Exploring the Role of Personality Traits in Followee Recommendation

Abstract

Purpose

Followee recommendation is a problem rapidly gaining importance in *Twitter* and other micro-blogging communities. To find interesting users to follow, most recommendation systems leverage different traditional factors such as graph topology, or user-generated content, among others, disregarding 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, this study aims at shedding some light on the impact of personality traits on followee selection.

Design/Methodology/Approach

A data analysis comparing the similarity among *Twitter* users and their followees regarding the personality traits of dimensions was performed. Three different similarity measures were analysed. First, an overall similarity considering the five personality dimensions of the Five-Factor model as a whole. Second, a dimension-to-dimension similarity considering each individual personality dimension independently of each other. Third, 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 may not accurately assess the actual matching between individuals. Instead, the performed data analysis showed the existence of relations among the individual dimensions, and stated the importance of considering each personality dimension in combination with the others.

Originality/Value

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 thus, social relationships, in the selection of followees in micro-blogging communities.

Keywords: Followee Recommendation; Twitter; Human Aspects Recommendation; Personality Traits

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 or even people to follow in the context of social networks by giving them personalised advice. Usually, recommendation systems present lists of suggested or interesting items to users, where items can also be people that they might be interested in, based on personal interests or even the opinion of like-minded users. Particularly, 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 who to follow. For example, a user might follow some users because they share mutual friends, others because they tweet interesting information, others because they share the same interests, and others because they are celebrities or popular individuals in the micro-blogging community, among other plausible explanations. 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 content and topology, neglecting 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 thus, social relationships, as it affects how people react and interact with other individuals. Due to this reason, this study aims at shedding some light on the impact of personality traits in the accurate prediction of followees. Particularly, this work focuses on analysing the impact of personality 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 user 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 & 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). 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 one states the existence of several relationship patterns depending on the personality of the involved users. It aims at verifying some of the relationship patterns described in (Selfhout et al. 2010, Cuperman & Ickes 2009), but in the context of micro-blogging platforms, such as *Twitter*. Finally, as Selfhout et al. (2010) and Cuperman & Ickes (2009) only analysed the effect of the separate dimensions, it is unknown whether the

combination of personality dimensions have an effect on social interactions. In this regard, the third sub-hypothesis states that 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* data used for experimentation. Section 3.2 analyses the followee preferences of the studied users regarding the personality of their followee. 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 originate internally in each person, and the temperamental and behavioural response patterns (Funder 2012, Adali & 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 & McCrae 1994, 1997, McCrae & Costa 1982, Moss & 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 & Ickes 2009), which could imply that people with similar personality would have similar interests.

Several works have aimed at finding a set of features or characteristics to describe personality. Tupes & Christal (1961, 1992) were the first authors that identified five recurrent features in personality. Subsequent works (Noller et al. 1987, McCrae & Costa Jr 1989) confirmed those findings and offered evidence of the existence of such features. The Five-Factor or Big Five model (Costa Jr

& McCrae 1992) is a hierarchical model that defines personality as a composition of five traits or dimensions. Although its theoretical foundations have been objected (Waller & 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 trust other people easily. 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 & Golbeck 2012, Golbeck et al. 2011) have cast the problem of determining personality as a classification or regression problem over directly observable information, such as text, conversations or conversational transcripts. Gottschalk & Gleser (1969), Rosenberg & Tucker (1979) among others, 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. Mairesse et al. (2007) 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 Concientiousness dimensions. Prediction accuracy ranged between 62% and 74% according to the personality trait under analysis. Results seemed to indicate that simple algorithms such as Naïve Bayes or regression tress tended to perform better than more complex algorithms for textual data. On the contrary, complex algorithms, such as 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 post, emotion tendency, and combinations of those features. Adali & 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' accounts statistics, such as 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. 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 sug-

gests 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. (2010, 2009) 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 received 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. On the contrary, 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. Cuperman & 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. On the contrary, neither the Neuroticism nor the Conscientiousness dimensions showed significant effects over friendship processes. These results support those

presented in (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 1 summarises the effects of the Big Five personality dimensions on those processes as stated in the literature (Selfhout et al. 2010, Cuperman & 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 for 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.

[Table 1 about here.]

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 determine whether results can be generalised to individuals of varying ages and life circumstances.

3 Research Method

This section presents the data collection used in the data analysis performed and describes how the personality scores of users were computed (Section 3.1). Additionally, it describes the different alternatives proposed for assessing the impact of personality in 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 10 followees, 10 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, followees, followers, user account information, and favourites were retrieved. Seed users were analysed to determine the influence of personality in the followee selection process. For this purpose, the same data was retrieved for their followees. All data was obtained through the *Twitter* API[3]. Table 2 summarises the general statistics of the data collection.

[Table 2 about here.]

This study uses the models and tool developed 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 Support Vector Machines (SVMs) 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   The five personality dimensions are considered as a whole. The scores for each of the personality traits are used for computing 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 alternative compares the score of a seed user and his/followees regarding each individual and independent personality dimension.

Cross-dimension Similarity   The personality scores of seed users in each dimension were divided 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 user's 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 the 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. On the contrary, 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

Fig. 1: Distribution of personality scores across the Big Five dimensions
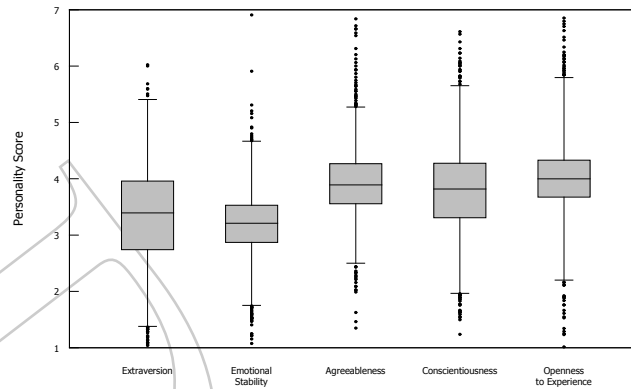
and Openness to Experience dimensions presented the most scattered distribution of outliers, which implied that users covered a larger range of scores than for the other dimensions. Although the majority of users was concentrated on neutral values, several users scored high on the Agreeableness and Openness to Experience dimensions. Conversely, regarding the Extraversion and Emotional Stability dimensions, the majority of outliers 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 & 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 of seed user (placed along the X-axis), the cosine similarity with each of their 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 are 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

Fig. 2: Overall personality similarity between seed users and their followees
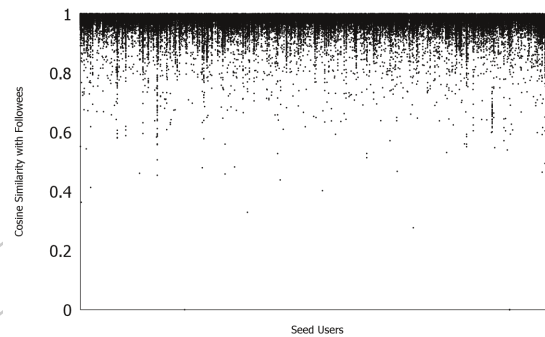
their score in the analysed dimension (plotted in Grey), whereas the personality score of their followees for the same dimension is depicted in the Y-axis. These figures allow to assess the impact of each personality dimension in the followee selection process.

Regarding the Extraversion dimension, Figure 3a 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 rather 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 3b 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.

Regarding the Emotional Stability dimension, Figure 3c 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 individuals, and as a result, tend to engage on fewer relationships but to carefully choose them. Interestingly, Neurotic users tended to choose followees with higher emotional stability scores.

Regarding the Conscientiousness dimension, Figure 3d 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

Fig. 3: Personality score distribution among the Big Five dimension between seed users and their followees

(a) Extraversion

(b) Agreeableness

(c) Emotional Stability (Neurocism)

(d) Conscientiousness

(e) Openness to Experience
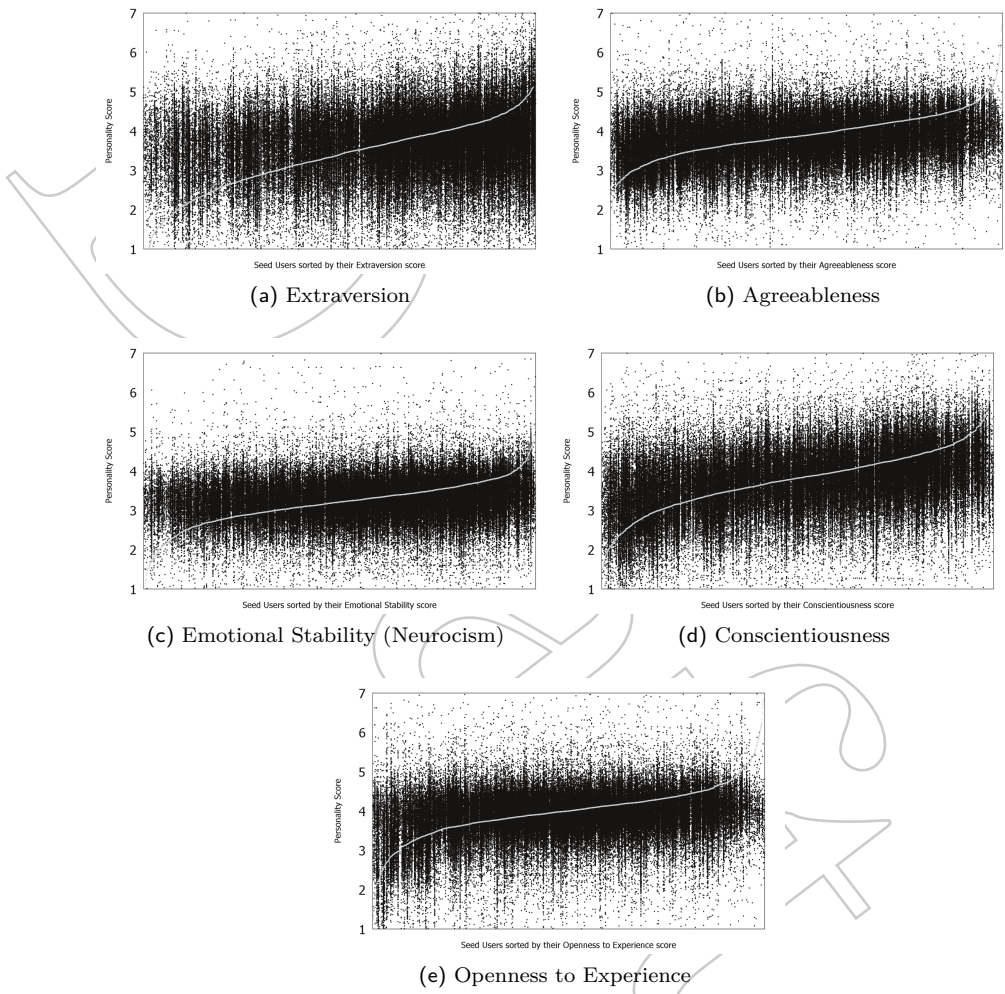
conscientious seed users. Finally, regarding the Openness to Experience dimension, Figure 4e shows that the majority of followees have a slightly tendency to be open to new experiences. However, as the personality score increased, the density of followees decreased.

4.3 Cross-Dimension Similarity

Figure 4 aims at analysing the effect of combining personality dimensions in the followee selection process. For each group of seed users, the statistical distribution of scores regarding the other personality dimensions are shown. Results revealed some variations regarding the followee preferences across the several personality groups. For example, regarding the Extraversion dimension (Figure 4a), 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.

Regarding the Agreeableness dimension (Figure 4b), 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 4c), all users tended to select followees with medium scores across all dimensions. Regarding the Conscientiousness dimension (Figure 4d), 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 4e), high-scored users tended to select followees with higher scores in the Agreeableness dimension than the followees selected by low-scored users.
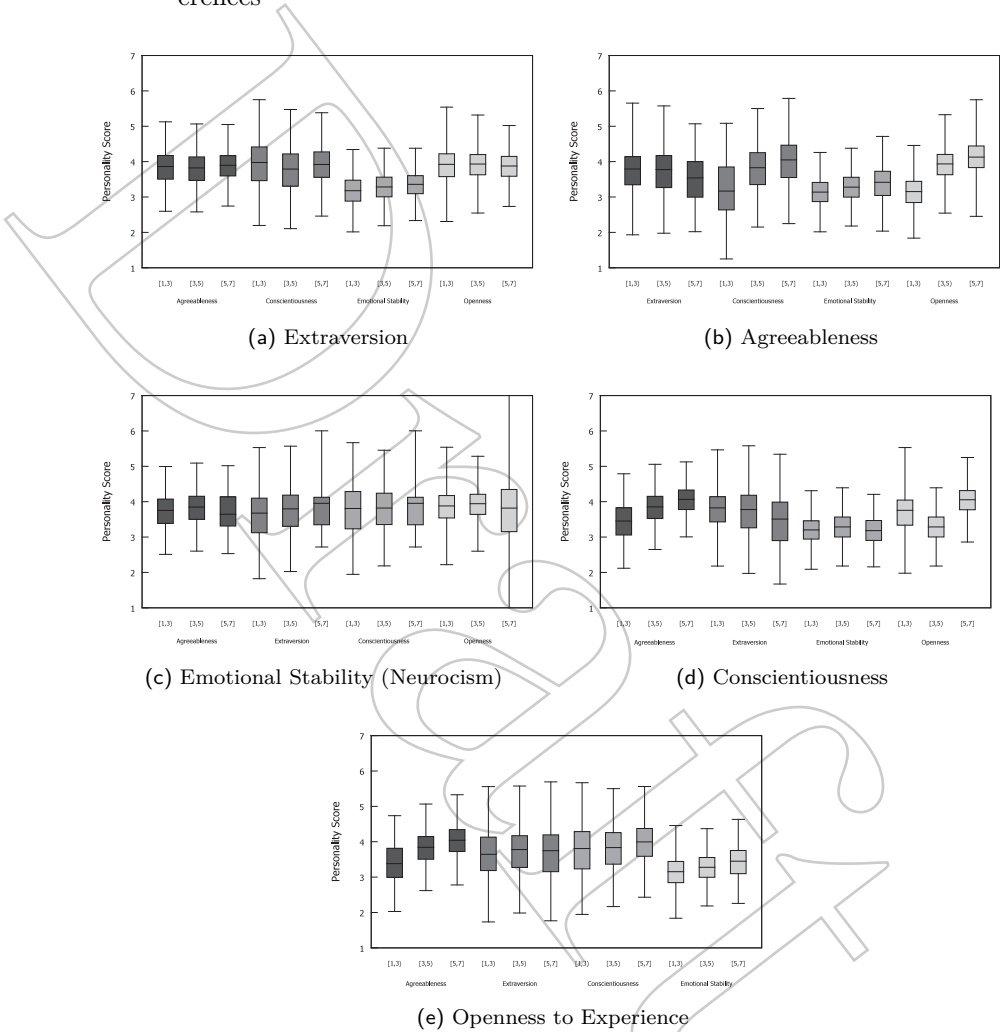
5   Discussion and Implications

This section presents an analysis of each of the proposed sub-hypothesis that guided this study, in relation with the obtained results and the literature. Finally, Section 5.4 summarises 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 considering all traits resulted in high scores when the scores for at least one of the dimensions were similar. Consequently, the overall similarity is influenced

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

(a) Extraversion

(b) Agreeableness

(c) Emotional Stability (Neurocism)

(d) Conscientiousness

(e) Openness to Experience

by the score of only one dimension, regardless the score of the other dimensions. 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, it is unsuitable for guiding the decision of a recommendation algorithm regarding the identification and ranking of possible followees. This finding agrees with those of Selfhout et al. (2010) who stated 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. A certain similarity match in the Openness to Experience dimension enhanced friendship selection between low-scored seed users and their followees. This tendency was also verified for the Extraversion dimension as Introverted users tended to mainly select Introverted and neutral followees instead of highly Extraverted followees. This agrees with the findings in (Cuperman & 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 stated 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 one in (Selfhout et al. 2010, Cuperman & Ickes 2009), results showed that personality homophily relations exposed by the authors verify 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 patters 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 whole range of scores. These findings agree with those in (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. On the contrary, 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 in (Cuperman & 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 & Ickes (2009) who stated that Neurotic users tended to engage with more stable users to use the other individuals 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, it could be 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 respond to the difference in the environment of both data analysis, 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 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 in (Selfhout et al. 2010, Cuperman & 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 distinguish each individual. For example, if only the individual effect of each personality dimension is considered, an individual who scores both 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 combination 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 in followee selection. For that purpose several hypothesis 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 thus social relationships, in the selection of followees in micro-blogging communities. The findings of this study allowed to verify each of the three sub-hypotheses that were proposed. 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 in (Selfhout et al. 2010, Cuperman & Ickes 2009) on micro-blogging platforms. Furthermore, it can be stated that personality has a distinctive impact on the conformation of social relations, particularly in the selection of followees. Table 3 summarises the findings of this study that confirm the results of both (Selfhout et al. 2010, Cuperman & Ickes 2009).

[Table 3 about here.]

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. Currently, most of the existing recommendation systems only rely on content and topological characteristics, neglecting how user's preferences and decisions are affected by psychological characteristics, such as personality. According to this study, adding personality could be beneficial for accurately guiding the search of potential followees, and thus, improving the quality of recommendations generated using content or topology. The findings could be used for developing strategies for quantitatively analysing personality, and thus, enriching followee recommendation systems. Second, several guidelines for assessing the personality matching can be derived from this study, which can be included in ad-hoc personality similarity definitions. Moreover, 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

of face-to-face relationship. 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.

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 has to be carefully evaluated. Traditionally, link prediction algorithms have been based on topological, textual analysis or other individual issues, disregarding the effect of psychological characteristics, such as personality, over the followee selection process. 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. Consequently, personality can be used to enhance followee recommendation systems by providing more accurate recommendations. However, this work presents some limitations. First, the tool used for automatically computing the personality scores of users has a maximum precision of 74%. This could hinder the accurate personality characterisation of users, and in turn affect the quality of recommendations. However, the automatic detection of personality is crucial 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 dataset 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.

References

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

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

Bai, S., Zhu, T. & 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* **117**(2), 187–215.

Byrne, D., Griffitt, W. & Stefaniak, D. (1967), 'Attraction and similarity of personality characteristics', *Journal of Personality and Social Psychology* **5**(1), 82–90.

Chen, H., Cui, X. & Jin, H. (2014), 'Top- followee recommendation over microblogging systems by exploiting diverse information sources', *Future Generation Computer Systems* (0), –.

Coltheart, M. (1981), 'The mrc psycholinguistic database', *Quarterly Journal of Experimental Psychology* **33**(4), 497–505.

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

Costa Jr, P. T. & McCrae, R. R. (1994), *Can personality change?*, American Psychological Association, chapter Set like plaster? Evidence for the stability of adult personality, pp. 21–40.

Costa Jr, P. T. & McCrae, R. R. (1997), Longitudinal stability of adult personality, *in* R. Hogan, J. A. Johnson & et al, eds, 'Handbook of Personality Psychology', Academic Press, Inc, San Diego, USA, pp. 269–290.

Cuperman, R. & 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* **97**(4), 667.

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

Funder, D. C. (2012), *Personality Puzzle*, W. W. Norton, Incorporated.

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

Gottschalk, L. A. & Gleser, G. C. (1969), *The measurement of psychological states through the content analysis of verbal behavior*, Univ. of California Press.

Heinström, J. (2003), 'Five personality dimensions and their influence on information behaviour.', *Inf. Res.* **9**(1).

Heinström, J. (2010), *From Fear to Flow: Personality and Information Reactions*, Chandos information professional series, Chandos.

Herlocker, J. L., Konstan, J. A., Terveen, L. G. & Riedl, J. T. (2004), 'Evaluating collaborative filtering recommender systems', *ACM Transactions on Information Systems* **22**(1), 5–53.

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

Mairesse, F., Walker, M., Mehl, M. & Moore, R. (2007), 'Using linguistic cues for the automatic recognition of personality in conversation and text', *Journal of Artificial Intelligence Research* **30**, 457–500.

McCrae, R. R. & Costa Jr, P. T. (1989), 'More reasons to adopt the five-factor model', *American Psychologist* **44**(2), 451–452.

McCrae, R. R. & 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* **43**(6), 1282–1292.

McPherson, M., Smith-Lovin, L. & Cook, J. M. (2001), 'Birds of a feather: Homophily in social networks', *Annual Review of Sociology* **27**(1), 415–444.

Moss, H. A. & Susman, E. J. (1980), *Constancy and change in human development*, Harvard University Press, chapter Longitudinal study of personality development, pp. 530–595.

Noller, P., Law, H. & Comrey, A. L. (1987), 'Cattell, Comrey, and Eysenck personality factors compared: More evidence for the five robust factors?', *Journal of Personality and Social Psychology* **53**(4), 775–782.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A. & Booth, R. J. (2007), 'The development and psychometric properties of LIWC2007', This article is published by LIWC Inc, Austin, Texas 78703 USA in conjunction with the LIWC2007 software program.
URL: *http://www.liwc.net/LIWC2007LanguageManual.pdf*

Pennebaker, J. W., Mehl, M. R. & Niederhoffer, K. G. (2003), 'Psychological aspects of natural language use: Our words, our selves', *Annual Review of Psychology* **54**(1), 547–577.

Rosenberg, S. D. & Tucker, G. J. (1979), 'Verbal behavior and schizophrenia: The semantic dimension', *Archives of General Psychiatry* **36**(12), 1331–1337.

Selfhout, M., Burk, W., Branje, S., Denissen, J., van Aken, M. & Meeus, W. (2010), 'Emerging late adolescent friendship networks and Big Five personality traits: A social network approach', *Journal of Personality* **78**(2), 509–538.

Selfhout, M., Denissen, J., Branje, S. & 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* **96**(6), 1152–1165.

Sherman, R. A., Nave, C. S. & Funder, D. C. (2012), 'Properties of persons and situations related to overall and distinctive personality-behavior congruence', *Journal of Research in Personality* **46**(1), 87–101.

Shevade, S. K., Keerthi, S. S., Bhattacharyya, C. & Murthy, Murthy, K. R. K. (2000), 'Improvements to the smo algorithm for svm regression.', *IEEE Transactions on Neural Networks and Learning Systems* **11**(5), 1188–1193.

Snijders, T. A., Steglich, C. E. & Schweinberger, M. (2007), *Longitudinal models in the behavioral and related sciences*, Lawrence Erlbaum, Mahwah, NJ, chapter Modeling the co-evolution of networks and behavior, pp. 41–71.

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

Tupes, E. C. & Christal, R. E. (1961), Recurrent personality factors based on trait ratings, Technical report, Lackland Air Force Base, TX: US Air Force.

Tupes, E. C. & Christal, R. E. (1992), 'Recurrent personality factors based on trait ratings', *Journal of Personality* **60**(2), 225–251.

Waller, N. G. & Ben-Porath, Y. S. (1987), 'Is it time for clinical psychology to embrace the five-factor model of personality?', *American Psychologist* **42**(9), 887–889.

Wu, H., Sorathia, V. & Prasanna, V. K. (2012), Predict whom one will follow: Followee recommendation in microblogs, *in* 'International Conference on Social Informatics (SocialInformatics)', pp. 260–264.

Notes

[1]http://www.liwc.net/index.php
[2]www.wjh.harvard.edu/~inquirer/Home.html
[3]https://api.twitter.com

TABLES   22

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, 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, higher peer acceptance.	There is no evidence of actual similarity among friends.
Neuroticism	More associated with maintaining relationships than forming them.	More associated with maintaining relationships than forming them.	There is no evidence of actual similarity among friends.

Tab. 1: Effects of personality dimensions on friendship processes

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

Tab. 2: Data collection general statistics

	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	Verify the lack actual similarity between friends as in (Selfhout et al. 2010).	The existence of social relation patterns was verified. Only followees with neutral or 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.	-	-

Tab. 3: Summary of the findings of this study and its relations with other studies