strict privacy policies of many social network services, included Facebook, put heavy restrictions on data sampling. In this work we analyse a network of Italian Facebook users related by their communicative exchanges, for instance posts and comments, rather than by friendship as other social network analysts did. For example Quercia et Al. 2012 [28] tested the hypothesis that people having many social contacts on Facebook are the ones who are able to adapt themselves to new forms of communication, present themselves in likable ways, and have propensity to maintain superficial relationships, but they found that there is no statistical evidence to support such a conjecture.

We are going to study how users' personality affects the way people interact and communicate in a OSN, rather than study the effect of personality on friendship connections. To do so we developed and tested a personality recognition system. In a previous work [10], we analysed the effects of one specific personality trait, emotional stability, in Twitter. Here we aim to go further in the research including all the five personality traits provided by the Big5, namely: extroversion (e), emotional stability (s), agreeableness (a), conscientiousness (c) and openness to experience (o).

The paper is structured as follows: in the remainder of this section we will provide an introduction to personality in psychology, to the Big5 and to previous and related work. Then, in the next sections, we will provide a description of the personality recognition system, how we tested its performance on Facebook data and we will introduce how we collected the Facebook dataset. In the end we will report and discuss the results of the experiment.

1.2. Personality

According to psychologists (DeYoung 2010 [14]) and neuroscientists (Adelstein et Al. 2011 [1]), personality is defined as an affect processing system that describes persistent human behavioural responses to broad classes of environmental stimuli. It characterises a unique individual and it is involved in communication processes and connected to how people interact one another.

The Standard Way to formalize personality in psychology is the Big5 factor model, introduced in by Norman in 1963 [24]. It emerged from empirical analyses of rating scales, and has become a standard over the years. The five bipolar personality traits, namely Extroversion, Emotional Stability, Agreeableness, Conscientiousness and Openness, have been proposed by Costa & MacCrae 1985 [13].

Extroversion is bound to energy, positive emotions, surgency, assertiveness, sociability and talkativeness. Emotional stability is bound to impulse control, and is sometimes referred by its low pole: neuroticism that is the tendency to experience unpleasant emotions easily, such as anger, anxiety, depression, or vulnerability. Agreeableness refers to the tendency to be compassionate and cooperative rather than suspicious and antagonistic towards others. Conscientiousness is the tendency to show self-discipline, act dutifully, and aim for achievement; planned rather than spontaneous behaviour, organized, and dependable. Openness to experience is bound to the appreciation for unusual ideas, to curiosity, and variety of experience. It often reflects the degree of intellectual curiosity, creativity and a preference for novelty and variety.

According to Digman 1990 [15], there has been a lot of studies in psychology that independently came to the conclusion that five are the right dimensions to describe personality. Despite there is a general agreement on the number of traits, there is no full agreement on their meaning, since some traits are vague. For example there is some disagreement about how to interpret the openness factor, which is sometimes called "intellect" rather than openness to experience.

The Big5 has been replicated in a variety of different languages and cultures, such as

Chinese (Trull & Geary 1997 [31]) and Indian (Lodhi et Al. 2002 [19]). Some researchers, such as Bond et Al. 1975 [4] and Cheung et Al. 2011 [11] suggest that the openness trait is particularly unsupported in asian cultures such as Chinese and Japanese, and that a different fifth factor is sometimes identified. Also the relationship between language and personality has been investigated (Mairesse et al. 2007 [22]), although there are very few applications for personality recognition in languages different from English. This is a good reason for experimenting with personality recognition in Italian.

1.3. Previous and Related Work

There are two main disciplines that are interested in the extraction of personality from OSN: one is computational linguistics, that extracts personality from text, and the other one is the community of social network analysts, that extract information about personality from network configuration (see for example [30]) as well as from other extralinguistic cues (see Bai et Al. 2012 [3]).

The computational linguistics community became interested in PRT first. In 2005 a pioneering work by Argamon et Al. [2] classified neuroticism and extroversion using linguistic features such as function words, deictics, appraisal expressions and modal verbs. Oberlander & Nowson 2006 [25] classified extroversion, stability, agreeableness and conscientiousness of blog authors' using n-grams as features and Naive Bayes (NB) as learning algorithm. Mairesse et Al. 2007 [22] reported a long list of correlations between Big5 personality traits and the features contained in two external resources: LIWC (see Pennebaker et Al. 2001 [27] for details) and RMC (see Coltheart 1981 [12] for details). The former includes features such as word classification, like "positive emotions" or "anger", while the latter includes scores like age of acquisition of word or word imageability. They obtained those correlations from psychological factor analysis on a corpus of Essays (see Pennebaker & king 1999 [26] for details) and developed a supervised system for personality recognition[1].

Luyckx & Daelemans 2008 [20] built a corpus for stylometry and personality prediction from text in Dutch using n-grams of Part-Of-Speech and chunks as features. They used the Myers-Briggs [6] Type Indicator schema, that includes 4 binary personality traits, in place of the Big 5. Unfortunately their results are not comparable to any other because of the different language and personality schema used. In a recent work, Iacobelli et Al. 2011 [17] tested different features, such as stop words or inverse document frequency. They found that word bigrams with stop words treated as boolean features yield very good results for predicting personality in a large corpus of blogs using Support Vector Machines (SVM) as learning algorithm. As is stated by the authors themselves, their model may overfit the data, since the bigrams extracted are very few in a very large corpus. Kermanidis 2012 [18] followed Mairesse 2007 and developed a supervised system based on low level features, such as Part-of-Speech and words associated to psychological states like in LIWC. She trained a SVM classifier on Modern Greek, obtaining good results and demonstrating that correlations between traits and language can be successfully ported to other languages.

Most of the computational linguists used accuracy (acc) as evaluation measure, but re-

[1]demo available online at `http://people.csail.mit.edu/francois/research/personality/demo.html`

cently there is a shift towards f-measure (f), that is representative not just of the precision of a system , but also of its coverage and replicability. Among social network analysts, Golbeck et Al. 2011 [16], predicted the personality of 279 users from Facebook using either linguistic (such as word count) and social network configuration features (such as friend count). Using M5 trees as learning algorithm (M5), they predicted personality trait scores rather than classes, and reported mean absolute error (mae) as evaluation measure. A sum-

Author	Alg.	Measure	Traits	Results (avg).
Argamon 2005	NB	acc	es	0.576
Oberlander 2006	NB	acc	esac	0.539
Mairesse 2007	SVM	acc	esaco	0.57
Iacobelli 2011	SVM	acc	esaco	0.767
Golbeck 2011	M5	mae	esaco	0.115
Celli 2012	-	pacc	esaco	0.631
Kermanidis 2012	SVM	f	esaco	0.687

Table 1. summary of PRT.

mary of the works in personality recognition is reported in table 1. Although the results of different scholars are not directly comparable one another because they used different datasets and different evaluation metrics, we can see that there has been an increase in performance in recent years, which went hand in hand with the renewed interest in PRT.

All the systems described adopt a supervised approach to PRT. This means that they retrieve a model from a finite, labeled set of data by using machine learning techniques, and then apply those models to other, larger, datasets. We suggest that the supervised approach in PRT has some major limitations due to 1) a high risk of overfitting; 2) poor domain and language portability and 3) great difficulties in the annotation of data for training.

We presented the first system for PRT that does not require supervision in Celli 2012 [9]. We exploited correlations between linguistic features and personality traits adapted from Mairesse et Al. 2007 for the prediction of personality in Italian. The advantage of this system is that it automatically adapts to the data at hand, avoiding the risk of overfitting and raising domain and language adaptation. The drawback is that it requires a small labeled set to evaluate the results post-hoc or to estimate accuracy (pacc), as we did in [9]. We report the details about our system in the next section.

2. Automatic Personality Recognition on Facebook Data

2.1. Description of the System

The system for PRT takes as input 1) unlabeled text data with authors; 2) a set of correlations between personality traits and linguistic features: we used the one reported in table 2. We picked up from Mairesse et Al. 2007 the correlations with cross-language features, for instance: punctuation, exclamation marks, parentheses, question marks, quotes, word repetition ratio and average word frequency. The system has two outputs: 1) one model of personality for each author and 2) a confidence score for each personality trait of models generated. This is based on the assumption that one user has one and only one complex personality, and that this personality emerges at various levels from written text as well as

feat.	extr.	em. st.	agree.	consc.	open.
punct.	-.08**	-.04	-.01	-.04	-10**
excl. marks	-.00	-.05*	.06**	.00	-.03
numbers	-.03	.05*	-.03	-.02	-.06**
parenth.	-.06**	.03	-.04*	-.01	.10**
quest. marks	-.06**	-.05*	-.04	-.06**	.08**
quotes	-.05*	-.02	-.01	-.03	.09**
repeat. ratio	-.05**	.10**	-.04*	-.05*	.09**
avg w. freq.	.05*	-.06**	.03*	.06**	.05**

Table 2. Features and correlations with personality traits.

from other extralinguistic cues.

Personality models are formalized as 5-characters strings, each one representing one trait of the Big5. Each character in the string can take 3 possible values: positive pole (y), negative pole (n) and omitted/balanced (o). For example a “ynoon” stands for an extrovert neurotic and not open minded person. Figure 1 represents the pipeline of the system. In the

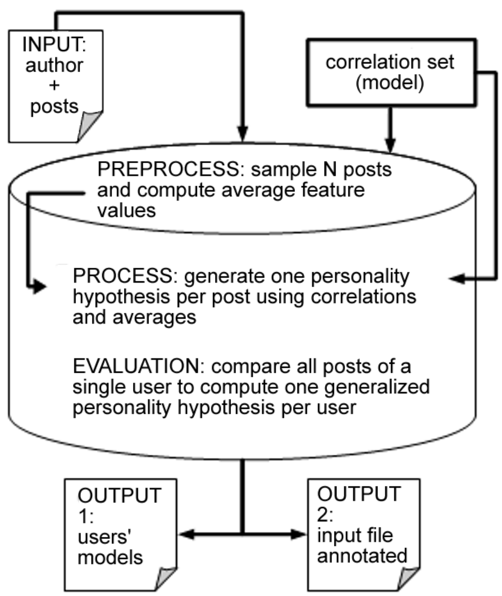

Figure 1. Personality Recognition System pipeline.

preprocessing phase the system samples a small portion of unlabeled data, (the amount can be defined by the user) and analyses the distribution of the features of the correlation set in portion of unlabeled data. With the information about the average feature usage in the dataset we are able to apply filters: for example if a user is found to use more punctuation than the average, the correlation with the punctuation fires, and increase or decrease the score associated to one or more personality traits. This strategy is useful in order to fit the correlation firing to the data, thus avoiding the portability problem.

In the processing phase the system generates one model for each written text, checking for matches of linguistic features provided in the correlation set. If it finds a feature value above the average the system increments or decrements the score associated to the personality trait, depending on a positive or negative correlation. From positive and negative trait scores the system can compute trait confidence scores (tc), defined as $tc = (y - n)$, where $y = \frac{ym}{P}$ and $n = \frac{nm}{P}$. Here *ym* is the count of matches for the positive pole of personality trait and *nm* is the count of matches for the negative pole of the trait. *P* is the count of posts.

In the evaluation phase the system compares all the models generated for each single post of each user and retrieves one model per user. This is based on the idea that, even if a user can express different aspects of personality in different posts, motivated by different goals and situations, we can still catch the personality that the user expresses most of the times by comparing all the posts. Then the system turns the personality scores into classes (if below 0 predicts a negative pole, if above 0 the positive one if it is equal to 0 predicts a "o"). In the evaluation phase the system also computes average confidence and variability for each user. Those measures are computed from the comparison of all models generated from each user's texts. Average confidence (c) gives a measure of the robustness of the personality model. It is defined as $c = \frac{tp}{M}$ where *tp* is the count of personality models matching within the same user (for example "y" and"y", "n" and "n", "o" and "o") and *M* is the total of the models generated for that user. Variability gives information about how much one user tends to write expressing the same personality traits in all the posts. It is defined as $v = \frac{c}{P}$ where *c* is the confidence score and *P* is the count of all user's texts. Most important: the system can evaluate personality only for users that have more than one post, the other users are discarded.

2.2. Testing the Personality Recognition System

We collected a small test set of Facebook data annotated with personality of users. To do so we run an experiment with 35 participants, who took the Big5 personality test. We asked the participants to leave the URL of their Facebook personal page and to write a short essay, minimum 15 lines and maximum 30, on any subject. The participants are all Italian native speaker students aged between 19 and 27, 10 males and 25 females. A couple of them are bilingual Italian-German speakers.

With their consent, we manually collected their public statuses from participants' Facebook pages. We sampled only text, discarding any other type of data. We could sample data only for 23 of them, which we actually put in the dataset. We built 2 datasets, one with Facebook data and the other one with the data from the essays. We produced the gold standard personality models for the users from the results of the Big5 test. We converted the scores of the big5 into the 3-class format used by the system. To do so we turned the scores above 50 into "y" and all the scores below or equal to 50 into "n".

We run the system on both the datasets, using the same features. Results, reported in table 3, reveal that the system achieves the best performance on Facebook data, even if it exploits correlations extracted from essays in English. This means that the way the system uses correlations is adaptable and suitable for social network data.

feature set	avg P	avg R	avg F
off	.45	.7	.558
fb	.547	.735	.628

Table 3. PRT on essays (off) and Facebook (fb) text data.

3. Collection of the Dataset

Sampling data from Facebook is hard. This is due to different factors, like the lack of a public timeline and the strict privacy policy. The former factor prevents from sampling data from users of which we do not have the friendship. The latter factor prevents from having access to the information of users with which we do not have the friendship.

The sampling pipeline can be seen in figure 2. We developed a crowler that exploits

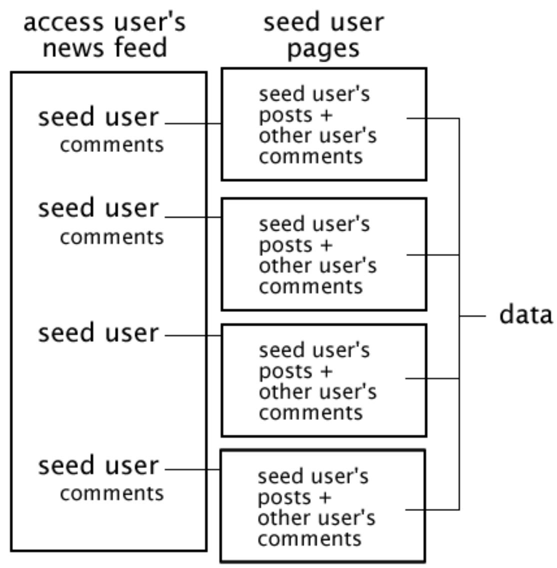

Figure 2. Sampling pipeline.

Facebook's graph API[2] in order to sample users' statuses. The system starts from the news feed of a "access user", who suscribed onto Facebook developer and can take the "access token" key for the API. From the timeline of the access user the system extracts some "seed users" and samples all the statuses and comments written either by the seed users and by the "related users" who interacted with them. The system collects a minimum of 2 posts or comments per user and keeps track of all the users' IDs sampled, in order to avoid duplicates. Finally we filtered out groups and fanpages and we kept only users.

The resulting dataset contains the egonetworks of the seed and related users, as depicted in figure 3. Seed users are linked to the related users with weighted "communicative exchanges" relationships. This means that the more a related user commented a seed user,

[2]http://developers.facebook.com/tools/explorer

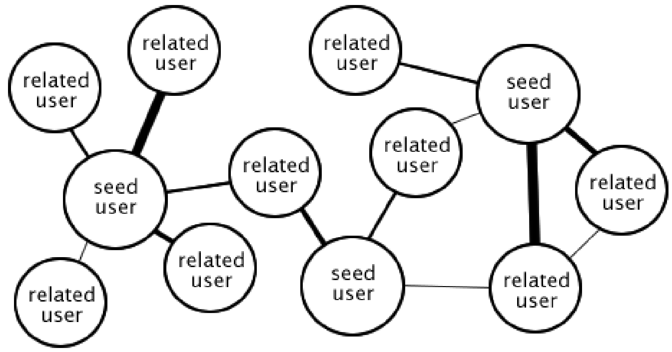

Figure 3. Egonetworks in the dataset.

the more the communicative relationship is considered strong. Although most works use friendships as network connections, we decided to use communicative exchanges because we are interested in the relationship between personality and communicative interactions. In the dataset there are more than 5000 posts and 1100 users. We annotated the personality of each user by means of our personality recognition system.

4. Experiments and Discussion

First of all we retrieved some statistics about the distribution of personality traits in the network and about its topology. The network has a diameter of 14, an average path length of 6.635, average degree centrality of 2.175 and average clustering coefficient of 0.017. This indicates that it is a small network where users have on average a couple of comment-relations each one and with low clustering level. Centrality measures and clustering coefficient have skewed distributions, meaning that a few users have high values and most of them have very low values. The distribution of personality traits, reported in table 4, highlights the low number of extroverted, mentally closed and uncooperative people in the

trait	y	o	n
extr.	6.2%	66.4%	27.4%
em. st.	13.7%	49.9%	36.4%
agree	31.9%	65%	3.1%
consc.	13.2%	50.4%	36.4%
open.	27.9%	62.2%	9.9%

Table 4. Distribution of personality traits in the network.

network. We suggest that this might be due to the personality of the access user ("noyyy"), that influences the selection of people who are in the network. We will refer to this problem as the "access user bias", that is related to the sampling procedure and does not take place in those networks, like Twitter, where there is a public timeline available.

We analysed the relationship between personality and interactions by computing the association between personality traits and some topology measures, like degree centrality, correlation coefficient and edge weight. In order to do that we measured association scores by computing $as = \frac{bti}{td}$, where *bti* are the 10 most frequent personality traits associated to each topology measure used, and *td* is the trait distribution reported in table 4.

Results, reported in table 5, show several interesting phenomena. First of all that introverted and open minded users have the highest degree centrality in the network. In other words they are the ones that are more central and more prone to catch conversations. It is not a surprise that open minded users are in this position, but it is very interesting to note that introvert people have a high degree centrality score too. A closer look to the data reveals that the open minded and introvert traits come often together in the dataset. We suggest this might be due again to the access user bias, as we found previous work [10] that there is a general tendency to have conversations between users that share the same traits. The highest edge weight scores are again associated to open minded and introverted users. This means that those users have the strongest links, in other words the highest number of comments. We interpret this as a consequence of the position those users occupy in the topology of the network. Also Agreeable and emotionally stable users have high degree centrality and edge weight scores, indicating that those personality traits play a role in being influential in a conversation network. The distribution of high edge weights is very skewed: there are very few strong links and really a lot of links with low weight.

degree centr.	extr.	em. st.	agree.	consc.	open.
y	0.774	1.387	2.687	2.167	**3.244**
o	0.215	0.381	0.22	0.472	0.077
n	**2.956**	1.701	0	1.308	0.485
edge weight	extr.	em. st.	agree.	consc.	open.
y	0	2.335	2.351	1.923	**3.405**
o	0.15	0.2	0.307	0.198	0.08
n	**3.284**	0.364	1.61	1.785	0
clustering c.	extr.	em. st.	agree.	consc.	open.
y	1.396	0.912	1.567	0.477	1.57
o	0.848	0.501	0.674	0.869	0.905
n	1.016	1.717	**2.032**	1.374	0

Table 5. Association scores.

The personality trait associated to high clustering coefficient scores is low agreeableness. If clustering coefficient is related to users' connectedness and links represent comment relationships here, we can interpret this fact as a hint that uncooperative users tend to participate in many conversations in order to debate in a polemic way. The distribution of clustering coefficient scores is very skewed too.

The outcomes of this experiment show the behaviour of a network of interacting users visible to the access user, hence are not generalizable, but yet interesting to study the role that personality traits play in social interactions in a micro network.

5. Conclusion

In this work we have sampled a network of communications between Italian users in Facebook, sampled from one access user's timeline. We automatically annotated it with personality traits in order to analyse how people's personality affects interactions online. We tested the system used for the annotation on a gold standard, obtained from 23 Italian Facebook users which took the Big5 personality test, and provided us with the public data from their timelines. The system proved to label correctly 62% of the data.

From the analysis of the most frequent traits associated to topology measures like degree centrality and correlation coefficients, emerged that open minded and introvert users have the highest degree centrality and the strongest links. We interpreted this evidence as introvert and open minded users (those traits come frequently together in the dataset) tend to be very interested to the information that passes through the network, and tend to post interesting (high commented) statuses. Another interesting result is that the users that have high correlation coefficient have low agreeableness. We interpreted this fact as as a hint that uncooperative users tend to participate in many conversations in order to debate in a polemic way.

The results show how people's personality can be successfully analysed with a quantitative approach on large scale data, yielding very interesting findings. It is not easy to interpret the results, but the same difficulty is found in much of the quantitative sociology based on big data [29]. We suggest that pairing personality recognition with sentiment analysis or topic extraction would make it more informative and easier to interpret. We also suggest that the comparison of personality recognition in communication exchanges and friendship relations, for example using the multi-layer model proposed by Magnani & Rossi 2011 [21], would bring out useful information.

The access user bias, that is due to the restrictions imposed by Facebook and to the lack of a public timeline, prevents from the generalization of those results. Yet it is interesting to observe that a micro network is filtered by the access user according to personality, among other factors. This underlines one more time the importance of personality recognition in the study of social networking.

References

[1] Adelstein J.S, Shehzad Z, Mennes M, DeYoung C.G, Zuo X-N, Kelly C, Margulies D.S, Bloomfield A, Gray J.R, Castellanos X.F, Milham M.P. (2011). Personality Is Reflected in the Brain's Intrinsic Functional Architecture. In *PLoS ONE* 6(11). pp. 1–12.

[2] Argamon, S., Dhawle S., Koppel, M., Pennebaker J. W. (2005). Lexical Predictors of Personality Type. In *Proceedings of Joint Annual Meeting of the Interface and the Classification Society of North America*. pp. 1–16.

[3] Bai.,S. Zhu.,T. Cheng.L. 2012. Big-Five Personality Prediction Based on User Behaviors at Social Network Sites. In *eprint arXiv:1204.4809*. Available at `http://arxiv.org/abs/1204.4809v1`.

[4] Bond, M.H. Nakazato, H.S. Shiraishi, D. (1975). Universality and distinctiveness in dimensions of Japanese Person Perception. In *Journal of Cross-Cultural Psychology*. 6. pp.346–355.

[5] Boyd, D. Hargittai, E. (2010). Facebook Privacy Settings: Who Cares? In: *First Monday* 15 (8). Available online at `http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3086/2589`.

[6] Briggs, I. Myers, P.B. (1980). *Gifts differing: Understanding personality type*. Mountain View, CA: Davies-Black Publishing.

[7] Bunloet, A., Saikeaw, K. R., Tengrungroj, M., Nalinthutsanai, N., Mungpooklang, T., Dabpookhiew, P., Winkam, T., Arayasilapatorn, N. Premgamone, A., Rattanasiri, A., Chaosakul, A. (2010). Analysis of Facebook Usage by College Students in Thailand. In *25th International Technical Conference on Circuit Systems, Computers and Communications (ITC-CSCC 2010)*. pp. 107–111.

[8] Catanese, S.A, De Meo P., Ferrara, E., Fiumara, G., Provetti, A. (2011). Crawling Facebook for Social Network Analysis Purposes. In *Proceedings of WIMS '11: International Conference on Web Intelligence, Mining and Semantics ACM*. pp. 1–8.

[9] Celli, F., (2012). Unsupervised Personality Recognition for Social Network Sites. In *Proceedings of ICDS*, pp. 59–62.

[10] Celli, F., Rossi, L. (2012). The role of Emotional Stability in Twitter Conversations. In *Proceedings of Workshop on Semantic Analysis in Social Media, in conjunction with EACL*, pp. 1–8.

[11] Cheung, F. M., van de Vijver, F. J. R., Leong, F. T. L. (2011). Toward a new approach to the study of personality in culture. In *American Psychologist*, Advance online publication. pp. 1–11.

[12] Coltheart, M. (1981). The MRC psycholinguistic database. In *Quarterly Journal of Experimental Psychology*, 33A, pp. 497-505.

[13] Costa, P.T., Jr. McCrae, R.R. (1985). The NEO Personality Inventory manual. In *Psychological Assessment Resources*. pp. 5–13.

[14] DeYoung, C.G. (2010). Toward a Theory of the Big Five. In *Psychological Inquiry*. 21: pp. 26–33.

[15] Digman, J.M. (1990). Personality structure: Emergence of the five-factor model. In *Annual Review of Psychology* 41: pp. 417–440.

[16] Golbeck, J. and Robles, C., and Turner, K. (2011). Predicting Personality with Social Media. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, pp. 253–262.

[17] Iacobelli, F., Gill, A.J., Nowson, S. Oberlander, J. (2011). Large scale personality classification of bloggers. In *Lecture Notes in Computer Science (6975)*, pp. 568–577.

[18] Kermanidis, K.L. (2012). Mining Authors' Personality Traits from Modern Greek Spontaneous Text. In *4th International Workshop on Corpora for Research on Emotion Sentiment & Social Signals, in conjunction with LREC12*. pp. 90–94.

[19] Lodhi, P. H., Deo, S., Belhekar, V. M. (2002). The Five-Factor model of personality in Indian context: measurement and correlates. In R. R. McCrae J. Allik (Eds.), *The Five-Factor model of personality across cultures*. N.Y.: Kluwer Academic Publisher. pp. 227–248.

[20] Luyckx K. Daelemans, W. (2008). Personae: a corpus for author and personality prediction from text. In: *Proceedings of LREC-2008, the Sixth International Language Resources and Evaluation Conference*. pp. 2981–2987.

[21] M. Magnani and L. Rossi. The ML-model for multi-layer social networks. In *Proceeding of 2011 International Conference on Advances in Social Networks Analysis and Mining, IEEE Computer Society*. pp. 5–12.

[22] Mairesse, F. and Walker, M. A. and Mehl, M. R., and Moore, R, K. (2007). Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. In *Journal of Artificial intelligence Research*, 30. pp. 457–500.

[23] D. Maynard and K. Bontcheva and D. Rout. (2012). Challenges in developing opinion mining tools for social media. In *Proceedings of @NLP can u tag usergeneratedcontent?! Workshop at LREC 2012*. pp. 15–22.

[24] Norman, W., T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality rating. In *Journal of Abnormal and Social Psychology*, 66. pp. 574–583.

[25] Oberlander, J., and Nowson, S. (2006). Whose thumb is it anyway? classifying author personality from weblog text. In *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics ACL*. pp. 627–634.

[26] Pennebaker, J. W., King, L. A. (1999). Linguistic styles: Language use as an individual difference. In *Journal of Personality and Social Psychology*, 77. pp. 1296-1312.

[27] Pennebaker, J. W., Francis, M. E., Booth, R. J. (2001). *Inquiry and Word Count: LIWC 2001*. Lawrence Erlbaum, Mahwah, NJ.

[28] Quercia, D., Lambiottez, R., Stillwell, D., Kosinskiy, M., Crowcroft, J. (2012). The Personality of Popular Facebook Users. In *Proceedings of ACM CSCW 2012*. pp 1–10.

[29] Scott, J. (2011). Social Network Analysis: developements, advances, and prospects. In *Social Network Analysis and Mining*, 1(1). pp. 21–26.

[30] Staiano J, Lepri B, Aharony N, Pianesi F, Sebe N, Pentland A.S. (2012). Friends dont Lie - Inferring Personality Traits from Social Network Structure. In *Proceedings of International Conference on Ubiquitous Computing*. pp. 324–334.

[31] Trull, T. J. Geary, D. C. (1997). Comparison of the big-five factor structure across samples of Chinese and American adults. *Journal of Personality Assessment* 69 (2). pp. 324–341.