How do personality traits relate on on-line social networks?

Structured Abstract:

Purpose: This work aims at analyzing the relationships between the personality traits of users who are related in online social networks. First, we tried to discover relation patterns between personality dimensions in conversations. We 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. We used the Five Factor Model of Personality or Big Five, which has been widely studied in the Psychology area.

Design/Methodology/Approach: One of the approaches to detect users’ personalities is by analyzing the language they use when they talk to others. Based on this assumption, we computed users’ personality from the conversations extracted from MySpace social network. Then, we analyzed the relationships among personality traits of related users to discover patterns.

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

Originality/value: The results we found provide some clues about how people relate within online social networks, particularly who they tend to relate to depending on their personality. The discovered patterns can be used in 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 relate in social media considering their personality.
1 Introduction

Personality is a subject that has been studied for decades by many researchers. It involves the particular combination of emotional, attitudinal, and behavioral 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 (Tupes and Christal, 1961) was the first that identified five recurrent factors in the analysis of personality; then Norman (Norman, 1963) replicated that work. Although the five factors remained hidden throughout the 1960s and 1970s, in the 1980s, however, researchers concluded that they are fundamental dimensions of personality, found in self-reports and ratings, in natural languages and theoretically based questionnaires (John, 1990).

There are many other 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/organized 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 a consensus in the Psychology area 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, 1987; Digman, 1990; Goldberg, 1992; Mount and Barrick, 1998; Aguilar, 2007; Feldt et al, 2010). Moreover, the FFM has been used in various domains. For example, it has been used by organizations 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). Also, personality affects the way in which we engage in social network. In this regard, some recent works have studied personality in social environments (Dolgova et al, 2010; Mehra et al, 2001; Özer and Benet-Martinez, 2006; Schrammel et al, 2009; Uesugui, 2011). However, the mentioned works are not oriented
to determining how users relate in social media according to their personality as ours, but how personality impacts on other aspects, such as social media adoption. To the best of our knowledge no previous works have analyzed 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 related within a social network, and analyzing how personality affects the way in which each individual communicates with others in such context. Particularly, we suggest that the similarity-attraction paradigm or homophily (Byrne, 1971; Clore and Byrne, 1974), which predicts that people tend to build up relationships with similar others, could be also verified in social networks in terms of personality.

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

Finally, we consider that characterizing users' behavior 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 relate to others can give some hints about which types of contacts a user would select, and thus, recommend him/her contacts or friends having a certain personality. With this aim, we tried to discover relationship patterns among related 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-Martinez and John, 1998; Costa and Mc Crae, 1992; Goldberg, 1992; John and Srivastava, 1999). Another approach is by analyzing the language people use to communicate with others (Mairesse and Walker, 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. Particularly, we based our study on a recent work that automatically detects users' personality 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 article is organized as follows. In Section 2 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 recognize users’ personality and how it works. In Section 3 we present our findings. In Section 4 we discuss the relationships we have found and their implications for the design and development of social software. Finally, in Section 5 we expose our conclusions, limitations of our study and future work.

2 Background and Related Works

In this section we introduce some key concepts in the study of personality (subsection 3.1), how it can be detected by analyzing interactions among individuals (subsection 3.2), and we discuss some researches that analyze the relationship between personality and social media (subsection 3.3).

2.1 Personality: Main concepts

Personality is the particular combination of emotional, attitudinal, and behavioral 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 FFM (Five Factor Model) or 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) summarizes 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 known as Big Five are the following (Norman, 1963):
• Extraversion vs. Introversion (sociable, assertive, playful vs. aloof, reserved, shy)

• Emotional stability vs. Neuroticism (calm, unemotional vs. insecure, anxious)

• Agreeableness vs. Disagreeableness (friendly, cooperative vs. antagonistic, fault finding)

• Conscientiousness vs. Unconscientiousness (self-disciplined, organized vs. inefficient, careless)

• 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 240-item NEO Personality Inventory (Costa and McCrae, 1992), Revised (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 (BFI) (Benet-Martinez and John, 1998; John and Srivastava, 1999), the 60-item NEO Five-Factor Inventory (NEO-FFI) (Costa and McCrae, 1992), and Goldberg’s instrument that comprised 100 trait descriptive adjectives (TDA) (Goldberg, 1992). In (John and Srivastava, 1999) the authors estimated that the BFI, NEO-FFI, and TDA take approximately 5, 15, and 15 minutes to complete, respectively. Recognizing the need for an even briefer measure of the Big Five, Saucier (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 analyzes 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 in (Allport and Odbert, 1936) abstracted terms from a dictionary, while Cattell (Cattell, 1946) formed them into synonym clusters and then created rating scales contrasting groups of adjectives. Then, in (Tuples and Christal, 1961) the authors 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 in (Mairesse et al, 2007), as described below.

2.2 Automatic detection of personality

Thus far, there has been little work on the automatic recognition of personality traits (Argamon et al., 2005; Oberlander & Nowson, 2006, Mairesse et al, 2007). Argamon’s work (Argamon et al, 2005) was one of the first published 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 Optimization (SMO). 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%: an 8% absolute improvement over their baseline. Oberlander and Nowson’s work (Oberlander and Nowson, 2006) follows Argamon’s approach, but they improved the performance and they included also the Agreeableness and Conscientiousness dimensions. Mairesse et al (Mairesse et al, 2007) made this classification using regression and ranking modeling techniques and language cues, and their results were the first to demonstrate statistically significant results for texts and to recognize personality from conversations.

Recently, some works have addressed the detection of personality from social networks. For example, in (Golbeck et al, 2011) the authors demonstrate that a user’s personality can be accurately predicted through the publicly available information on their Facebook profile. In (Quercia et al, 2011), the authors analyze 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’ utterances so that the conclusions can be applied to multiple social services. In this work we used a software application to detect users’ personality, named Personality Recognizer, based on the models proposed in (Mairesse et al, 2007). The Personality Recognizer is a Java 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 from whom it is needed to know his/her personality. 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, 2007; Mehl et al, 2006; Pennebaker and King, 1999), and n-grams (Oberlander and Gill, 2006). In (Pennebaker and King, 1999) the authors identified many linguistic features associated with each of the Big Five personality traits. They used their Linguistic Inquiry and Word Count (LIWC) 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. In (Oberlander and Gill, 2006) the authors studied correlates of emotional stability: they found that neurotics use more concrete and frequent words. In (Coltheart, 1981) the author deploys the MRC psycholinguistic database, which contains statistics for over 150,000 words, such as frequency of use, and familiarity.

The Personality Recognizer software use models trained from experiments made in the works mentioned before and also extra tests described in (Mairesse and Walker, 2007). Based on this training the recognizer 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 sentence; few articles; few negations. There are also other levels analyzed, such as lexicon or speech, with cues for each personality trait. But, not all of these features are helpful to recognize all personality types. The authors made a feature selection process in order

1 http://people.csail.mit.edu/francois/research/personality/recognizer.html

2 http://www.java.com/
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% of precision to classify Extraversion and Emotional stability, respectively; and 65% 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.

2.3 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 the performance on team work in a variety of domains. Dolgova et al. (Dolgova et al, 2006) 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’ personality. 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 process. On the other hand, Mehra et al. (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 propose to use self-monitoring orientation (one of the many personality variables) to predict the individual’s position in the social network and how this affects the performance when working in groups. Uesugi (Uesugi, 2011) studied the relation of personality traits in the FFM and SNS usage as well as attitudes towards protecting privacy.

There are also other researches that have indicated that the structural position of an individual in social networks is in part shaped by his/her personality (Dolgova et al, 2006), and how this affects the relation with other group members. Klein et al. (Klein et al, 2004) hypothesized 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 et al. (Ozer et al, 2006) analyze people's personality as the cause of consequential outcomes at individual, interpersonal and social/institutional level. The authors discovered a relation between personality traits and consequential outcomes at these levels. Correa et al. (Correa et al. 2010) studied personality traits as important factors for the engagement of users with social media. The authors investigated the relationship between personality and social media use, mainly focusing on whether people use social media or not. In their study, the authors found, for example, that extraversion is positively related to social media use, whereas emotional stability is negatively related.

Differently to these previous works, we use personality traits as a factor that could affect how people communicate in social networks and whom they tend to relate with.

3 Analyses and findings

In this section we first describe the main goals of our study (subsection 3.1), then the dataset used (subsection 3.2), different methodologies for automatic personality detection (subsection 3.3) and finally, different approaches for pattern discovery (subsection 3.4).

3.1 Goals

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

For our study, we used a dataset created with data from MySpace³, which is a popular social networking site. MySpace offers to its registered users, among other things, the possibility to participate in discussion forums about several predefined topics. In this way, everyone can start a new thread in these forums with a question, and participate freely in a thread created by another user expressing his/her opinion or knowledge. Thus, there is a relation between the person who started a thread and the people who participated in it, given their shared interest.

The threads that form this dataset have been chosen from three different forum topics: Campus Life, News & Politics and Movies. The dataset has been automatically crawled by Fundación Barcelona Media⁴ and it contains about 380,000 comments to 16346 threads dealing with one of the topics mentioned before.

We created a database table that contains information about how many posts each user made in all the threads (33407 users in total). More than 18000 users have only posted once and almost 5000 users have made two posts. However, for our analysis we only used those users who have made more than 2 posts (10117 users) because we needed enough text to obtain good results in the personality recognition step. We also found users that posted only spam, publicity, and text with strange symbols or too short text messages, which caused Personality Recognizer to return an invalid value. We excluded such users, keeping a total of 5002 users.

3.3 Detection of users’ personality

3.3.1 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 Recognizer 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.

³ http://www.myspace.com/

⁴ http://caw2.barcelonamedia.org/node/7
calculated from the user who wrote the text. For example, considering the extroversion dimension, a value of 0 indicates that person has a tendency to being introvert, whereas a value of 7 indicates that a person has a tendency to being extrovert.

Since some users have written strange symbols or words in other languages, and some of them have also introduced 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’ personality separated in five graphics corresponding to the big five dimensions. Table 1 summarizes the average and deviation values for these distributions.

### 3.3.2 Methodology 2: Considering initial posts

Taking into account that a thread could only be started with an initial post that talks about some topic or making some question, we decided to do the same analysis than in Section 3.3.1 but using only these starting texts. We did this analysis because one of the identified language cues for introvert/extrovert at conversational behavior is whether a person listens to the 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 2 shows a summary of the personalities obtained considering only initial texts.
3.3.3 Methodology 3: Considering different threads

Considering that it is demonstrated that personality tends to be stable across time and over different situations (Cobb-Clark and Schurer, 2011), we compared Personality Recognizer’s 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) and deviation (STD) of the personalities of each user, and then we calculated the average and deviation of all STDs. 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 3 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.

3.3.4 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 1, we can observe that almost all people in this dataset are open to experiences. As regards extraversion, we can see that there are also more extroverted people than introverted ones, but the corresponding distribution is centered between values 4 and 5. On the other hand, the emotional stability dimension is the only one having more people with low values than people with high ones.

Comparing Table 2 with Table 1, we can see that the average of extraversion, agreeableness and consciousness have increased a little, but the standard deviation also climbed up. With this information we cannot affirm that people who start a conversation have in average high value of extraversion.

Table 2. Average and deviation of personality's dimensions distribution calculated from initial texts.

Finally, analyzing Table 3 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.

Table 3. Average and deviation of: (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.

3.4 Discovering relationships between related 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 some 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 analyzed correlations between the numbers of posts corresponding to different personality traits. Then, in order to verify that people tend to relate with personality-similar ones (homophily), and also to discover relationships between users’ personalities, we performed two different analyses. First, we analyzed all the relations included in the dataset used, looking for who people tend to make a relation with depending on their personality. The second type of analysis we made consisted in mining association rules.

### 3.4.1 Methodology 1: Analyzing relationships between posts

In order to discover behavioral patterns inside a conversation, we made an analysis that relates 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 2 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 relation. For example, given that each thread has a number of introverted posts and a number of extroverted posts, we tried to find significant 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.

### 3.4.2 Methodology 2: Quantitative analysis

In this analysis, each relation is composed by the personality of a user and the personality of a second one. 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 of relations. We supposed that if the similarity-attraction paradigm is true in social networks, we would observe many relations between people with similar personality. Table 4 shows the results obtained for each dimension separately.
3.4.3 **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 \( X \subseteq I \) and \( Y \subseteq I \) and \( X \cap Y = \emptyset \). \( X \) is the antecedent of the rule and \( Y \) is the consequent. The rule has support \( s \) in \( D \) if \( s\% \) of the transactions in \( D \) contain \( X \cup Y \). The rule \( X \rightarrow Y \) holds in \( D \) with confidence \( c \) if \( c\% \) 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 \( \text{minsup} \)) and minimum confidence (called \( \text{minconf} \)).

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\(^5\) tool and the Apriori algorithm to discover association rules using a value of \( \text{minconf}=0.7 \) (70\%) and a value of \( \text{minsup}=0.1 \) (10\%).

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 use a subset of the pruning rules proposed in (Shah et al, 1999). Basically, these pruning rules state that given the rules \( A \rightarrow B \rightarrow C \) 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

---

\(^5\) http://www.knime.org
consequent is more specific. Thus, the redundant rule can be deleted provided that both rules have similar confidence values.

### 3.4.4 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 recognized as emotional stable (low score in this dimension) and neurotic posts (high score). This relation means that there is, in average, almost double number of emotionally stable posts over neurotic posts. Figure 2 shows the relation distribution and its lineal tendency ($R^2=0.96$).

![Fig. 2. Relation between number of emotionally stable posts (X axis) and average of neurotic posts (Y axis).](image)

We also found a relation in the number of posts in the extraversion and conscientiousness dimensions. Figure 3 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. On the other hand, introverts tend to be more reserved and less outspoken in groups.

Figure 4 shows the relation in the conscientiousness dimension. We found that more than double of conscientious posts over unconscientious ones are present in a conversation. The relations found can explain how
people contribute to a discussion thread in different ways, depending on their personalities.

![Graph showing the relationship between introverted posts and extroverted posts.](image1)

**Fig. 3.** Relation between number of introverted posts (X axis) and average of extroverted posts (Y axis).

![Graph showing the relationship between non-conscientious posts and conscientious posts.](image2)

**Fig. 4.** Relation between the number of non-conscientious posts (X axis) and conscientious posts (Y axis).

Considering the second methodology, in Table 4 we observe, for example, that from the total number of relations considered, 80% of these relations involved extroverted people, while only 19% involved introverted and extroverted people. As regards agreeableness, from the total number of relations 75% involved agreeable people, while 22% involved people with different values in this dimension. With respect to
openness to experience, 96% of the relationships involved people open to experience, while 4% included people with different values in this personality trait. Finally, for emotional stability we found that 58% of the relations correspond to stable people, while 36% to stable and neurotic users. No significant values were obtained for the conscientiousness dimension.

As regards association rules, some of the more interesting ones we found are the following:

R1. open_1=openn $\rightarrow$ openn_2=openn conf: 0.94 sup: 0.96
R1 means that people open to experience tend to relate to others who are open to experience too. This rule has a support of 96% and a confidence value of 94%.

R2. agree_1=agree $\rightarrow$ openn_2=openn conf: 0.94 sup: 0.84
R2 indicates that agreeable users tend to relate to users who are open to experience.

R3. extra_1=extro $\rightarrow$ openn_2=openn conf: 0.94 sup: 0.88
This rule suggests that extroverted people tend to relate to people who are open to experience in social networks.

R4. consc_1=consc, open_1=openn $\rightarrow$ extra_2=extro conf: 0.74 sup: 0.28
Rule 4 indicates that conscious and opened to experience people tend to relate to extroverted people.

R5. extra_1=extro $\rightarrow$ extra_2=extro conf: 0.74 sup: 0.80
This rule suggests that extroverted people tend to relate to extroverted people.

R6. agree_1=agree $\rightarrow$ extra_2=extro conf: 0.74 sup: 0.76
Rule 6 suggests that agreeable users tend to relate to extroverted users.

R7. agree_1=agree $\rightarrow$ agree_2=agree conf: 0.71 sup: 0.75
Rule 7 suggests that agreeable users tend to relate to agreeable users.
Rule 8 indicates that emotionally stable people tend to relate to agreeable people.

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

Rule 9 indicates that extroverted people tend to relate to agreeable people.

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

Rule 10 indicates that emotionally stable people tend to relate to extroverted people.

Some of these rules (rules 1, 5 and 7) coincide with the patterns discovered in the analysis reported in Table 4 and others provide information about how different traits relate. For example, from rule 6 we can conclude that agreeable users tend to relate with extroverted users. From rule 8 we can infer that emotionally stable users tend to relate to agreeable users.

The discovery of association rules among user personality traits not only helps us to understand how users relate in social sites, but also has a direct application in recommender systems technology.
4 Discussion and Implications

The main goal of our analyses was discovering relationship patterns between the personalities of users who are related in social networks. Our findings suggest that there is a tendency for people with some personality dimension value to make a relation with other ones with a very similar value in that personality dimension (homophily). This could be verified for users who are open to experience, 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 relate to extroverted people, and emotionally stable users tend to relate 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 some social network, such as Facebook or Twitter, for research purposes (Hannon, et al., 2010; Kwak et al., 2010). However, most of these works use only 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 some thread, 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 user’s posts. The validation of the patterns found in other social media in which users relate through textual posts remains as 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 years due to the explosive growth of registered users in social sites, which hinders the user task of finding interesting people to contact with.
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 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 such candidates. If the target user (the one receiving the recommendations) personality matches the antecedent of one or more association rules, more confidence can be given to those candidates whose personality matches the consequent of 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 Correa et al.’s theory (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 relate to it. In relation to this finding, we found that most users included in the dataset got high values in openness to experiences and extraversion dimensions.

5 Conclusions, Limitations and Future Work

In this work we showed the results of different analyses. Firstly, we calculated the personality of each user using text from different threads, in order to check if 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, and this fact could validate this theory.

For our main goal, we created a relation between two users when both of them have 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 another one 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 analyzing 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. As regards 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, a 74% of precision is reported. This result affects the precision of our approach when finding relationship patterns among users’ personality. However, an automatic tool for personality detection is crucial in social media since users are not willing to complete long questionnaires. Also, 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 like group behavior, prosodic features or highlighted text. As a future work, we plan to make some more experiments considering these issues.

6 References


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


### Table 1. Average and deviation of personality's dimensions distribution.

<table>
<thead>
<tr>
<th></th>
<th>agree</th>
<th>consc</th>
<th>emoti</th>
<th>extra</th>
<th>openn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>4.67</td>
<td>4.14</td>
<td>3.11</td>
<td>4.19</td>
<td>5.61</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.06</td>
<td>1.05</td>
<td>1.10</td>
<td>1.42</td>
<td>1.08</td>
</tr>
</tbody>
</table>

### Table 2. Average and deviation of personality's dimensions distribution calculated from initial texts.

<table>
<thead>
<tr>
<th></th>
<th>agree</th>
<th>consc</th>
<th>emoti</th>
<th>extra</th>
<th>openn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>4.83</td>
<td>4.31</td>
<td>3.11</td>
<td>4.26</td>
<td>5.46</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.08</td>
<td>1.11</td>
<td>1.16</td>
<td>1.61</td>
<td>1.11</td>
</tr>
</tbody>
</table>

### Table 3. Average and deviation of (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.

<table>
<thead>
<tr>
<th></th>
<th>Agree</th>
<th>consc</th>
<th>emoti</th>
<th>extra</th>
<th>Open</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average (A)</td>
<td>0.67</td>
<td>0.69</td>
<td>0.73</td>
<td>0.74</td>
<td>0.58</td>
</tr>
<tr>
<td>Standard Deviation (A)</td>
<td>0.44</td>
<td>0.44</td>
<td>0.45</td>
<td>0.43</td>
<td>0.38</td>
</tr>
<tr>
<td>Average (B)</td>
<td>0.63</td>
<td>0.66</td>
<td>0.70</td>
<td>0.70</td>
<td>0.54</td>
</tr>
<tr>
<td>Standard Deviation (B)</td>
<td>0.85</td>
<td>0.88</td>
<td>0.92</td>
<td>0.89</td>
<td>0.73</td>
</tr>
</tbody>
</table>
### Table 4. Analysis of personalities in users’ relations.

<table>
<thead>
<tr>
<th></th>
<th>Extro.</th>
<th>Intro.</th>
<th>Neuro.</th>
<th>Stable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extro.</td>
<td>80%</td>
<td>10%</td>
<td>6%</td>
<td>17%</td>
</tr>
<tr>
<td>Intro.</td>
<td>9%</td>
<td>1%</td>
<td>19%</td>
<td>58%</td>
</tr>
</tbody>
</table>

(A) Extraversion

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Agree.</td>
<td>75%</td>
<td>10%</td>
<td>42%</td>
<td>22%</td>
</tr>
<tr>
<td>Disagree.</td>
<td>12%</td>
<td>2%</td>
<td>20%</td>
<td>16%</td>
</tr>
</tbody>
</table>

(B) Emotional Stability

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Openne.</td>
<td>96%</td>
<td>2%</td>
<td>96%</td>
<td>2%</td>
</tr>
<tr>
<td>NotOpenne.</td>
<td>2%</td>
<td>0%</td>
<td>2%</td>
<td>0%</td>
</tr>
</tbody>
</table>

(C) Agreeableness

(D) Conscientiousness

(E) Openness to experience