The current issue and full text archive of this journal is available at
www.emeraldinsight.com/1468-4527.htm

Received 13 June 2012
First revision accepted
27 February 2013

How do personality traits affect communication among users in online social networks?

Jose Maria Balmaceda, Silvia Schiaffino and Daniela Godoy
Department of Computer Science,
Universidad Nacional del Centro de la Provincia de Buenos Aires Tandil,
Argentina

Abstract

Purpose – The purpose of this work is to analyse the relationships between the personality traits of linked users in online social networks. First the authors tried to discover relation patterns between personality dimensions in conversations. They also wanted to verify some hypotheses: whether users' personality is stable throughout different conversation threads and whether the similarity-attraction paradigm can be verified in this context. They used the five factor model of personality or Big Five, which has been widely studied in psychology.

Design/methodology/approach – One of the approaches to detect users' personalities is by analysing the language they use when they talk to others. Based on this assumption the authors computed users' personality from the conversations extracted from the MySpace social network. Then the authors analysed the relationships among personality traits of users to discover patterns.

Findings – The authors found that there are patterns between some personality dimensions in conversation threads, for example, agreeable people tend to communicate with extroverted people. They confirmed that the personality stability theory can be verified in social networks. Finally the authors could verify the similarity-attraction paradigm for some values of personality traits, such as extroversion, agreeableness, and openness to experience.

Originality/value – The results the authors found provide some clues about how people communicate within online social networks, particularly who they tend to communicate with depending on their personality. The discovered patterns can be used in a wide range of applications, such as suggesting contacts in online social networks. Although some studies have been made regarding the role of personality in social media, no similar analysis has been done to evaluate how users communicate in social media considering their personality.

Keywords Social networks, Social media, Personality traits, Association rules, Users' behavior

Paper type Research paper

Introduction

Personality is a subject that has been studied for decades by many researchers. It involves the particular combination of emotional, attitudinal, and behavioural response patterns of an individual. Many researchers have tried to find a group of traits that describe the personality of an individual. Tupes and Christal's work (1961) was the first that identified five recurrent factors in the analysis of personality; then Norman (1963) replicated that work. Although the five factors were not elaborated further throughout the 1960 s and 1970 s, in the 1980 s, however, researchers concluded that they are fundamental dimensions of personality, found in self-reports and ratings, in natural languages and theory based questionnaires (John, 1990). There are many other

This work has been partially funded by ANPCYT (Argentina) through Project PICT-2011-0366.

Online Information Review
Vol. 38 No. 1, 2014
pp. 136-153
© Emerald Group Publishing Limited
1468-4527
DOI 10.1108/OIR-06-2012-0104

works that have offered evidence for the existence of five personality traits (Noller *et al.*, 1987; Waller and Ben-Porath, 1987; Zuckerman *et al.*, 1989). Nowadays the five-factor model (FFM) of personality is considered correct in its representation of the structure of traits.

The FFM or "Big Five" factors of personality are five broad domains or dimensions of personality that are used to describe human personality (Costa and McCrae, 1992). The Big Five factors are openness to experience (inventive/curious vs consistent/cautious), conscientiousness (efficient/organised vs easy-going/careless), extraversion (outgoing/energetic vs solitary/reserved), agreeableness (friendly/compassionate vs cold/unkind), and neuroticism (sensitive/nervous vs secure/confident). We decided to use this model because it is the one that has gained consensus in the psychology field and it is one of the most widely used in academic research (Tupes and Christal, 1961; Norman, 1963; Digman and Inouye, 1986; McCrae and Costa, 1989; Digman, 1990; Goldberg, 1992; Mount and Barrick, 1998; Aguilar *et al.*, 2007; Feldt *et al.*, 2010). Moreover the FFM has been used in various domains. For example it has been used by organisations when hiring personnel. Different studies have linked personality to job performance and proficiency (Mount and Barrick, 1998), and to innovation and leadership (Steel *et al.*, 2012). Personality also affects the way in which we engage in social networks. In this regard some recent works have studied personality in social environments (Dolgova *et al.*, 2010; Mehra *et al.*, 2001; Ozer and Benet-Martínez, 2006; Schrammel *et al.*, 2009; Uesugi, 2011). However these works did not aim to determine how users communicate in social media according to their personality, but instead how personality affects other aspects, such as social media adoption. To the best of our knowledge no previous works have analysed the role of personality in social media relationships.

In this work we aim at studying whether there is a relationship among the personalities of users communicating within a social network, and analysing how personality affects the way in which each individual communicates with others in such a context. In particular we suggest that the similarity-attraction paradigm or homophily (Byrne, 1971; Clore and Byrne, 1974), which predicts that people tend to build relationships with similar others, could also be verified in social networks in terms of personality.

Moreover we analysed whether the personality stability theory (Cobb-Clark and Schurer, 2011), which says that personality tends to be stable across time and over different situations, could also be verified in social networks, taking into account the diversity of topics users write about.

Finally we consider that characterising users' behaviour in social networks creates opportunities for better interface design, richer studies of social interactions, and improved design of content distribution systems (Benevenuto *et al.*, 2009). Specifically understanding how users communicate with others can give some hints about which types of contacts a user would select, and thus, help to recommend contacts or friends who have a certain personality. With this aim, we tried to discover relationship patterns among linked users' personalities.

There are different ways of assessing personality. The most widely used one is through questionnaires. Several rating instruments have been developed to measure the Big Five dimensions (Benet-Martínez and John, 1998; Costa and McCrae, 1992; Goldberg, 1992; John and Srivastava, 1999). Another approach is by analysing the language people use to communicate with others (Mairesse *et al.*, 2007). The lexical hypothesis argues that all important individual differences are encoded in trait terms and that by decoding them we can determine an individual's personality.

In this work we used the second approach. In particular we based our study on a recent work that automatically detects users' personalities from text and conversations, following the lexical approach. Knowing that there is evidence that personality interacts with, and affects, aspects of linguistic production (Watson and Clark, 1992), we tried to identify relationship patterns in social networks, particularly in MySpace.

The rest of the paper is organised as follows. In the next section we provide some background knowledge by describing the main concepts regarding personality, how it can be automatically detected from users' conversations and the relationship between personality and social media. We also describe the software application we used to recognise users' personality and how it works. In the subsequent section we present our findings. In the following section we discuss the relationships we have found and their implications for the design and development of social software. Finally we provide our conclusions, limitations of our study and future work.

Background and related works

In this section we introduce some key concepts in the study of personality, how it can be detected by analysing interactions among individuals, and we discuss some studies that analyse the relationship between personality and social media.

Personality: main concepts

Personality is the particular combination of emotional, attitudinal, and behavioural response patterns of an individual. Different models or theories have been proposed. Despite some known limits (Eysenck, 1991; Paunonen and Jackson, 2000), over the last 50 years the Big Five model has become a standard in psychology. The FFM is a hierarchical model of personality traits with five broad factors, which represent personality at the broadest level of abstraction. Each bipolar factor (e.g. extraversion vs introversion) summarises several more specific facets (e.g. sociability), which in turn, subsume a large number of even more specific traits (e.g. talkative, outgoing). The FFM suggests that most individual differences in human personality can be classified into five broad, empirically derived domains.

The five dimensions are the following (Norman, 1963):

(1) extraversion vs introversion (sociable, assertive, playful vs aloof, reserved, shy);

(2) emotional stability vs neuroticism (calm, unemotional vs insecure, anxious);

(3) agreeableness vs disagreeableness (friendly, cooperative vs antagonistic, fault finding);

(4) conscientiousness vs unconscientiousness (self-disciplined, organised vs inefficient, careless); and

(5) openness to experience (intellectual, insightful vs. shallow, unimaginative).

There are two typical paths to the FFM. The first, and also the most widely used, is through questionnaires. Several rating instruments have been developed to measure the Big Five dimensions. The most comprehensive instrument is Costa and McCrae's

(1992) revised 240-item NEO Personality Inventory (NEO-PI-R), which enables the measurement of the Big Five domains and six specific facets within each dimension. Taking about 45 minutes to complete, the NEO-PI-R is too lengthy for many research purposes and so a number of shorter instruments are commonly used. Three well-established and widely used instruments are the 44-item Big Five Inventory (Benet-Martínez and John, 1998; John and Srivastava, 1999), the 60-item NEO Five-Factor Inventory (Costa and McCrae, 1992), and Goldberg's instrument that comprised 100 trait descriptive adjectives (Goldberg, 1992). John and Srivastava (1999) estimated that they take approximately 5, 15, and 15 minutes to complete, respectively. Recognising the need for an even briefer measure of the Big Five, Saucier (1994) developed a 40-item instrument derived from Goldberg's 100-item set. However, completing questionnaires can be a tedious and error-prone activity for users, and thus automatic techniques are needed.

The second path to the FFM is the lexical approach that analyses the terms in the natural language used by an individual to communicate. The lexical hypothesis argues that all important individual differences are encoded in trait terms in natural language; by decoding these terms, we can discover the basic dimensions of personality. The decoding has to be able to determine what feelings are being shown by the speaker. Some pioneering works in this direction were the following. Allport and Odbert (1936) abstracted terms from a dictionary, while Cattell (1946) formed them into synonym clusters and then created rating scales contrasting groups of adjectives. Then Tupes and Christal (1961) obtained observer ratings on these 35 scales and factored them.

In our research we used a recent work that automatically detects users' personality from texts and conversations, following the lexical approach. Thus, knowing that there is evidence that personality affects aspects of linguistic production (Watson and Clark, 1992), we tried to identify relational patterns between users in social networks taking into account their conversations. The tool we used applies the models proposed by Mairesse *et al.* (2007) as described in the following.

Automatic detection of personality

Thus far there has been little work on the automatic recognition of personality traits (Argamon *et al.*, 2005; Oberlander and Nowson, 2006, Mairesse *et al.*, 2007). Argamon *et al.* (2005) published one of the first works on automatic detection of personality. The authors focused on determining two dimensions of personality (neuroticism and extraversion) from casual written text applying techniques such as Naive Bayes and Sequential Minimal Optimisation. They considered four different sets of lexical features for this detection: a standard function word list, conjunctive phrases, modality indicators, and appraisal adjectives and modifiers. For both dimensions they reported binary classification accuracy of around 58 per cent: an 8 per cent absolute improvement over their baseline. Oberlander and Nowson's (2006) work follows Argamon's approach, but they improved the performance and they added agreeableness and conscientiousness dimensions. Mairesse *et al.* (2007) made this classification using regression and ranking modelling techniques and language cues, and their results were the first to demonstrate statistically significant results for texts and to recognise personality from conversations.

Recently some works have addressed the detection of personality from social networks. For example, Golbeck *et al.* (2011) demonstrate that a user's personality can be accurately predicted through the publicly available information on their Facebook profile. Quercia *et al.* (2011) analyse the relationship between personality and different types of Twitter users, including popular and influential users. Their approach is based on three elements publicly available on Twitter profiles: following, followers, and listed counts. One of the conclusions of this study is that personality can be inferred from public data.

Since most social networks are rich in text, we based personality prediction on users' writing so that the conclusions can be applied to multiple social network services. In this work we used a software application to detect users' personality, named Personality Recogniser (http://people.csail.mit.edu/francois/research/personality/recogniser.html), based on the models proposed by Mairesse *et al.* (2007). The Personality Recogniser is a Java (www.java.com/) application that computes estimates of personality scores along the five dimensions we are considering.

The application carries out the detection process from text written by a person whose personality is being assessed. The basis of this method is that there is a correlation between a range of linguistic variables and personality traits, across a wide range of linguistic levels, including acoustic parameters (Scherer, 1979), lexical categories (Fast and Funder, 2008; Mehl *et al.*, 2006; Pennebaker and King, 1999), and n-grams (Oberlander and Gill, 2006). Pennebaker and King (1999) identified many linguistic features associated with each of the Big Five personality traits. They used their linguistic inquiry and word count tool to count word categories of essays written by students whose personality had been assessed using a questionnaire. The authors found significant correlations between their linguistic dimensions and personality traits. Oberlander and Gill (2006) studied correlates of emotional stability: they found that neurotics use more concrete and frequent words. Coltheart (1981) deploys the MRC psycholinguistic database, which contains statistics for over 150,000 words, such as frequency of use and familiarity.

The Personality Recogniser software uses models trained from experiments made in the works mentioned before and also extra tests described by Mairesse *et al.* (2007). Based on this training the Recogniser extracted a list of features related to each personality trait that could be identified from text. For example some identified language cues for extraversion at syntax level are: many verbs, adverbs and pronouns; few words *per se*ntence; few articles; few negations. There are also other levels analysed, such as lexicon or speech, with cues for each personality trait. However not all of these features are helpful for recognising all personality types. Mairesse *et al.* (2007) performed a feature selection process in order to use just those features that provided relevant information about personality. Then they explored the use of classification, regression and ranking models. In their experiments the best performance was obtained by a classification model that reached 74 and 73 per cent of precision to classify extraversion and emotional stability, respectively; and 65 per cent of precision to classify openness to experience. These are the best results reached by automatic techniques to detect personality we have found in the literature.

Personality and social media

Some works in the literature studied the relationship between users' personality and different aspects of social media. For example, personality has been used as a factor that could improve teamwork performance in a variety of domains. Dolgova *et al.*

(2010) suggest that there is a potential mismatch between social network structure in different stages of the innovation process, and this mismatch is caused by individuals' personalities. The authors proposed a conceptual framework that helps to understand why people create markedly different patterns of social relations in the workplace and how this relation formation process and personality influence innovation. Mehra *et al.* (2001) examined how different personalities relate to social structure, and how social structure and personality combine to predict work performance. Specifically the authors proposed to use self-monitoring orientation (one of the many personality variables) to predict an individual's position in the social network and how this affects performance when working in groups. Uesugi (2011) studied the relation between personality traits in the FFM and SNS usage as well as attitudes towards protecting privacy.

There are also other studies that have indicated that the structural position of an individual in social networks is in part shaped by their personality (Dolgova *et al.*, 2010), and how this affects relations with other group members. Klein *et al.* (2004) hypothesised that individuals' demographic characteristics, values, and personality influence their acquisition of central positions in their teams' social networks. Finally personality could be seen as a factor that helps us to understand why there are differences in the way each individual reacts and interacts in a social environment. Ozer and Benet-Martínez (2006) analysed people's personality as the cause of consequential outcomes at individual, interpersonal and social/institutional levels. The authors discovered a relation between personality traits and consequential outcomes at these levels. Correa *et al.* (2010) studied personality traits as important factors for the engagement of users with social media. The authors investigated the relation between personality and social media use, mainly focusing on whether people use social media or not. They found, for example, that extraversion is positively related to social media use, whereas emotional stability is negatively related.

In contrast to these previous works, we use personality traits as a factor that could affect how and with who people communicate in social networks.

Analyses and findings

In this section we first describe the main goals of our study, then the dataset used, different methodologies for automatic personality detection and finally, different approaches for pattern discovery.

Goals

The goal of our analysis was determining whether there are patterns in the personality types of linked users in a social network. As stated before, we wanted to verify whether people tend to communicate with others with similar personalities in social networks (homophily). We studied each of the five dimensions separately. Moreover we tried to discover whether there are relations between different dimensions of personality or between values in the same dimension. In addition we wanted to determine whether the stability theory is verified in social networks, not only through time but also over different discussion topics.

Dataset

For our study we used a dataset created with data from MySpace (www.myspace.com/), which is a popular social networking site. MySpace offers its registered users, among other things, the possibility to participate in discussion forums about several predefined topics. In this way anyone can start a new thread in these forums with a question, and participate freely in a thread created by another user expressing their opinion or knowledge. Thus there is a link between the person who started a thread and the people who participated in it, given their shared interest.

The threads that form this dataset were chosen from three different forum topics: campus life, news and politics, and movies. The dataset was automatically crawled by Fundación Barcelona Media (http://caw2.barcelonamedia.org/node/7) and it contains about 380,000 comments on 16,346 threads dealing with one of those three topics.

We created a database table that contains information about how many posts each user made in all the threads (33,407 users in total). More than 18,000 users have only posted once and almost 5,000 users have made two posts. However, for our analysis we only used those users who have made more than two posts (10,117 users) because we needed enough text to obtain good results in the personality recognition step. We also found some users posted only spam, publicity, and text with strange symbols or too short text messages, which caused Personality Recogniser to return an invalid value. We excluded such users, keeping a total of 5,002 users.

Detection of users' personality

Methodology 1: Considering all posts. First we obtained the texts of all the posts made by each user and we joined them in a unique text in order to execute a detection process for each user. Using the Personality Recogniser Software we calculated the personality traits for users. This application returns five float point values between 0 and 7 that represent the five personality factor values calculated for the user who wrote the text. For example, considering the extroversion dimension, a value of 0 indicates that person has a tendency to be introverted, whereas a value of 7 indicates that a person has a tendency to be extroverted.

When users included strange symbols or words in other languages, or used HTML code, the detection returned invalid values. Thus we deleted those instances that contained values smaller than 0 or bigger than 7. Figure 1 shows the distribution of users' personalities separated in five graphics corresponding to the Big Five dimensions. Table I summarises the average and deviation values for these distributions.

Methodology 2: Considering initial posts. Taking into account that a thread could only be started with an initial post that talks about a topic or asking a question, we decided to do the same analysis as previously but using only these starting texts. We did this analysis because one of the identified language cues for introversion/extroversion in conversational behaviour is whether a person listens to a conversation or starts it. The description of personality traits says that extroverted people usually start conversations, whereas introverted ones prefer to listen to others. Table II shows a summary of the personalities obtained considering only initial texts.

Methodology 3: Considering different threads. Considering that it has been demonstrated that personality tends to be stable across time and over different situations (Cobb-Clark and Schurer, 2011), we compared Personality Recogniser's

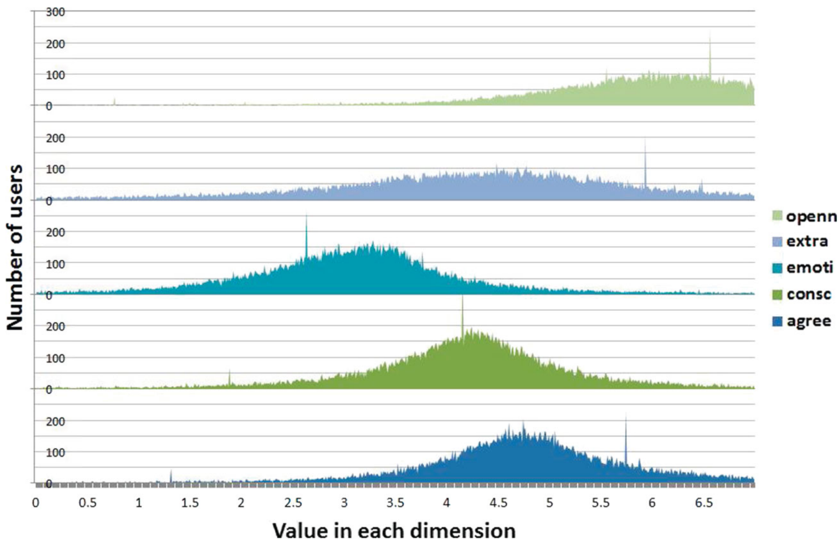

Figure 1.
Distribution of the five personality dimensions among users

Table I.
Average and deviation of personality dimension distribution (all posts)

	Agreeableness	Conscientiousness	Emotional stability	Extraversion	Openness
Average	4.67	4.14	3.11	4.19	5.61
Std Dev.	1.06	1.05	1.10	1.42	1.08

Table II.
Average and deviation of personality dimensions distribution (initial texts)

	Agreeableness	Conscientiousness	Emotional stability	Extraversion	Openness
Average	4.83	4.31	3.11	4.26	5.46
Std Dev.	1.08	1.11	1.16	1.61	1.11

results with input text from different threads, in which people could act in a different way, depending on the thread topic.

To carry out this evaluation, we calculated N_i-times the personality of each user (N_i is the number of threads in which $user_i$ has participated), using all the text written by $user_i$ in the $thread_j$ (j between 1 and N_i). We calculated the average (AVG_i) and deviation (STD_i) of the personalities of each user, and then we calculated the average and deviation of all STD_i. We also calculated the differences between the personality of each user using text from each thread and the personality obtained from all posts (subsection 3.3.1). Table III shows these results, where each cell has a value between 0 and 7. A big value in a cell means that there is a big difference between the values of the personality trait.

Findings. In Figure 1 we can see that each dimension has a different distribution, central points and concentrations. For example the extraversion distribution is the least concentrated and lowest because users have different values in this dimension. In contrast agreeableness and conscientiousness have the most concentrated distribution. In Table I we can observe that almost all people in this dataset are open to experiences.

Regarding extraversion we can see that there are also more extroverted people than introverted ones, but the corresponding distribution is centred between values 4 and 5. However the emotional stability dimension is the only one that has more people with low values than people with high ones.

Comparing Table II with Table I we can see that the averages of extraversion, agreeableness and conscientiousness have increased a little, but the standard deviation also increased. With this information we cannot affirm that people who start a conversation have on average a high value of extraversion.

Finally analysing Table III we found that the values of average and deviation are very low, both between threads and between each thread and the value of personality calculated from all texts. This fact suggests that the stability personality theory is verified.

Discovering relations between linked users' personalities

We consider that there is a relation between two users when they both have posted something in the same thread, because this action means that these users have shared ideas or knowledge. For example if a user writes a post in a thread, we consider that there is a relation between this user and the other users that have posted before in the same thread.

We used different approaches to find patterns. First we analysed correlations between the numbers of posts corresponding to different personality traits. Then, in order to verify that people tend to be linked to personality-similar ones (homophily), and also to discover relationships between users' personalities, we performed two different analyses. First we analysed all the relations included in the dataset, looking for who people tend to communicate with depending on their personality. The second type of analysis we made consisted in mining association rules.

Methodology 1: Analysing relationships between posts. In order to discover behavioural patterns inside a conversation, we made an analysis that links people who participated in the same thread taking into account only the text written in each post to detect the personality. We used text from threads that had more than two posts in order to exclude those threads that cannot be considered a conversation. Our objective was to find a relation between personality traits inside a discussion, and for this, we calculated the personality of users who wrote each text, and made an analysis with the remaining texts in the same thread. For each personality trait, for example extroversion, we counted the number of introverted and extroverted posts in a thread and then looked for relations. For example, given that each thread has a number of introverted posts, and a number of extroverted posts, we tried to find significant

Table III. Average and deviation

	Agreeableness	Conscientiousness	Emotional stability	Extraversion	Openness
Average (A)	0.67	0.69	0.73	0.74	0.58
Std Dev. (A)	0.44	0.44	0.45	0.43	0.38
Average (B)	0.63	0.66	0.70	0.70	0.54
Std Dev. (B)	0.85	0.88	0.92	0.89	0.73

Note: (A) differences between values of personality calculated from text of different threads; (B) differences between values of personality calculated from text of each thread and the global value of personality for that user relations between these traits. The same analysis was done with the other personality dimensions. We considered the middle range (values between 3 and 4) of each dimension as neutral and those posts were not taken into account.

Methodology 2: Quantitative analysis. In this analysis each relation is composed of the personalities of two users. For each post in a thread, there are a number of candidate relations equal to the number of users who has posted a message in the same thread before, most of which are not statistically significant. From this dataset we obtained more than 6 million relations. We supposed that if the similarity-attraction paradigm is true in social networks, we would observe many relations between people with similar personalities. Figure 2 shows the results obtained for each dimension separately.

>*Methodology 3: Discovering association rules*. Association rules (Agrawal and Srikant, 1994) imply an association relationship among a set of items in a given domain. Association rule mining is commonly stated as follows: Let I be a set of items and D be a set of transactions, each consisting of a subset X of items in I. An association rule is an implication of the form X $\rightarrow$ Y, where $X \subseteq I$, $Y \subseteq I$ and $X \cap Y = ø$. X is the antecedent of the rule and Y is the consequent. The rule has support s in D if s per cent of the transactions in D contains $X \cup Y$. The rule X $\rightarrow$ Y holds in D with confidence c if c per cent of transactions in D that contain X also contain Y. Given a transaction database D, the problem of mining association rules is to find all association rules that satisfy: minimum support (called *minsup*) and minimum confidence (called *minconf*).

In our analysis each transaction consists of the different values of the personality traits of two users involved in a relationship. Thus we tried to discover association relationships between personality traits in user 1 and personality traits in user 2. We used the Knime (www.knime.org) tool and the Apriori algorithm to discover association rules using a value of $minconf = 0.7$ (70 per cent) and a value of $minsup = 0.1$ (10 per cent).

	Extro.	Intro.
Extro.	80%	10%
Intro.	9%	1%

(A) Extraversion

	Neurotic	Stable
Neurotic	6%	17%
Stable	19%	58%

(B) Emotional stability

	Agree.	Disagree.
Agree.	75%	10%
Disagree.	12%	2%

(C) Agreeableness

	Consc.	Unconsc.
Consc.	42%	22%
Unconsc.	20%	16%

(D) Conscientiousness

	Open	Not open
Open	96%	2%
Not open	2%	0%

(E) Openness to experience

Figure 2.
Analysis of personalities in users' relations

The Apriori algorithm, although one of the most widely used for association mining, returns many rules that might be irrelevant for our purposes. To filter out rules, we use templates or constraints (Klementinen *et al.*, 1994) that select those rules that are relevant to our goals. For example we are interested in those association rules having as antecedent personality traits of user 1 and as consequent personality traits corresponding to user 2. Rules containing other combinations of attributes are not considered. To eliminate redundant rules, we used a subset of the pruning rules proposed by Shah *et al.* (1999). Basically these pruning rules state that given the rules A, B $\rightarrow$ C and A $\rightarrow$ C, the first rule is redundant because it gives little extra information. Thus it can be deleted if the two rules have similar confidence values. Similarly, given the rules A $\rightarrow$ B and A $\rightarrow$ B, C, the first rule is redundant since the second consequent is more specific. Thus the redundant rule can be deleted provided that both rules have similar confidence values.

Findings: Relations discovered. With the first methodology we discovered that there is a high lineal relation in the number of posts in three of the personality traits. For example, considering the neuroticism dimension, we found a lineal relation between the number of posts recognised as emotionally stable (low score in this dimension) and neurotic posts (high score). This relation means that there is, on average, almost double the number of emotionally stable posts over neurotic posts. Figure 3 shows the relation distribution and its lineal tendency ($R^2 = 0.96$).

We also found a relation in the number of posts in the extraversion and conscensiouness dimensions. Figure 4 shows that the number of extroverted posts is close to double the number of introverted posts, and this could be because extraverts tend to enjoy human interactions and to be enthusiastic and talkative. Conversely introverts tend to be more reserved and less outspoken in groups.

Figure 5 shows the relation in the conscientiousness dimension. We found that there were more than twice as many conscientious posts as unconscientious ones in conversations. The relations found can explain how people contribute to a discussion thread in different ways, depending on their personalities.

Considering the second methodology, in Figure 2 we observe, for example, that from the total number of relations considered, 80 per cent involved extroverted people, while only 19 per cent involved introverted and extroverted people. Regarding agreeableness, of the total number of relations 75 per cent involved agreeable people, while 22 per cent involved people with different values in this dimension. With

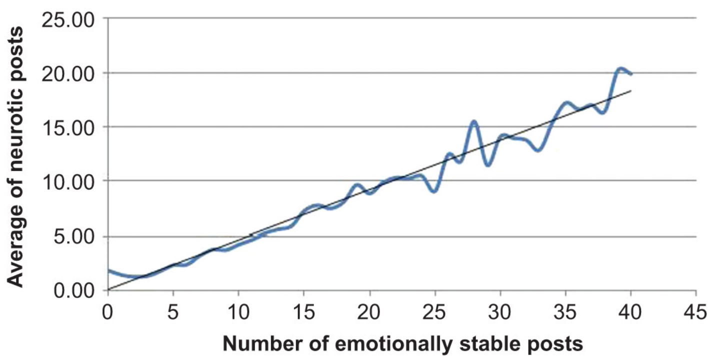

Figure 3.
Relation between number of emotionally stable posts (X axis) and average of neurotic posts (Y axis)

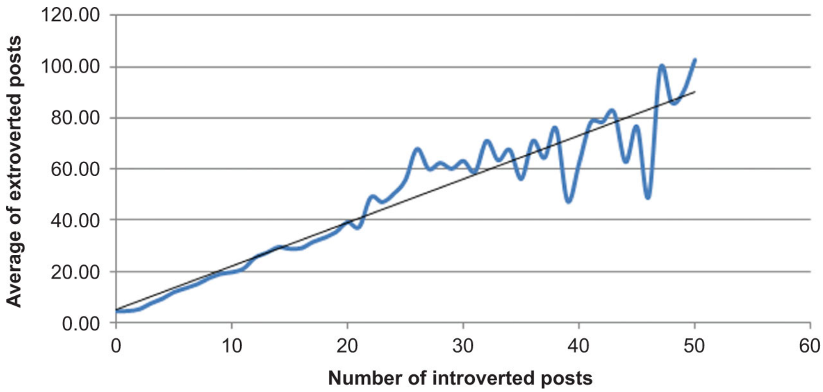

Figure 4.
Relation between number of introverted posts (X axis) and average of extroverted posts (Y axis)

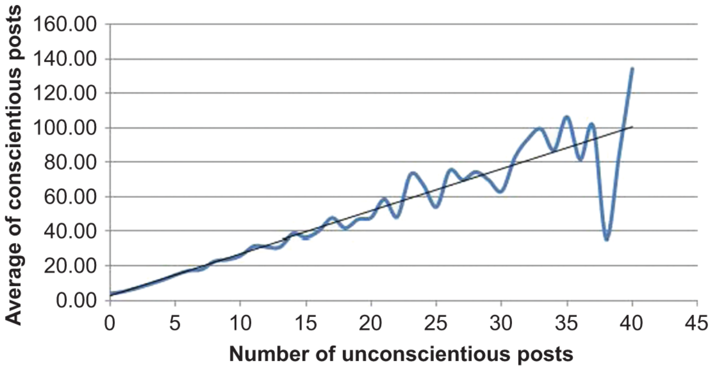

Figure 5.
Relation between the number of unconscientious posts (X axis) and conscientious posts (Y axis)

respect to openness to experience, 96 per cent of the relationships involved people open to experience, while 4 per cent included people with different values in this personality trait. Finally for emotional stability we found that 58 per cent of the relations correspond to stable people, while 36 per cent to stable and neurotic users. No significant values were obtained for the conscientiousness dimension.

With regard to association rules, some of the more interesting ones we found are the following:

R1. open_1 = openn → openn_2 = openn conf: 0.94 sup: 0.96

R1 means that people open to experience tend to communicate with others who are open to experience too. This rule has a support of 96 per cent and a confidence value of 94 per cent.

R2. agree_1 = agree → openn_2 = openn conf: 0.94 sup: 0.84

R2 indicates that agreeable users tend to communicate with users who are open to experience.

R3. extra_1 = extro → openn_2 = openn conf: 0.94 sup: 0.88

This rule suggests that extroverted people tend to communicate with people who are open to experience in social networks.

R4. consc_1 = consc, open_1 = openn → extra_2 = extro conf: 0.74 sup: 0.28

Rule 4 indicates that conscious and open to experience people tend to communicate with extroverted people.

R5. extra_1 = extro → extra_2 = extro conf: 0.74 sup: 0.80

This rule suggests that extroverted people tend to communicate with extroverted people.

R6. agree_1 = agree → extra_2 = extro conf: 0.74 sup: 0.76

Rule 6 suggests that agreeable users tend to communicate with extroverted users.

R7. agree_1 = agree → agree_2 = agree conf: 0.71 sup: 0.75

Rule 7 suggests that agreeable users tend to communicate with agreeable users.

R8. emoti_1 = stable → agree_2 = agree conf: 0.71 sup: 0.67

Rule 8 indicates that emotionally stable people tend to communicate with agreeable people.

R9. extra_1 = extro → agree_2 = agree conf: 0.71 sup: 0.79

Rule 9 indicates that extroverted people tend to communicate with agreeable people.

R10. emoti_1 = stable → extra_2 = extro conf:0.74 sup: 0.68

Rule 10 indicates that emotionally stable people tend to communicate with extroverted people.

Some of these rules (Rules 1, 5 and 7) coincide with the patterns discovered in the analysis reported in Figure 2 and others provide information about how different traits interact. For example from Rule 6 we can conclude that agreeable users tend to communicate with extroverted users. From Rule 8 we can infer that emotionally stable users tend to communicate with agreeable users.

The discovery of association rules among user personality traits not only helps us to understand how users communicate on social network sites, but also has a direct application in recommender systems technology.

Discussion and implications

The main goal of our analyses was discovering relationship patterns between the personalities of users who are linked in social networks. Our findings suggest that there is a tendency for people with a certain personality dimension value to communicate with others who have a very similar value in that personality dimension (homophily). This was verified for users who are open to experience, extrovert, agreeable and, with less precision, for those who are emotionally stable. We also discovered relations between different dimensions of personality.

According to some of the patterns discovered, agreeable people tend to communicate with extroverted people, and emotionally stable users tend to communicate through discussion threads with agreeable people. As far as we are concerned, no similar analysis has been done based on other social networks.

In the literature we can find different works that have used datasets from social networks, such as Facebook or Twitter, for research purposes (Hannon *et al.*, 2010; Kwak *et al.*, 2010). However most of these works only use information about actions made by users, such as logging in, comments, number of readings, among others. Most of these works do not use text for assessing personality. We decided to use a dataset from MySpace because it is formed by information about users, discussion threads and their topics, the posts made by these users in threads, and mainly, it contains full text written by the users. In addition users interact with each other in discussions, allowing us to extract user relationships. Our study and the technique we used could also be easily applied to any other dataset that contains full text of users' posts. The validation of the patterns found in other social media in which users communicate through textual posts remains for future work.

The patterns we have discovered can be used to recommend contacts or friends to users based on their personality or to include new people in a thread to contribute new ideas to the discussion topic. The problem of recommending users in the social web has gained interest in the last few years due to the explosive growth of registered users on social sites, which hinders the user task of finding interesting people to contact. Several approaches have recently been presented for Twitter, Facebook and other social networks (Kazemi and Nematbakhsh, 2011; Armentano *et al.*, 2012; Roth *et al.*, 2010).

In this context personality can play an important role for enriching content-based or topology-based approaches for people recommendation. In particular associations can be directly applied to the problem of determining the confidence in recommending a contact to a certain user. Let us suppose a set of users has been identified as potential candidates using some recommendation strategy. Then personality matching can be applied to help in the ranking of the candidates. If the target user's (the one receiving the recommendations) personality matches the antecedent of one or more association rules, more confidence can be placed in those candidates whose personality matches the consequent of the rules. Thus personality is added as an additional factor to take into account in the recommendation process of suggesting friends.

Finally the values obtained for the openness dimension could validate the theory of Correa *et al.* (2010). Their results revealed that extraversion and openness to experiences were positively related to social media use, whereas emotional stability was negatively related to it. In relation to this finding, we found that most users included in the dataset scored high values in openness to experiences and extraversion dimensions.

Conclusions, limitations and future work

In this work we showed the results of different analyses. First we calculated the personality of each user using text from different threads, in order to check whether the stability personality theory holds in social networks when writing about distinct topics. We obtained very low differences between personality values from different threads for the same user, which could validate this theory.

For our main goal we created a relation between two users when both of them had posted something in the same thread. We discovered interesting patterns for some of the personality dimensions considering all the relationships between traits existing in the dataset.

From the results obtained, we can conclude that personality affects the way in which a person interacts with others in social networks, and that there are relationships between some personality dimensions in this context. As mentioned before, this information can be used to provide recommendations of contacts or potential users that can contribute to a given discussion topic.

There are some limitations in our approach that should be mentioned. We had to determine relations by analysing users posting to a certain thread. In other social networks, such as Facebook, this information is directly extracted from the user profile. However, Facebook datasets do not include users' texts to compute their personalities. Regarding the dataset used, the distribution of personality traits is unbalanced. For example most users were open to experience. This fact could affect the pattern discovery process.

Another limitation of our proposal is its dependence on the precision of the personality recognition software. Currently 74 per cent precision is reported. This result affects the precision of our approach when finding relationship patterns among users' personalities. However an automatic tool for personality detection is crucial in social media since users are not willing to complete long questionnaires. Moreover there is an important volume of textual data available to infer personality.

In addition the personality recognition process could be improved by using other characteristics such as group behaviour, prosodic features or highlighted text. As future work we plan to conduct some more experiments considering these issues.

References

Agrawal, R. and Srikant, R. (1994), "Fast algorithms for mining association rules", *Proceedings of the 20th International Conference on Very Large Data Bases, Morgan Kaufmann, San Francisco, CA*, pp. 487-499.

Aguilar, R.A., De Antonio, A. and Imbert, R. (2007), "Searching Pancho's soul: an intelligent virtual agent for human teams", *Proceedings of the 4th Congress of Electronics, Robotics and Automotive Mechanics*, IEEE, Los Alamitos, CA, pp. 568-571.

Allport, G.W. and Odbert, H.S. (1936), "Trait names: a psycho-lexical study", *Psychological Monographs*, Vol. 47 No. 1, p. 171.

Argamon, S., Dhawle, S., Koppel, M. and Pennebaker, J. (2005), "Lexical predictors of personality type", paper presented at the Joint Annual Meeting of the Interface and the Classification Society of North America, 8-12 June, St Louis, MO.

Armentano, M., Godoy, D. and Amandi, A. (2012), "Topology-based recommendation of users in micro-blogging communities", *Journal of Computer Science and Technology*, Vol. 27 No. 3, pp. 624-634.

Benet-Martínez, V. and John, O.P. (1998), "Los Cinco Grandes across cultures and ethnic groups: multitrait-multimethod analyses of the Big Five in Spanish and English", *Journal of Personality and Social Psychology*, Vol. 75 No. 3, pp. 729-750.

Benevenuto, F., Rodrigues, T., Cha, M. and Almeida, V. (2009), "Characterizing user behavior in online social networks", *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference in Chicago*, ACM, New York, NY, pp. 49-62.

Byrne, D. (1971), *The Attraction Paradigm*, Academic Press, New York, NY.

Cattell, R.B. (1946), *The Description and Measurement of Personality*, World Book, Yonkers, NY.

Clore, G.L. and Byrne, D.A. (1974), "A reinforcement-affect model of attraction", in Huston, T.L. (Ed.), *Foundations of Interpersonal Attraction*, Academic Press, New York, NY, pp. 143-170.

Cobb-Clark, D.A. and Schurer, S. (2011), "The stability of big-five personality traits", *Economics Letters*, Vol. 115 No. 1, pp. 11-15.

Coltheart, M. (1981), "The MRC psycholinguistic database", *Quarterly Journal of Experimental Psychology*, Vol. 33 No. 4, pp. 497-505.

Correa, T., Hinsley, A.W. and Gil de Zùniga, H. (2010), "Who interacts on the web? The intersection of users' personality and social media use", *Computers in Human Behavior*, Vol. 26 No. 2, pp. 247-253.

Costa, P.T. Jr and McCrae, R.R. (1992), *NEO PI-R Professional Manual*, Psychological Assessment Resources, Odessa, FL.

Digman, J.M. (1990), "Personality structure: emergence of the five-factor model", *Annual Review of Psychology*, Vol. 41, pp. 417-440.

Digman, J.M. and Inouye, J. (1986), "Further specification of the five robust factors of personality", *Journal of Personality and Social Psychology*, Vol. 50 No. 1, pp. 116-123.

Dolgova, E.O.W., van Bosch, F.A.J. and Van den Volberda, H.W. (2010), *The Interaction between Personality, Social Network Position and Involvement in the Innovation Process*, Erasmus University, Rotterdam, pp. 1-31, available at: https://faculty.fuqua.duke.edu/oswc/2010/Proposals/Dolgova.pdf (accessed 24 June 2013)

Eysenck, H.J. (1991), "Dimensions of personality: 16, 5 or 3? Criteria for a taxonomic paradigm", *Personality and Individual Differences*, Vol. 12 No. 8, pp. 773-790.

Fast, L.A. and Funder, D.C. (2008), "Personality as manifest in word use: correlations with self-report, acquaintance-report, and behavior", *Journal of Personality and Social Psychology*, Vol. 94 No. 2, pp. 334-346.

Feldt, R., Angelis, L., Torkar, R. and Samuelsson, M. (2010), "Links between the personalities, views and attitudes of software engineers", *Information and Software Technology*, Vol. 52 No. 6, pp. 611-624.

Golbeck, J., Robles, C. and Turner, K. (2011), "Predicting personality with social media", *Proceedings 2011 Annual Conference Extended Abstracts on Human factors in Computing Systems*, ACM, New York, NY, pp. 253-262.

Goldberg, L.R. (1992), "The development of markers for the Big-Five factor structure", *Psychological Assessment*, Vol. 4 No. 1, pp. 26-42.

Hannon, J., Bennett, M. and Smyth, B. (2010), "Recommending Twitter users to follow using content and collaborative filtering approaches", *Proceedings of the 4th ACM Conference on Recommender Systems*, ACM, New York, NY, pp. 199-206.

John, O.P. (1990), "The 'Big Five' factor taxonomy: dimensions of personality in the natural language and in questionnaires", in Pervin, L. (Ed.), *Handbook of Personality Theory and Research*, Guilford, New York, NY, pp. 66-100.

John, O.P. and Srivastava, S. (1999), "The Big Five trait taxonomy: history, measurement, and theoretical perspectives", in Pervin, L.A. and John, O.P. (Eds), *Handbook of Personality: Theory and Research*, Guilford Press, New York, NY, pp. 102-138.

Kazemi, A. and Nematbakhsh, M. (2011), "Finding compatible people on social networking sites, a semantic technology approach", *Proceedings 2nd International Conference on Intelligent Systems, Modelling and Simulation*, IEEE, Alamitos, CA, pp. 306-309.

Klein, K.J., Saltz, J.L. and Mayer, D.M. (2004), "How do they get there? An examination of the antecedents of centrality in team networks", *Academy of Management Journal*, Vol. 47 No. 6, pp. 952-963.

Klementinen, M., Mannila, H., Ronkainen, P., Toivonen, H. and Verkamo, A.I. (1994), "Finding interesting rules from large sets of discovered association rules", *Proceedings of the Third International Conference on Information and Knowledge Management*, ACM, New York, NY, pp. 401-407.

Kwak, H., Lee, C., Park, H. and Moon, S. (2010), "What is Twitter, a social network or a news media?", *Proceedings of the 19th International Conference on World Wide Web*, ACM, New York, NY, pp. 591-600.

McCrae, R. and Costa, P. (1989), "Reinterpreting the Myers-Briggs type indicator from the perspective of the five-factor model of personality", *Journal of Personality*, Vol. 57 No. 1, pp. 17-41.

Mairesse, F., Walker, M., Mehl, M. and Moore, R. (2007), "Using linguistic cues for the automatic recognition of personality in conversation and text", *Journal of Artificial Intelligence Research*, Vol. 30 No. 1, pp. 457-500.

Mehl, M.R., Gosling, S.D. and Pennebaker, J.W. (2006), "Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life", *Journal of Personality and Social Psychology*, Vol. 90 No. 5, pp. 862-877.

Mehra, A., Kilduff, M. and Brass, D.J. (2001), "The social networks of high and low self-monitors: implications for workplace performance", *Administrative Science Quarterly*, Vol. 46 No. 1, pp. 121-146.

Mount, M.K. and Barrick, M.R. (1998), "Five reasons why the 'big five' article has been frequently cited", *Personnel Psychology*, Vol. 51 No. 4, pp. 849-857.

Noller, P., Law, H. and Comrey, A.L. (1987), "Cattell, Comrey and Eysenck personality factors compared: more evidence for the five robust factors?", *Journal of Personality and Social Psychology*, Vol. 53 No. 4, pp. 775-782.

Norman, W.T. (1963), "Toward an adequate taxonomy of personality attributes: replicated factor structure in peer nomination personality ratings", *Journal of Abnormal and Social Psychology*, Vol. 66 No. 6, pp. 574-583.

Oberlander, J. and Gill, A.J. (2006), "Language with character: a stratified corpus comparison of individual differences in e-mail communication", *Discourse Processes*, Vol. 42 No. 3, pp. 239-270.

Oberlander, J. and Nowson, S. (2006), "Whose thumb is it anyway?", *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics*, ACL, Stroudsburg, PA, pp. 627-634.

Ozer, D.J. and Benet-Martínez, V. (2006), "Personality and the prediction of consequential outcomes", *Annual Review of Psychology*, Vol. 57, pp. 401-421.

Paunonen, S.V. and Jackson, D.N. (2000), "What is beyond the Big Five? Plenty!", *Journal of Personality*, Vol. 68 No. 5, pp. 821-836.

Pennebaker, J.W. and King, L.A. (1999), "Linguistic styles: language use as an individual difference", *Journal of Personality and Social Psychology*, Vol. 77 No. 6, pp. 1296-1312.

Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J., Stillwell, D. and Crowcroft, J. (2011), "Our Twitter profiles, ourselves: predicting personality with Twitter", *Proceedings of the IEEE Third International Conference on Privacy, Security, Risk and Trust and IEEE Third International Conference on Social Computing*, IEEE, Los Alamitos, CA, pp. 180-185.

Roth, M., Ben-David, A., Deutscher, D., Flysher, G., Horn, I., Leichtberg, A., Leiser, N., Matias, Y. and Merom, R. (2010), "Suggesting friends using the implicit social graph", *Proceedings 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, ACM, New York, NY, pp. 233-242.

Saucier, G. (1994), "Mini-markers: a brief version of Goldberg's unipolar Big-Five markers", *Journal of Personality Assessment*, Vol. 63 No. 3, pp. 506-516.

Scherer, K.R. (1979), "Personality markers in speech", in Scherer, G.R. and Giles, H. (Eds), *Social Markers in Speech*, Cambridge University Press, Cambridge, pp. 147-209.

Schrammel, J., Köffel, C. and Tscheligi, M. (2009), "Personality traits, usage patterns and information disclosure in online communities", *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*, British Computer Society, Cambridge, pp. 169-174.

Shah, D., Lakshmanan, L., Ramamritham, K. and Sudarshan, S. (1999), "Interestingness and pruning of mined patterns", paper presented at 1999 ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, 30 May, Philadelphia, PA.

Steel, D., Rinne, T. and Fairweather, J. (2012), "Personality, nations, and innovation: relationships between personality traits and national innovation scores", *Cross-Cultural Research: The Journal of Comparative Social Science*, Vol. 46 No. 1, pp. 3-30.

Tupes, E.C. and Christal, R.E. (1961), "Recurrent personality factors based on trait ratings", *USAF ASD Technical Report*, US Air Force, Lackland Air Force Base, San Antonio, TX, pp. 61-97.

Uesugi, S. (2011), "Effects of personality traits on usage of social networking service", *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*, ACM, New York, NY, pp. 629-634.

Waller, N.G. and Ben-Porath, Y.S. (1987), "Is it time for clinical psychology to embrace the five-factor model of personality?", *American Psychologist*, Vol. 42 No. 9, pp. 887-889.

Watson, D. and Clark, L.A. (1992), "On traits and temperament: general and specific factors of emotional experience and their relation to the five factor model", *Journal of Personality*, Vol. 60 No. 2, pp. 441-476.

Zuckerman, M., Bernieri, F., Koestner, R. and Rosenthal, R. (1989), "To predict some of the people some of the time: in search of moderators", *Journal of Personality and Social Psychology*, Vol. 57 No. 2, pp. 279-293.

About the authors

Jose Maria Balmaceda is a PhD Candidate and Teaching Assistant at the Universidad Nacional del Centro de la Provincia de Buenos Aires (UNCPBA), Tandil, Argentina. He holds a PhD scholarship awarded by the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). His research interests are computer-supported collaborative work, team roles, and artificial intelligence techniques for user profiling.

Silvia Schiaffino is a Researcher at CONICET and a Member of the Instituto de Sistemas Tandil. She is also a full-time professor in the Department of Computer Science at UNCPBA. She obtained her master's degree in systems engineering (2001) and her PhD in computer science (2004) at the same university. Her main research interests are personalisation and adaptation, recommender systems and artificial intelligence techniques for user profiling. Professor Silvia Schiaffino is the corresponding author and can be contacted at: silvia.schiaffino@isistan.unicen.edu.ar

Daniela Godoy is a Researcher at CONICET and a Member of the Instituto de Sistemas Tandil, as well as a full-time professor in the Department of Computer Science at UNCPBA. She obtained her master's degree in systems engineering (2001) and her PhD in computer science (2005) at the same university. Her research interests include intelligent agents, user profiling and text mining.

To purchase reprints of this article please e-mail: **reprints@emeraldinsight.com**
Or visit our web site for further details: **www.emeraldinsight.com/reprints**