Exploring the role of personality traits in followee recommendation

Antonela Tommasel, Alejandro Corbellini, Daniela Godoy and
Silvia Schiaffino

ISISTAN Research Institute, UNICEN – CONICET, Tandil, Argentina

Abstract

Purpose – Followee recommendation is a problem rapidly gaining importance in Twitter as well as in other micro-blogging communities. To find interesting users to follow, most recommendation systems leverage different factors such as graph topology or user-generated content, among others. Those systems mostly disregard, however, the effect of psychological characteristics, such as personality, over the followee selection process. As personality is considered one of the primary factors that influence human behaviour, the purpose of this paper is to shed some light on the impact of personality traits on followee selection.

Design/methodology/approach – The authors performed a data analysis comparing the similarity among Twitter users and their followees regarding personality traits. The authors analysed three different similarity measures. First, the authors computed an overall similarity considering the five personality traits or dimensions of the Five-Factor model as a whole. Second, the authors computed the dimension-to-dimension similarity considering each individual personality trait independently of each other. Third, the authors computed a cross-dimension similarity considering each personality dimension in relation to the others.

Findings – This study showed that personality should be considered as a distinctive factor in the process of followee selection. However, personality dimensions should not be analysed as a whole as the overall personality similarity might not accurately assess the actual matching between individuals. Instead, the performed data analysis showed the existence of relations among the individual dimensions. Thus, the importance of considering each personality trait with respect to others is stated.

Originality/value – This study is among the firsts to study the impact of personality, one of the primary factors that influence human behaviour and social relationships, in the selection of followees in micro-blogging communities.

Keywords Personality traits, Twitter, Followee recommendation, Human aspects of recommendation

Paper type Research paper

1. Introduction

Recommendation systems are present in a wide range of applications that expose users to huge collections of items or services (Herlocker *et al.*, 2004). Such systems aim at reducing the information overload and assisting users in efficiently finding interesting items by giving them personalised advice. Usually, recommendation systems present lists of suggested or interesting items to users based on personal interests or on the opinion of like-minded users. In the context of social networks, items can be people a user might be interested in following. Followee recommendation is a problem rapidly gaining importance in micro-blogging communities due to the extremely large and exponentially increasing volume of online activity and active users. Several approaches have been proposed to suggest users worth following based on different criteria (Armentano *et al.*, 2011; Wu *et al.*, 2012; Chen *et al.*, 2014).

The decision to start following another user in social networks might not depend on a unique factor or characteristic, i.e. users might have several reasons for choosing whom to follow. For example, a user might follow users because they share mutual friends, they tweet interesting information, they share the same interests than him/her, or they are celebrities or popular individuals in the micro-blogging community. Understanding the reasons for choosing followees becomes essential in the design of personalised recommendation strategies to suggest users worth following.

Interestingly, most of the existing recommendation systems rely on the content of posts/tweets and on the social network topology, without considering how users' preferences and decisions are affected by psychological characteristics, such as personality. Personality is considered one of the primary factors that influence human behaviour and social relationships, as it affects how people react and interact with others. Due to this reason, this study aims at shedding some light on the impact of personality traits on the accurate prediction of followees. Particularly, it focuses on analysing the impact of personality on followee selection in Twitter, an information-oriented social network that does not encourage the formation of high-quality social ties.

This work hypothesises that there is a relation between users' personality and social interactions in the context of social networks, as it was stated for face-to-face relationships in (Selfhout *et al.*, 2010; Cuperman and Ickes, 2009). In turn, several sub-hypotheses can be derived from this general hypothesis. The first one states that social interactions regarding personality are guided by the principle of homophily (McPherson *et al.*, 2001), which in our context indicates that a user would follow users having the same personality than him/her. In this regard, the study aims at verifying whether an overall assessment of personality is sufficient for accurately assessing the relationships between users. The second sub-hypothesis states the existence of several relationship patterns depending on the personality of the involved users. Particularly, it aims at verifying some of the relationship patterns described by Selfhout *et al.* (2010) and Cuperman and Ickes (2009), but in the context of micro-blogging platforms, such as Twitter. For example, we analysed whether extraverted users tend to initiate more friend relations, or in this case, followee relations. Finally, as Selfhout *et al.* (2010) and Cuperman and Ickes (2009) only analysed the effect of the separate dimensions, it is unknown whether the combination of personality dimensions has an effect on social interactions. In this regard, the third sub-hypothesis aims at verifying whether there is a tendency of users exhibiting certain characteristics regarding one specific dimension, to relate with other with certain characteristics regarding the remaining personality dimensions.

The rest of this paper is organised as follows. Section 2 presents background concepts regarding personality and its influence over social relationships. Section 3 describes the Twitter dataset used for experimentation and the research method used. Section 4 presents the obtained results. Then, in Section 5 we discuss the findings of our study and some practical implications. Finally, Section 6 summarises the conclusions drawn from this study.

2. Literature review
Recommendation systems aim at assisting users in efficiently and effectively finding interesting items or other users in social networks by giving them personalised recommendations. However, most social recommendation systems disregard the effect of psychological characteristics over the followee selection process, relying almost exclusively on content and topological factors. In this section, background concepts regarding personality and its influence over social relationships are introduced.

2.1 Personality

Psychology theories define personality as the combination of emotional, attitudinal and interpersonal processes that internally originate in each person, and temperamental and behavioural response patterns (Funder, 2012; Adali and Golbeck, 2012; Heinström, 2003). As a result, personality is one of the primary factors that influence human behaviour as it can moderate how people react, behave and interact with other individuals. Several authors (Costa Jr and McCrae, 1994, 1997; McCrae and Costa, 1982; Moss and Susman, 1980) have agreed that personality remains stable during adulthood exhibiting considerable continuity and stability over time. Consequently, a single assessment can be sufficient to infer or predict individuals' personality in the short to medium term. Social environments, such as micro-blogging sites, can encourage the manifestation of personality as they satisfy the basic psychological needs, including relatedness to other individuals, competence and autonomy (Sherman *et al.*, 2012). Additionally, there is a connection between personality and the tastes and interests of individuals regarding, for example, social behaviour and affective experience (Cuperman and Ickes, 2009), which could imply that people with similar personality have similar interests.

Several works aimed at finding a set of features or characteristics to describe personality. Tupes and Christal (1961, 1992) were the first authors that identified five recurrent features in personality. Subsequent works (Noller *et al.*, 1987; McCrae and Costa Jr, 1989) confirmed those findings and offered evidence of the existence of such features. The Five-Factor or Big Five model (Costa Jr and McCrae, 1992) is a hierarchical model that defines personality as a composition of five traits or dimensions. Although its theoretical foundations have been questioned (Waller and Ben-Porath, 1987; Block, 1995), the model is acknowledged to define some of the essential aspects of personality. The Big Five divides personality into five dimensions: agreeableness, extraversion, openness to experience, conscientiousness and neuroticism. Agreeableness refers to being sympathetic, cooperative and helpful towards others. Individuals with high scores in this dimension tend to be optimistic and to easily trust other people. Extraversion refers to being outgoing, friendly, assertive and energetic. Individuals with high scores in this dimension tend to display high degrees of sociability and talkativeness. Openness to experience refers to being curious, intelligent and imaginative. Individuals with high scores in this dimension tend to have a strong intellectual curiosity, a preference for novelty and variety, and an artistic and sophisticated taste. Conscientiousness refers to being organised, persevering, disciplined, achievement-oriented and responsible. Individuals with high scores in this dimension tend to be extremely reliable, high achievers, hard workers and planners. Finally, neuroticism refers to being anxious, insecure, moody and sensitive. This dimension assesses the degree of emotional stability, anxiety and impulse control.

Generally, to accurately assess personality, individuals are required to explicitly answer a personality questionnaire. However, explicit tests have several drawbacks. First, as individuals are self-reporting their personality, they reflect their own view of themselves, and not necessarily their actual behaviour (Selfhout *et al.*, 2009). Second, the tests are impractical to perform personality analysis in the context of social domains for online recommendation algorithms.

As the outer behaviour represents a manifestation of personality, several works (Bai *et al.*, 2012; Mairesse *et al.*, 2007; Adali and Golbeck, 2012; Golbeck *et al.*, 2011) cast the problem of determining personality as a classification or regression problem over directly observable information, such as text, conversations or conversational

transcripts. Some works (Gottschalk and Gleser, 1969; Rosenberg and Tucker, 1979) have provided evidence suggesting that people's mental states and personality can be predicted by the words they use. Pennebaker *et al.* (2003, 2007) designed a text analysis software, known as Linguistic Inquiry and Word Count (LIWC)[1], which analyses the usage of different categories of words across a wide array of texts. Particularly, it analyses the usage of positive or negative emotions, self-references and causal words, among over 70 language features. Subsequent works used the LIWC features to analyse the correlation between linguistic markers and user personality. In Mairesse *et al.* (2007) the authors aimed at recognising the Big Five personality dimensions from text and conversational features. The authors claimed that there is a correlation between linguistic markers and user personality. For example, extraverted users tend to use social and emotional words, first person pronouns and present tense verbs, whereas introverted users tend to use a richer and more precise vocabulary. The approach considered the LIWC features, and the 14 Medical Research Council (MRC) psycholinguistic features (Coltheart, 1981), both including syntactic and semantic information. Experimental evaluation showed that the best prediction results were obtained for the openness to experience dimension, whereas the worst results were obtained for the extraversion and conscientiousness dimensions. Prediction accuracy ranged between 62 and 74 per cent according to the personality trait under analysis. Results seemed to indicate that simple algorithms such as Naïve Bayes or regression trees tended to perform better than more complex algorithms for textual data. Complex algorithms, such as support vector machine (SVM), tended to perform better for conversational data and big-data corpora. The authors concluded that models of observed personality could outperform models of self-assessed personality.

In social networks, observable information for personality assessment could comprise the publicly available information of profiles (e.g. Facebook profiles), the intensity and number of social interactions of an individual, reciprocity of relations (e.g. follower-followee relations on Twitter) and linguistic features. For example, Bai *et al.* (2012) represented users by a set of features such as friends count, weekly usage, gender, age, home-town, emotion count in tweets, emotion tendency and combinations of those features. Adali and Golbeck (2012) defined prediction features based on the behaviour towards friends and followers, the intensity and number of social interactions (e.g. number of re-tweets, conversations, messages per conversation), reciprocity of friendship relations, the textual analysis of messages and the LIWC features. Experimental evaluation was based on 71 users who answered the Big Five Inventory. The usage of two regression algorithms for prediction showed that the time between messages, text length and re-tweet rate were the most informative features. It was also found that user personality could be predicted by using either behavioural features or the LIWC features.

Golbeck *et al.* (2011) defined several features based on users' account statistics, such as the number of followees, mentions, replies, hashtags and URLs, and the content of tweets, including the LIWC, MRC psycholinguistic and the general inquire[2] features. Experimental evaluation was based on 50 Twitter users who were asked to answer a 45-question version of the Big Five Inventory. The usage of two regression algorithms for prediction showed that the openness to experience dimension was the easiest to predict, whereas neuroticism was the most difficult one. Content-based features proved to be more useful than statistical features for predicting personality. These results confirmed the findings of Mairesse *et al.* (2007).

2.2 How does personality affect social relationships?

Social networks are representations of relational patterns between individuals that, as real-world social relationships, evolve over time. Changes can be due to structural and network mechanisms such as reciprocity and transitivity, structural competition, or they could be related to the characteristics of the individuals (Snijders *et al.*, 2007). Different sociological theories have been postulated to explain the preference of individuals for interacting with others. Homophily (McPherson *et al.*, 2001) states that social interactions between similar individuals occur at a higher rate than among dissimilar ones. The theory suggests that cultural, behavioural and genetic information that flows through social networks tends to be alike. The reinforcement-affect theory (Selfhout *et al.*, 2009) states that similar values reinforce feelings, views and opinions, which trigger an implicit response that increases attraction. Byrne *et al.* (1967) stated that the degree of similarity and liking among individuals are linearly correlated, and thus dismissed the idea of opposite attraction.

Selfhout *et al.* (2009, 2010) studied the influence of the Big Five personality dimensions on the friendship selection process. Selfhout *et al.* (2009) aimed at studying the relation between the perceived similarity, actual similarity and peer-rated similarity during friendship selection in a naturalistic setting. Experimental evaluations were based on 205 psychology students divided into groups, who were asked to answer personality questionnaires and to describe the relationship with each member of the group. Regarding the personality dimensions (Selfhout *et al.*, 2010), results suggested that several Big Five dimensions have an important and differentiated role in friendship selection. Agreeable individuals tended to receive more friend nominations. The agreeableness, extraversion and openness to experience dimensions emerged as significant predictors of friendship ties, i.e. individuals tended to select friends with similar personality scores among those dimensions. The neuroticism and conscientiousness dimensions were more related to maintaining relationships than to establishing new ones. Results also suggested that individual dimension similarities have a greater impact on friendship selection than overall similarity across the five dimensions.

Regarding the actual and perceived personality similarity (Selfhout *et al.*, 2009), results reinforced the need to consider multiple personality assessments, as only perceived and peer-reported similarity were correlated with greater friendship intensity. Conversely, actual similarity between individuals was not associated with friendship intensity. Furthermore, the authors found a causality relation between similarity and attraction. Perceived and peer-rated similarity seemed to enhance friendship formation. Simultaneously, being be-friended enhanced similarity perceptions. Other works (Cuperman and Ickes, 2009) studied the effect of the Big Five personality dimensions on the formation of dyads. The effect of each dimension was separately studied in dyads of individuals with similar scores across dimensions, i.e. the effect of each dimension in combination with the others was not assessed. Experimental evaluation was based on 174 psychology students divided into dyads. The most significant effects were those associated to the extraversion and agreeableness dimensions, suggesting that individuals with high scores in the extraversion or openness to experience dimensions tended to be more interested in interacting with new acquaintances, and thus, forming new friendships. High-scored individuals in the agreeableness dimension were interested in continuing the new friendships in the future, even with individuals with lower scores in such dimension. Additionally, personality similarity allowed positive interactions in those dyads composed

by either extraverted or introverted individuals, when compared with dissimilar dyads. However, neither the neuroticism nor the conscientiousness dimensions showed significant effects over friendship processes. These results support those presented by Selfhout *et al.* (2010) regarding which dimensions are the most influential in the friendship selection processes.

In summary, the different personality dimensions affect the extent to which individuals are selected as friends, the friendship selection process, the size of the friends group and the similarity between friends across the personality dimensions. Table I summarises the effects of the Big Five personality dimensions on those processes as stated in the literature (Selfhout *et al.*, 2010; Cuperman and Ickes, 2009). Besides social interactions, Heinström (2003, 2010) stated that personality can also affect the behaviour of individuals regarding the information seeking process. The author established that the impact of personality on the information seeking process is dependent on the unique combination of traits that distinguish each individual. For example, extraversion was related to a preference for searching thought-provoking information over information that confirmed previous ideas. Openness to experience was related to broad information seeking and incidental information acquisition. Finally, both neuroticism and conscientiousness were associated with a preference of documents that confirmed previous information.

Dimension	Selecting friends	Being selected as a friend	Selecting similar friends
Agreeableness	Associated with more reciprocal friends	Tend to show more pro-social and altruist behaviour, and thus are more attractive as potential friends. Associated with more reciprocal friends, higher peer acceptance. Attract more individuals than extraverts	Individuals who select others with similar scores of agreeableness may benefit more from the relation than when selecting others who differ in terms of agreeableness
Extraversion	Tend to experience positive affect in social situations. Tend to be motivated to select friends. Associated with more reciprocal friends. The most important factor for selecting friends	Associated with more reciprocal friends. Associated with higher peer acceptance	Individuals tend to select others with similar extraversion scores
Openness to experience	Are more interested in interacting with new friends	No evidence of enhancing being selected as friends	Although it is neither a desirable or undesirable personality trait, a certain match may enhance friendship selection
Conscientiousness	Associated with more reciprocal friends	Associated with more reciprocal friends and higher peer acceptance	There is no evidence of actual similarity among friends
Neuroticism	More associated with maintaining relationships than with forming them	More associated with maintaining relationships than with forming them	There is no evidence of actual similarity among friends

The findings of the presented approaches are limited in their generality. All of the studies relied on samples of psychology students, thereby constraining the validity of results to individuals reporting similar age range, socio-economic status and college standards. Consequently, it remains to be determined whether results can be generalised to individuals of varying ages and life circumstances.

3. Research method

In this section we present the data collection used in the performed data analysis and we describe how the personality scores of users were computed (Section 3.1). Additionally, we describe the different alternatives proposed for assessing the impact of personality on social relationships (Section 3.2).

3.1 Data collection and processing

The Twitter dataset used for this study was obtained by crawling a set of 1,852 seed users selected from the set of users in (De Choudhury *et al.* 2010). All of the selected users had at least ten followees, ten published tweets and listed the language account as English. These restrictions were imposed to obtain meaningful user profiles for content analysis. For those users, all tweets, favourites, followees, followers and user account information were retrieved. Seed users were analysed to determine the influence of personality in the followee selection process. For this purpose, the same data were retrieved for their followees. We obtained all the data using the Twitter API[3]. Table II summarises the general statistics of the data collection.

This study uses the models and tool described by Mairesse *et al.* (2007) for automatically computing personality scores for the Big Five model. The tool was tailored according to the performance requirements posed by this study. SMOreg (Shevade *et al.*, 2000), an implementation of SVM for regression, was the selected model for computing the personality scores as it was reported to obtain the most accurate results for big-data corpora (Mairesse *et al.*, 2007).

3.2 Assessing user similarity based on personality

To verify the proposed hypotheses, three alternatives for measuring the similarity between a user and his/her followees regarding personality are proposed.

Overall similarity: to compute this measure, the five personality dimensions are considered as a whole. The scores of all the personality traits are used for computing

Total number of seed users	1,852
Total number of second-level users (followees)	545,286
Total number of tweets (seed users)	2,307,920
Average number of tweets per user (seed users)	1,247
Total number of tweets (followees)	1,058,285,978
Average number of tweets per user(followees)	1,941
Total number of favourite tweets (seed users)	316,419
Average number of favourite tweets per user (seed users)	171
Total number of favourite tweets (followees)	213,139,602
Average number of favourite tweets per user(followees)	391
Total number of followee relations (seed users)	780,220
Average number of followee relations per user (seed users)	422
Total number of followee relations (followees)	1,539,661,626
Average number of followee relations per user (followees)	2,824

Table II.
Data collection general statistics

the cosine similarity between two users, and thus computing an overall personality similarity between them.

Dimension-to-dimension: Selfhout *et al.* (2010) stated that the actual overall similarity regarding the five personality dimensions was neither a predictor of new social ties nor friendship intensity, thus highlighting the potential importance of examining each individual trait. In this context, this measure compares the score of a seed user and his/her followees regarding each individual and independent personality dimension.

Cross-dimension similarity: the personality scores of seed users in each dimension was discretised into three groups: low (scores ranging between 1 and 3), medium (scores ranging between 3 and 5) and high (score ranging between 5 and 7). Then, each group of seed users' scores is compared to the score of the followees for each of the other personality dimensions. For example, the extraversion score of seed users is compared to the agreeableness, conscientiousness, emotional stability and openness to experience scores of their followees.

4. Results and analysis

The distribution of personality scores across the Big Five dimensions for the 1,852 seed users is depicted in Figure 1. The predicted scores are similar across all dimensions, tending to be concentrated along the neutral values, i.e. the median value of the score distribution. Thus, the majority of the selected seed users did not manifest extreme personalities. The most disperse scores were found for the extraversion dimension. The quartile distribution of that dimension appeared to be scattered across a wider range of scores than for the other dimensions. The emotional stability, agreeableness and openness to experience dimensions presented more concentrated quartile distributions, implying that the majority of scores ranged between 3.5 and 4. The agreeableness and openness to experience dimensions presented the most scattered distribution of outliers, which implied that users covered a larger range of scores than for the other dimensions. Although most users showed neutral values, several users scored high on the agreeableness and openness to experience dimensions. Conversely, regarding the extraversion and emotional stability dimensions, the majority of outliers
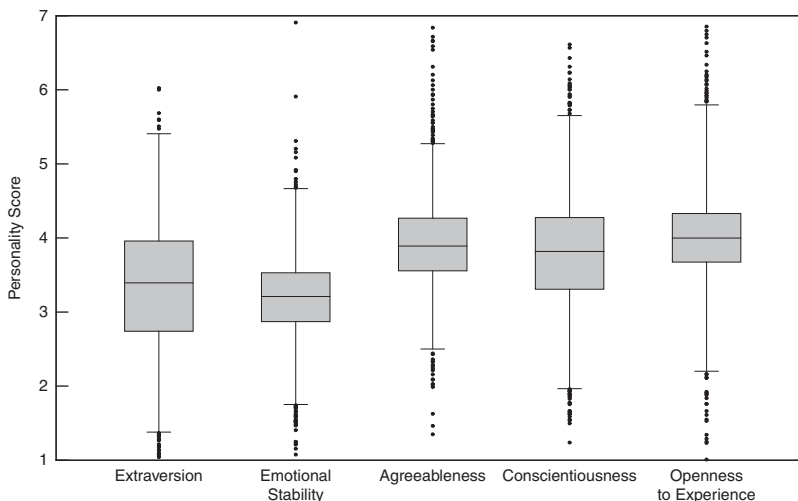

Figure 1.
Distribution of personality scores across the Big Five dimensions

was concentrated along low scores, showing the presence of introverted and highly neurotic users. The rest of this section analyses the similarity among seed users and their followees by the three alternatives proposed in Section 3.2.

4.1 Overall similarity
Most of the approaches presented in the literature (Tkalcic *et al.*, 2009; Hu and Pu, 2011) analysed personality by considering the overall similarity score between users using the cosine similarity among all dimensions. However, this measure might result inadequate for accurately assessing similarity between users as it tends to achieve high values, as Figure 2 shows. The figure depicts for each seed user (placed along the *X*-axis), the cosine similarity with each of his/her followees (placed along the *Y*-axis).

As the figure shows, the cosine similarity yields high values in most cases as high scores are obtained when at least one of the compared dimensions is similar. In other words, the overall similarity might be influenced by the score of only one dimension, regardless the score of the other dimensions. Consequently, the overall personality similarity might not accurately assess the actual similarity between users across the individual personality dimensions.

4.2 Dimension-to-dimension similarity
Figure 3 presents the personality scores of the seed users and their followees for each individual dimension. Seed users are sorted in ascending order according to their score in the analysed dimension (plotted in grey), whereas the personality score of their followees for the same dimension is depicted in the *Y*-axis. These figures allow to assess the impact of each personality dimension in the followee selection process.

Regarding the extraversion dimension, Figure 3(a) shows that high-scored seed users tended to relate with all kinds of users, i.e. followees are not concentrated along a specific score, but are distributed along the whole range of scores. Additionally, the density of followees increased as the personality score increased, implying that extraverted users tended to have more followees than introverted ones.

Regarding the agreeableness dimension, Figure 3(b) shows that the majority of followees did not present extreme scores, suggesting that users tend to relate with similar users in this personality trait. Interestingly, seed users with extreme personality scores accounted for the lowest number of followees. This implied that both high- and low-scored users in the agreeableness dimension tended to relate with fewer users than users with neutral scores.
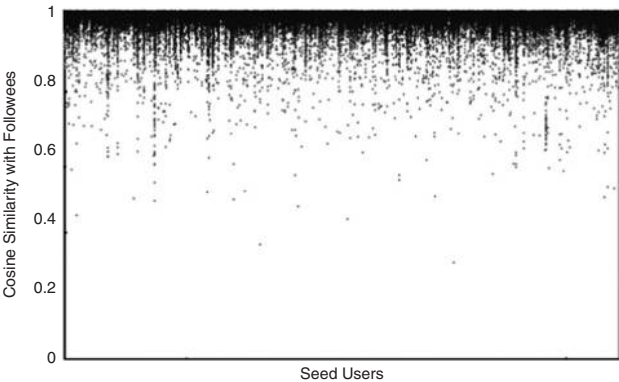

Figure 2.
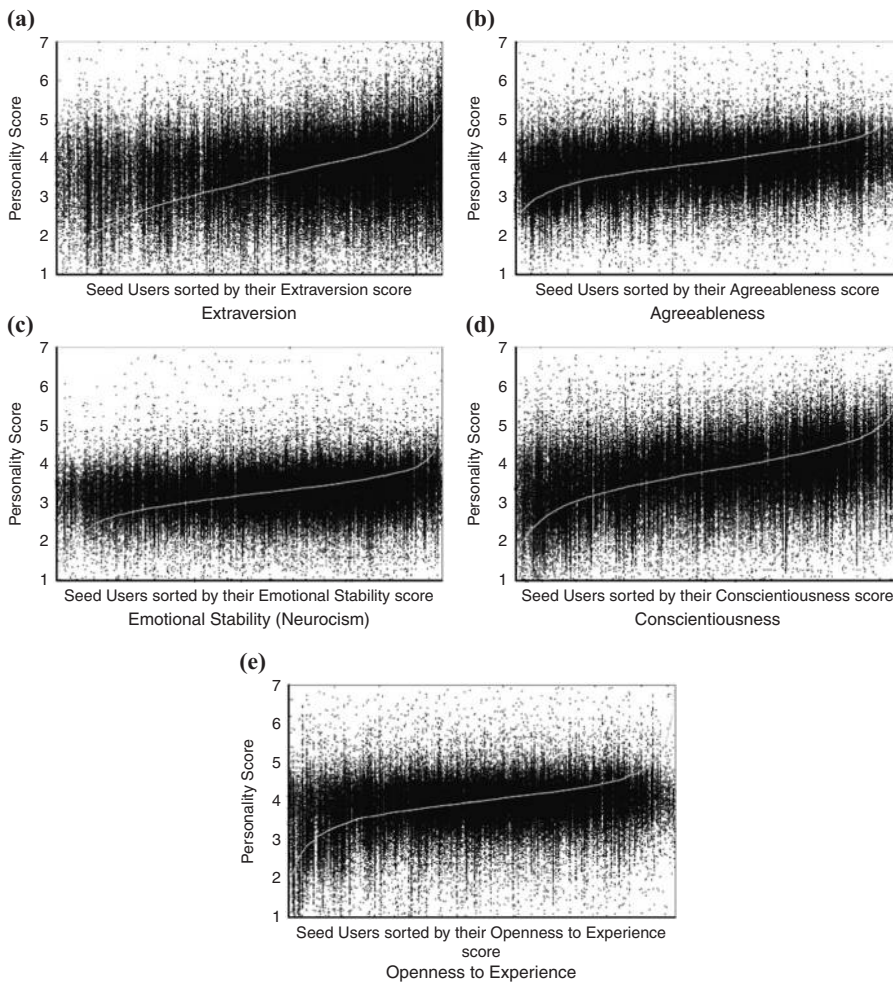Overall personality similarity between seed users and their followees

(a) Extraversion

(b) Agreeableness

(c) Emotional Stability (Neurocism)

(d) Conscientiousness

(e) Openness to Experience

Regarding the emotional stability dimension, Figure 3(c) shows that the majority of followees tended to concentrate on the neutral and lower scores, implying that they are not emotionally stable, i.e. they present neurotic characteristics. Few followees appeared to be emotionally stable. Furthermore, the number of followees seemed to increase as the emotional stability of seed users increased, i.e. neurotic users tended to relate with fewer user than emotional stable ones. Neurotic individuals are more sensitive than emotionally stable ones, and as a result, they tend to engage on fewer relationships and to carefully choose them. Interestingly, neurotic users tended to choose followees with higher emotional stability scores.

Regarding the conscientiousness dimension, Figure 3(d) shows a wide dispersion of followees across the personality scores. This personality dimension exhibited both the most diverse and highest personality scores. Interestingly, as the seed users conscientiousness score increased, their followees score also increased. The highest followees' scores corresponded to the followees of the most conscientious seed users.

Finally, regarding the openness to experience dimension, Figure 4(e) shows that most followees have a slightly tendency to be open to new experiences. However, as the personality score increased, the density of followees decreased.

4.3 Cross-dimension similarity
Figure 4 aims at analysing the effect of combining personality dimensions on the followee selection process. For each group of seed users, the figure shows the statistical distribution of scores regarding the other personality dimensions. Results revealed some variations regarding the followee preferences across the several personality groups. For example, regarding the extraversion dimension (Figure 4(a)), the preference of users was stable across the score groups. However, there was a tendency of users to select followees with lower emotional stability scores, regarding the scores of the other dimensions.
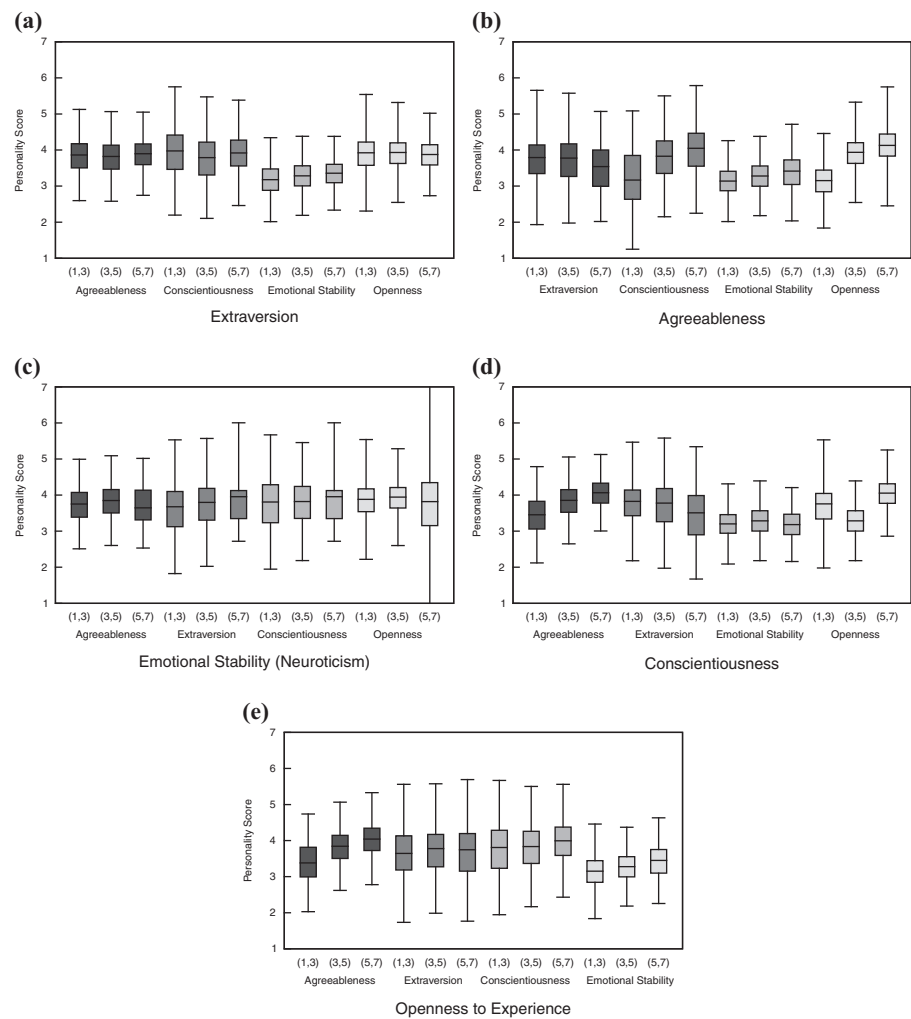

Figure 4.
Relation between the Big Five dimensions and followee personality preferences

Regarding the agreeableness dimension (Figure 4(b)), low-scored users tended to select followees with lower scores in the conscientiousness dimension than the followees selected by medium or high-scored users. Additionally, low-scored users in the agreeableness dimension tended to select followees with lower scores in the openness to experience dimension than the followees selected by medium or high-scored users.

Regarding the emotional stability dimension (Figure 4(c)), all users tended to select followees with medium scores across all dimensions. Regarding the conscientiousness dimension (Figure 4(d)), low-scored users tended to select followees belonging to a wider range of openness to experience scores than the followees selected by high-scored users. Finally, regarding the openness to experience dimension (Figure 4(e)), high-scored users tended to select followees with higher scores in the agreeableness dimension than the followees selected by low-scored users.

5. Discussion and implications

In this section, we present an analysis of each of the proposed sub-hypothesis that guided our study, in relation with the obtained results and the literature. Finally, we summarise the implications and possible applications of the presented data analysis.

5.1 Homophily in personality traits

The first sub-hypothesis aimed at verifying whether social interactions regarding personality are guided by the principle of homophily in the context of social networks. Furthermore, it aimed at verifying whether an overall assessment of personality is sufficient for accurately assessing the similarity between users.

Regarding the overall assessment of personality, computing the similarity simultaneously considering all the traits resulted in high scores when at least one of the dimensions was similar. Consequently, the overall similarity is influenced by the score of only one dimension. These results allowed to conclude that the overall personality similarity might not accurately assess the actual similarity between users across the individual personality dimensions. Thus, we consider that it is unsuitable for guiding the decision of a recommendation algorithm regarding the identification and ranking of potential followees. This finding agrees with those of (Selfhout *et al.*, 2010) stating that the effect of each personality dimension on friendship relations is higher and more important than the overall effect of the five dimensions considered as a whole.

Regarding the tendency of users to connect with similar followees, relations seemed to be guided by their similarity for at least three personality traits. Low-scored seed users tend to select users that are alike in the openness dimension. This tendency was also verified for the extraversion dimension, as introverted users tended to select introverted and neutral followees instead of highly extraverted followees. This agrees with the findings of Cuperman and Ickes (2009), which stated that introverted users tend to feel more comfortable with friends with similar personality scores, and prefer to continue their friend relations with other introverted users rather than with extraverted ones. Also, the agreeableness dimension confirmed the existence of homophily as seed users tended to relate with followees with similar scores, as stated in (Selfhout *et al.*, 2010). Finally, regarding the conscientiousness dimension, there was no evidence of similarity between seed users and their followees as described in (Selfhout *et al.*, 2010).

These results validated the hypothesis of the existence of homophily in social networks regarding the selection of users with similar personalities. Interestingly, although this data analysis was performed in an environment different from the ones

by Selfhout *et al.* (2010) and Cuperman and Ickes (2009), results showed that personality homophily relations can be verified in the context of the follower/followee relations in social networks.

5.2 Personality patterns in social relationships
The second sub-hypothesis aimed at verifying the existence of relationship patterns between users depending on their personality traits. In this regard, several patterns were discovered between users exhibiting certain scores across the different dimensions.

Regarding the extraversion dimension, high-scored seed users tended to relate with all kinds of users as their followees were distributed along the entire range of scores. These findings agree with those of Selfhout *et al.* (2010), which stated that extraverted users tend to experience more positive affect in social situations, and thus they are more motivated to be socially active, and initiate more friend relations (or in this case, followee relations). Introverted users tend to be shier, and thus do not engage on numerous friendship relations. Interestingly, introverted users tended to mainly select introverted and neutral followees instead of highly extraverted followees. This agrees with the findings of Cuperman and Ickes (2009), which stated that introvert users tend to feel more comfortable with friends with similar personality scores, and prefer to continue their friend relations with other introverted users rather than with extraverted ones.

According to Selfhout *et al.* (2010) individuals who select others with similar agreeableness scores might benefit more from the relation than when selecting others with dissimilar scores. The data analysis confirmed such pattern, as seed users tended to select followees with similar scores in the agreeableness dimension.

Regarding the emotional stability dimension, there was a tendency of low-scored users to relate with users with higher scores. This agrees with the findings of Cuperman and Ickes (2009) who stated that neurotic users tended to engage with more stable users to use their behaviour as a guide for their own.

The conscientiousness dimension exhibited both the most diverse and highest personality scores. Seed users tended to select followees with similar or higher scores. Additionally, there was a tendency of only selecting users with neutral or high scores. This agrees with the results of Selfhout *et al.* (2010). The authors argued that users with low scores have poorer self-control and tend to disclose inappropriate information. Thus, those users tend to engage in prejudiced relations, making them an unattractive followee choice for other users.

Finally, regarding the openness to experience dimension, we observed that the number of followees decreased as the score of seed users increased. This contrasted with the theories that users who are open to new experiences tend to engage on more friendships (Selfhout *et al.*, 2010). This might be due to the difference in the environment of both studies, and the distinctive nature of the follower/followee relationships in contrast to friendship ties. However, the same authors claimed that this dimension could be dismissed as they stated that it is concerned with the individual differences in mind functioning and structure, and thus it is of little importance to social relationships.

In summary, the data analysis verified the existence of several patterns regarding the selection of followees with certain personality characteristics. Particularly, social relation patterns for four personality dimensions were found. Those patterns confirm the validity of the findings presented by Selfhout *et al.* (2010) and Cuperman and Ickes (2009) in the context of Twitter, and thus validate the second sub-hypothesis defined in this study.

5.3 Cross-dimensional effect of personality

Considering that personality dimensions can be relatively independent from each other, each individual personality dimension might not be sufficient for accurately assessing followee preferences. According to Heinström (2003) the final impact of personality on behaviour is dependent on the unique combination of traits that distinguishes each individual. For example, if only the individual effect of each personality dimension is considered, an individual who scores low in the extraversion and high in the openness to experience dimensions would be recommended to follow both low- and high-scored followees in the emotional stability dimension. Moreover, an individual who scores medium in both the agreeableness and conscientiousness dimensions would be recommended to follow both low- and high-scored followees in the openness to experience dimension.

In some cases, conflicting personality traits might neutralise the impact of personality, whereas in other situations, a strong personality characteristic might dominate and override other tendencies (Heinström, 2003). In this context, further multi-variables studies that consider all the combinations of the different dimensions are needed to correctly recommend followees.

5.4 Findings and implications

The main goal of this study was to shed some light on the impact of personality traits on followee selection. For that purpose several hypotheses were defined. This study is among the firsts to study the impact of personality, which is considered as one of the primary factors that influence human behaviour and social relationships, on the selection of followees in micro-blogging communities. The findings of this study allowed to verify each of the three sub-hypotheses that were tested. Consequently, it can be stated that the general hypothesis regarding the existence of a relation between user personality and social interactions in the context of Twitter was also verified. More importantly, the performed data analysis confirmed the existence of patterns between personality and face-to-face friendships demonstrated by Selfhout *et al.* (2010) and Cuperman and Ickes (2009) on micro-blogging platforms. Furthermore, it was shown that personality has a distinctive impact on the conformation of social relations, particularly in the selection of followees. Table III summarises the findings of this study that confirm the results of Selfhout *et al.* (2010) and Cuperman and Ickes (2009).

The findings of this study could have several implications. First, as personality can condition the selection of followees, it could be inserted into a followee recommendation system. According to this study, adding personality to these systems could be beneficial for accurately guiding the search of potential followees, and thus, improving the quality of recommendations generated using content and/or topology. The findings could be the cornerstone of strategies for quantitatively analysing personality in combination with other factors. Second, several guidelines for assessing the personality matching can be derived from this study, which can be included in ad hoc personality similarity definitions. For example, the data analysis allowed to discard the cosine similarity as a useful measure of personality similarity. Third, this study demonstrated that virtual social relations in micro-blogging communities are guided by the same principles than face-to-face relationships. Consequently, theories and findings regarding traditional social relationships could be successfully applied also in the context of follower/followee relations in Twitter. These results are interesting considering that Twitter is an information-oriented social network, unlike Facebook, which is a friendship-oriented social network.

	Homophily	Patterns in social relations	Cross-dimensional effect
Extraversion	The hypothesis of homophily was verified Introverts tended to relate with other introverts	The existence of social relation patterns was verified Extroverted users tended to relate with all kinds of followees Introverted users tended to mainly select Introverted and neutral followees	The need of analysing the combination of traits in case of conflicting scores was verified
Aggreableness	The hypothesis of homophily was verified	The existence of social relation patterns was verified	The need of analysing the combination of traits in case of conflicting scores was verified
Emotional stability	–	The existence of social relation patterns was verified Neurotic users tended to relate with followees with higher scores	The need of analysing the combination of traits in case of conflicting scores was verified
Conscientiousness	The lack of actual similarity between friends was verified as in (Selfhout *et al.*, 2010)	The existence of social relation patterns was verified Only followees with neutral of high scores tended to be selected	The need of analysing the combination of traits in case of conflicting scores was verified
Openness to experience	The hypothesis of homophily was verified A certain similarity match enhanced the followee selection	–	–

Table III.
Summary of the findings of this study and its relations with other studies

6. Conclusions

In the context of recommendation systems, the selection of potential followees to suggest is a crucial issue given the exponential number of active users in micro-blogging communities. Thus, the criteria used to guide the search of candidate users and rank them have to be carefully evaluated. Traditionally, link prediction algorithms have been based on topological characteristics of networks, textual analysis or other individual issues, but not psychological characteristics such as personality. However, because of its effect over people's reactions and interactions with other individuals, personality is considered one of the primary factors that influence human behaviour. Consequently, such criteria need to be adapted and combined accordingly to the user behaviour in the community.

This study has shown that personality should be considered as a distinctive factor in the followee selection process as it can enhance the accuracy of followee recommendations. However, personality dimensions should not be analysed as a whole as its overall similarity may not accurately assess the actual personality matching between individuals. Instead, the performed data analysis showed the existence

of relations among the individual traits. For example, introverted users tended to mainly select introverted and neutral followees instead of highly extraverted followees. Also, the data analysis stated the importance of considering the combination of the different personality dimensions. This study showed the existence of several relations between the score in one dimension and the selection of followees with specific scores in the other dimensions. For example, regarding the agreeableness dimension, low-scored users tended to select followees with lower scores in the conscientiousness dimension than the followees selected by medium or high-scored users.

In conclusion, this study showed that personality has a distinctive effect over the followee selection process which can be leveraged to enhance followee recommendation systems. However, this work presents some limitations. First, the tool used for automatically computing the personality scores of users has a maximum reported precision of 74 per cent. This could hinder the accurate personality characterisation of users, and in turn affect the quality of recommendations. Still, the automatic detection of personality is vital in social media due to the massive number of users, and their reluctance to complete long questionnaires. Interestingly, although the precision results might seem low, Mairesse *et al.* (2007) stated that their precision outperformed models of self-assessed personality. Second, in spite of including a large sample of Twitter users, the data set did not include users with extreme behaviours regarding any of the personality traits. Therefore, the exact effect of personality in the followee selection process for that type of users' needs to be further studied.

Notes

1. www.liwc.net/index.php

2. www.wjh.harvard.edu/~inquirer/Home.html

3. https://api.twitter.com

References

Adali, S. and Golbeck, J. (2012), "Predicting personality with social behavior", *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Istanbul*, pp. 302-309.

Armentano, M., Godoy, D. and Amandi, A. (2011), "A topology-based approach for followees recommendation in Twitter", *Proceedings of the Intelligent Techniques for Web Personalization & Recommendation (ITWP'11), Barcelona*, pp. 22-29.

Bai, S., Zhu, T. and Cheng, L. (2012), "Big-five personality prediction based on user behaviors at social network sites", CoRR abs/1204.4809.

Block, J. (1995), "A contrarian view of the five-factor approach to personality description", *Psychological Bulletin*, Vol. 117 No. 2, pp. 187-215.

Byrne, D., Griffitt, W. and Stefaniak, D. (1967), "Attraction and similarity of personality characteristics", *Journal of Personality and Social Psychology*, Vol. 5 No. 1, pp. 82-90.

Chen, H., Cui, X. and Jin, H. (2014), "Top-followee recommendation over microblogging systems by exploiting diverse information sources", *Future Generation Computer Systems*, 26 May, available at: http://dx.doi.org/10.1016/j.future.2014.05.002

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

Costa, P.T. Jr and McCrae, R.R. (1992), *Revised NEO Personality Inventory (NEO PI-R) and NEP Five-factor Inventory (NEO-FFI): Professional Manual*, Psychological Assessment Resources, Lutz, FL.

Costa, P.T. Jr and McCrae, R.R. (1994), "Set like plaster? Evidence for the stability of adult personality", in Heatherton, T. and Weinberger, J. (Eds), *Can Personality Change?, American Psychological Association*, Washington, DC, pp. 21-40.

Costa, P.T. Jr and McCrae, R.R. (1997), "Longitudinal stability of adult personality", in Hogan, R., Johnson, J.A. *et al.* (Eds), *Handbook of Personality Psychology*, Academic Press, Inc., San Diego, CA, pp. 269-290.

Cuperman, R. and Ickes, W. (2009), "Big five predictors of behavior and perceptions in initial dyadic interactions: personality similarity helps extraverts and introverts, but hurts 'disagreeables'", *Journal of Personality and Social Psychology*, Vol. 97 No. 4, pp. 667-684.

De Choudhury, M., Lin, Y.-R., Sundaram, H., Candan, K.S., Xie, L. and Kelliher, A. (2010), "How does the data sampling strategy impact the discovery of information diffusion in social media?", *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM'10), Washington, DC, 23-26 May*.

Funder, D.C. (2012), *Personality Puzzle*, W.W. Norton & Company, Inc, New York, NY.

Golbeck, J., Robles, C., Edmondson, M. and Turner, K. (2011), "Predicting personality from Twitter", *IEEE Third International Conference on Social Computing (SocialCom/PASSAT)*, *Boston, MA*, pp. 149-156.

Gottschalk, L.A. and Gleser, G.C. (1969), *The Measurement of Psychological States Through the Content Analysis of Verbal Behavior*, University of California Press, Oakland, CA.

Heinström, J. (2003), "Five personality dimensions and their influence on information behaviour", *Information Research,*, Vol. 9 No. 1, available at: www.informationr.net/ir/9-1/paper165.html

Heinström, J. (2010), *From Fear to Flow: Personality and Information Reactions*, Chandos Information Professional Series, Chandos, Witney, Oxford.

Herlocker, J.L., Konstan, J.A., Terveen, L.G. and Riedl, J.T. (2004), "Evaluating collaborative filtering recommender systems", *ACM Transactions on Information Systems*, Vol. 22 No. 1, pp. 5-53.

Hu, R. and Pu, P. (2011), "Enhancing collaborative filtering systems with personality information", *Proceedings of the 5th ACM Conference on Recommender Systems (RecSys' 11), Chicago, IL*, pp. 197-204.

McCrae, R.R. and Costa, P.T. (1982), "Self-concept and the stability of personality: cross-sectional comparisons of self-reports and ratings", *Journal of Personality and Social Psychology*, Vol. 43 No. 6, pp. 1282-1292.

McCrae, R.R. and Costa, P.T. Jr (1989), "More reasons to adopt the five-factor model", *American Psychologist*, Vol. 44 No. 2, pp. 451-452.

McPherson, M., Smith-Lovin, L. and Cook, J.M. (2001), "Birds of a feather: homophily in social networks", *Annual Review of Sociology*, Vol. 27 No. 1, pp. 415-444.

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.

Moss, H.A. and Susman, E.J. (1980), "Longitudinal study of personality development", in Brim, G. (Ed.), *Constancy and Change in Human Development*, Harvard University Press, Cambridge, MA, pp. 530-595.

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.

Pennebaker, J.W., Mehl, M.R. and Niederhoffer, K.G. (2003), "Psychological aspects of natural language use: our words, our selves", *Annual Review of Psychology*, Vol. 54 No. 1, pp. 547-577.

Pennebaker, J.W., Chung, C.K., Ireland, M., Gonzales, A. and Booth, R.J. (2007), "The development and psychometric properties of LIWC2007", Conjunction with the LIWC2007 Software Program, LIWC Inc., Austin, TX, available at: www.liwc.net/LIWC2007 LanguageManual.pdf (accessed 15 April 2015).

Rosenberg, S.D. and Tucker, G.J. (1979), "Verbal behavior and schizophrenia: the semantic dimension", *Archives of General Psychiatry*, Vol. 36 No. 12, pp. 1331-1337.

Selfhout, M., Denissen, J., Branje, S. and Meeus, W. (2009), "In the eye of the beholder: perceived, actual, and peer-rated similarity in personality, communication, and friendship intensity during the acquaintanceship process", *Journal of Personality and Social Psychology*, Vol. 96 No. 6, pp. 1152-1165.

Selfhout, M., Burk, W., Branje, S., Denissen, J., van Aken, M. and Meeus, W. (2010), "Emerging late adolescent friendship networks and Big Five personality traits: a social network approach", *Journal of Personality*, Vol. 78 No. 2, pp. 509-538.

Sherman, R.A., Nave, C.S. and Funder, D.C. (2012), "Properties of persons and situations related to overall and distinctive personality-behavior congruence", *Journal of Research in Personality*, Vol. 46 No. 1, pp. 87-101.

Shevade, S.K., Keerthi, S.S., Bhattacharyya, C. and Murthy, K.R.K. (2000), "Improvements to the SMO algorithm for SVM regression", *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 11 No. 5, pp. 1188-1193.

Snijders, T.A., Steglich, C.E. and Schweinberger, M. (2007), "Modeling the co-evolution of networks and behavior", in Satorra, A. and Oud, J. (Eds), *Longitudinal Models in the Behavioral and Related Sciences*, Lawrence Erlbaum, Mahwah, NJ, pp. 41-71.

Tkalcic, M., Kunaver, M., Tasic, J. and Košir, A. (2009), "Personality based user similarity measure for a collaborative recommender system", *Proceedings of the 5th Workshop on Emotion in Human-Computer Interaction-Real World Challenges*, pp. 30-37.

Tupes, E.C. and Christal, R.E. (1961), "Recurrent personality factors based on trait ratings", technical report, Lackland Air Force Base, San Antonio, TX.

Tupes, E.C. and Christal, R.E. (1992), "Recurrent personality factors based on trait ratings", *Journal of Personality*, Vol. 60 No. 2, pp. 225-251.

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.

Wu, H., Sorathia, V. and Prasanna, V.K. (2012), "Predict whom one will follow: followee recommendation in microblogs", *International Conference on Social Informatics (Social Informatics)*, pp. 260-264.

About the authors

Antonela Tommasel is a PhD Candidate at the Universidad Nacional del Centro de la Pprovincia de Buenos Aires (UNCPBA), and a Member of the ISISTAN Research Institute, Tandil, Argentina. She is also a Teacher Assistant at the same university. Her research interests include recommender systems, text mining and social web. She obtained her system engineer degree (2012) at the UNCPBA. Antonela Tommasel is the corresponding author and can be contacted at: antonela.tommasel@isistan.unicen.edu.ar

Alejandro Corbellini is a PhD Candidate at the Universidad Nacional del Centro de la Provincia de Buenos Aires (UNCPBA), and a Member of the ISISTAN Research Institute, Tandil, Argentina. He is also a Teacher Assistant at the same university. His research interests

include recommender systems, graph processing and storage systems and distributed systems. He obtained his system engineer degree (2011) at the UNCPBA.

Daniela Godoy is a Researcher at the CONICET and a Member of the ISISTAN Research Institute, Tandil, Argentina. She is also a Full-Time Professor in the Department of Computer Science at the UNCPBA, Tandil, Argentina. 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.

Silvia Schiaffino is a Researcher at the 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.