

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/94124/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Noe, Nyala, Whitaker, Roger and Allen, Stuart 2016. Personality homophily and the local network characteristics of Facebook. Presented at: The 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, San Francisco, CA, USA, 18 - 21 August 2016. Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2016). IEEE, pp. 386-393. 10.1109/ASONAM.2016.7752263

Publishers page: <http://dx.doi.org/10.1109/ASONAM.2016.7752263>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



# Personality Homophily and the Local Network Characteristics of Facebook

Nyala Noë\*, Roger M. Whitaker<sup>†</sup>, and Stuart M. Allen<sup>‡</sup>  
 School of Computer Science and Informatics, Cardiff University

Cardiff, United Kingdom

Email: \*NoeN@cardiff.ac.uk, <sup>†</sup>WhitakerRM@cardiff.ac.uk, <sup>‡</sup>AllenSM@cardiff.ac.uk

**Abstract**—Social networks are known to form on the basis of homophily, where nodes with some type of similar characteristics are more likely to be connected. Some of the most fundamental human characteristics are reflected by an individual’s personality, which represents a persistent disposition governing a human’s outlook and approach to diverse situations. While taking into account demographics of age and gender, we assess the extent to which personality homophily is evident in the local network features of Facebook. Using a large sample obtained from the MyPersonality dataset, we find that a range of network-based features correlate with personality facets of individuals. In particular, extraversion had a positive effect on an individual’s network size, while neuroticism had a negative effect. Additionally, extraversion and openness were positively related to transitivity, which was moderated by gender. Finally, we found that conscientiousness, agreeableness and extraversion were homophilous: people with higher similarity on these facets were more strongly connected. This was additionally mediated by gender for agreeableness: personality similarity had an effect for male-only and mixed pairs, but not for female-only pairs. Personality similarity was also stronger among closed triangles, compared to open ones. These results support the idea that inherent attraction between individuals, on the basis of personality, drives the roles we play within our online social networks.

## I. INTRODUCTION

Instinctive human behavior contributes to the interesting characteristics of social networks. In particular, assortative mixing, or homophily, the tendency of similar nodes to be connected, is a core characteristic of social networks [1], [2]. Similarity makes it easier to predict the behaviors of others, which in turn eases communication. Similarity also increases trust and solidarity. These beneficial effects of similarity contribute to explaining why people have a preference for homophilous ties, with similarity between friends being stronger than between non-friends [2].

The basis for homophilic attraction is diverse, with many different ways for individuals to establish meaningful commonality. Examples include gender, age, race, religion, and education [3], and even the commonality of location decisions [4], [5]. A fundamental but often overlooked aspect is the inherent disposition of individuals. This is represented by human personality, which is the set of pervasive traits that affect our general disposition and approach to behavior, attitudes and decisions in wide ranging settings [6], [7]. This extends to how individuals choose to relate to each other and sustain relationships.

In this paper we focus on the issue of similarities in personality being the basis for homophily, so-called *personality homophily*, as observed through online social connections in Facebook. While personality and online activity has received much interest [8]–[13], insights into personality homophily have received less attention [5], [14]. In particular, from a network perspective, the role of links supporting personality homophily is not yet understood. Our aim in this paper is to examine presence of local structural features in online networks relative to personality homophily. Using the “MyPersonality” project dataset [15] we examine a considerable sample of the Facebook social network consisting of 313,699 unique nodes, 627,503 edges and an average clustering of 0.14. Motivated by hypotheses developed from the literature, we investigate the differences in likelihood of connection for particular personality traits and investigate the role and connectivity of individuals as a function of their personality. We also consider other network characteristic such as degree and transitivity, to further highlight the relationship between personality and network structure, and especially, network connectivity. This is explored while controlling for age and gender.

## II. RELATED WORK

The explosion of online social networks such as Facebook has allowed new ways in which relationships can be maintained, providing insight into fundamental human behavior [16]. From this perspective, personality has been shown as influential in online settings, especially in how people use social media [10], [11], [17]–[19] and the Internet [8], [9], [20] to the extent that it is possible to predict user’s personality traits based on their online interactions [21], [22]. Personality is measured in a number of different ways, ranging from self-report, to peer-report or observational methods. The Five-Factor Model of personality consists of 5 factors: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. It is one of the most widely used and validated personality models to date [23], [24]. Openness to experience refers to an individual’s curiosity and willingness to engage in new experiences, conscientiousness highlights how organized a person is, extraversion quantifies how sociable and outgoing someone is, while agreeableness relates to friendliness, and neuroticism to anxiety [23].

Previous literature already extensively covers the influence of individual personality on network characteristics, but the in-

fluence of personality homophily has not been well addressed. Regarding individual personality, not all effects appear to be robust across replications with different samples, methods and social networking platforms. Several studies have used network characteristics to predict personality [25] or have used Facebook data, such as status updates, for predictive models of personality [26], [27] or for developing an automatic classification system of personality traits based on network characteristics [28]. Other studies have specifically focused on the interplay between personality and network characteristics [29]–[32]. For example, the correlation between extraversion and network size has been repeatedly supported across different networking platforms (i.e., friends on Facebook: [11], [13], [26], [32]; online communities: [33] or followers on Twitter: [34]), while one Facebook study could not replicate this effect [17].

We develop hypotheses based on existing literature, initially considering relationships between individual personality traits and network characteristics, *with hypotheses italicised*.

#### A. Network size

Starting with extraversion and network size, it is prudent to confirm whether or not a positive correlation exists that has been evident in previous work [11], [13], [26], [32] (Hypothesis 1a). However we extend this by considering other personality facets, and *hypothesize that there is a basis for neuroticism to have a negative effect on network size (Hypothesis 1a)*. This arises from a communication study that identified neurotic users as unpopular interaction partners in an online discussion board [14], especially between fellow high scorers on the facet. Network size was also a significant feature to predict Neuroticism in Markovikj et al’s personality prediction model [26].

#### B. Effects of age and gender

Gender plays a wider role in social media use, with females and males using Facebook functions differently depending on their personality and gender [10], [11]. In a study on social media use and identity, extraverts were most likely to construe their offline self as their ‘real-me’, while neurotic users were more likely to construe their online self as their ‘real-me’ [8]. The effect for extraversion was significant for both males and females, while the effect for neurotic users was driven by female users [8].

Age is another demographic factor that might confound effects of personality. Personality is assumed to remain relatively stable over the lifespan, but some variation occurs. For example, neuroticism has been found to decrease with age in women, while agreeableness and conscientiousness have a tendency to increase [35], [36]. Another study found differences in all Big five facets over the lifespan [37]. We therefore take into account age as a possible confounding factor in our sample, as the MyPersonality sample is quite young, with a mean age of 24. In fact age might also influence network characteristics, as number of friends on Facebook

have been shown to decrease as age increases [32]. Additionally, age and gender have been found to be homophilous themselves [3], [38]. Individuals are more likely to befriend others of the same gender and who are of a similar age [3]. Furthermore, young people tend to expand their social networks, while older people tended to focus on maintaining fewer, but closer ties. Finally, younger people tend to form social connections with both same and opposite gender friends, especially during their reproductive active years. Older people tend to focus on maintaining same-gender ties instead [38]. Gender and age are therefore important control factors that should be taken into account when studying social networks online and offline, considering the gender differences that exist in how social media is used [11], the age-related differences in personality scores [36], and how gender affects fundamental network characteristics, such as transitivity [31]. We therefore hypothesize that differences exist in personality scores for both gender and age (Hypothesis 1b) *and have the potential to moderate personality homophily effects (Hypothesis 2d)*.

#### C. Local connectivity - Transitivity

Transitivity is an interesting connectivity feature in social networks, as it is an indication of the embeddedness of a node in its network. In particular, sociable individuals tend to be at the centre of large, loosely connected networks. This would explain why openness and extraversion have previously been negatively associated with transitivity [28], [30], [31]. A negative effect of extraversion on transitivity can possibly be explained through the associated higher degree of nodes while the negative effect of openness to experience on transitivity appears to be only significant for men [31]. This study had a smaller sample size (4305 nodes verses 313,669 nodes here) and therefore we reconsider whether openness to experience and extraversion are negatively related to transitivity for our case study (Hypothesis 1c). As in [31], we control for the possible effects of gender and degree, hypothesizing that *the effects of openness to experience on transitivity are moderated by gender, while effects of extraversion on transitivity are mediated by degree (Hypothesis 1d)*.

#### D. Local connectivity - Assortativity

Many socio-demographic characteristics have been found to be assortative (i.e., homophilous). The strongest homophily effects are observed for race and ethnicity, followed by age, religion, education, occupation and gender [3]. Additionally, other factors have also been identified as homophilous in online social networks, such as music taste and residential location [39]. Even psychological dispositions, such as happiness or loneliness, have been found to be homophilous in social networks [3], [40].

Notably fewer insights have emerged on personality homophily. This has been established for behavior in communication networks where users high on extraversion, agreeableness and openness prefer to communicate with similar others on online discussion boards [14]. Interestingly, these same personality facets have been found to be homophilous

in offline friendship networks as well [41]. Both studies also found that extraverts were more likely to initiate contact, while agreeable people were the most popular candidates for friendship requests [41] and as online interaction partners [14]. Openness to experience further proved to be homophilous in a spatial network as well [5]. People who scored high on openness to experience were likely to go to the same locations. This same study also found an interesting effect for conscientiousness, which is usually an overlooked facet in the personality and social media literature. People high in conscientiousness were also more likely to frequent the same venues as similar others [5]. An interesting effect was observed for neuroticism, which proved to be disassortative, in both communication networks [14] and location-based networks [5].

Based on these observations, *we hypothesize that Facebook nodes with similar scores in openness to experience, agreeableness, and extraversion are more strongly connected (Hypothesis 2a)*. Although there are considerable differences in sampling (size and generality), it is reasonable to expect some consistency with [14], [41]. We also consider an additional measure of connectedness: triangle closure. If similarity in personality indeed predicts connectedness between two nodes, we also expect nodes in closed triangles (in which all nodes are directly connect) to be more similar than in open triangles (in which two nodes are only indirectly connected through a broker). This additional measure will allow us to distinguish between triads of connected nodes (in closed triangles) and triads of indirectly connected nodes (in open triangles). We hypothesize that personality similarity is higher in closed triangles than for open triangles for openness, extraversion and agreeableness (Hypothesis 2b). Additionally, we will explore whether similarity in conscientiousness is related to triangle closure and strength of connectedness, although it must be noted that location-based and social networks are fundamentally different. We keep the hypothesis two-tailed because of the lack of prior support for this facet (Hypothesis 2c).

### E. Summary of Hypotheses

- 1) Network characteristics and demographics
  - a) extraversion is positively related to network size, and neuroticism is negatively related to network size
  - b) Personality scores differ between genders and across age
  - c) openness to experience and extraversion are negatively related to transitivity
  - d) effects of openness to experience on transitivity are moderated by gender, while effects of extraversion on transitivity are mediated by network size
- 2) Assortativity
  - a) nodes with similar scores in openness to experience, agreeableness, and extraversion are more strongly connected

- b) personality similarity is higher in closed triangles than in open triangles for openness, extraversion and agreeableness
- c) personality similarity differs for closed and open triangles, and varies with strength of connectedness for conscientiousness
- d) gender moderates the effect of personality similarity on connectedness

## III. METHODS

The data used in this paper was provided by collaboration with the MyPersonality project [15]. The nature of the data does not allow for any causal inferences to be made, but does allow for relationships between personality and social network structure to be uncovered. Two main datasets were used, the triad dataset and the ego-network dataset. The triad dataset contained information about triangles of friends from Facebook who had filled out a personality questionnaire and provided additional demographic (gender, age) and geographic information. The personality questionnaires used ranged from 20 to 100 items and were all derived from the International Personality Item Pool (IPIP) [42]. Each item was answered on a 5-point Likert scale and responses were averaged across all items to give a single score on each of the five facets for each participant. Together, the triads formed a network, denoted graph  $G$ , of 313,699 nodes and 627,503 edges, see table I for a breakdown of the characteristics of  $G$ .

TABLE I  
NETWORK CHARACTERISTIC VALUES FOR NETWORK  $G$

characteristic	value
number of nodes	313,699
number of edges	627,503
average clustering	.14
graph density	$3.18 * 10^{-6}$
graph transitivity	.15

The ego-network dataset contained information about ego networks and pre-computed values for node transitivity, network size, and node brokerage were of interest for our hypotheses. The intersection of the two datasets results in a sample consisting of 9,659 nodes. The variables considered for each node were the five personality scores from the personality questionnaire (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism), their connected alters' personality scores, their network size, brokerage, and transitivity. The full dataset of 313,699 cases will be referred to as the triad dataset, while the reduced dataset of  $N = 9,659$  will be referred to as the ego-networks dataset. The ego-networks dataset is used to answer Hypotheses 1c and 1d. The triad dataset of 313,699 was used to answer hypotheses 1b, and 2a to 2d. A breakdown of the definitions and equations for the different network characteristics used is provided in Section III-A.

### A. Ego-centric network characteristics

Ego-centric networks offer information about a central actor, ego, and all its connections, referred to as alters. The below definitions and equations apply to every single ego network from the MyPersonality ego-networks dataset [15]. The open-source statistics software *R* was used to compute most of the network characteristics [43]. The following definitions are applied.

- 1) **Network size** : total number of nodes in ego’s network, including ego.
- 2) **Transitivity**: this is the fraction of possible triangles that exist involving ego. Transitivity for ego node  $e$  is defined as:

$$\frac{2 * e\_triangles}{e\_degree * (e\_degree - 1)} \quad (1)$$

where  $e\_triangles$  denotes the number of triangles including  $e$  and  $e\_degree$  is the degree of  $e$ .

- 3) **Strength of connectedness**: This was calculated between two nodes as the relative number of common neighbours they have, normalised by their combined network size, defined as:

$$\frac{|N(u) \cap N(v)|}{|N(u)| + |N(v)|} \quad (2)$$

where  $u$  and  $v$  are two connected nodes in the network graph  $G$  (see Table I), and  $N$  refers to the node’s degree. Note that the degree here refers to the degree of the node in graph  $G$  rather than the actual network size of the node provided by the triad dataset.

- 4) **Triangle closure**: Triangle closure is a dichotomous variable, which takes the value of 0 for open triangles and of 1 for closed triangles. Open triangles are triads in the Facebook dataset that have a broker with two unconnected friends. Closed triangles have brokers with two directly connected friends.
- 5) **Personality difference score**: We used absolute differences to measure personality similarity between pairs of connected nodes. The equation below is applied to each triad in the dataset, for each facet separately, resulting in five similarity measures for each pair.  $F$  represents the facet, while  $u$  and  $v$  are any two connected nodes:

$$|F(u) - F(v)| \quad (3)$$

Note that a lower score on this measure represents higher similarity.

- 6) **personality variance**: We used variance to measure personality similarity in open and closed triangles. The equation below is applied to each triangle in the dataset, for each facet separately, resulting in five similarity measures for each triangle.  $x$  represents the facet score for nodes  $u$ ,  $v$ , and  $w$ , while  $\bar{x}$  represents their mean:

$$\frac{\sum(x - \bar{x})^2}{n - 1} \quad (4)$$

Note that a lower variance score represents higher similarity.

A high transitivity score is an indication of a better connected network. Strength of connectedness and personality variances measure the relative similarity through connectivity, while triangle closure further provides local connectivity information.

### B. Descriptive Statistics

Table II provides descriptive statistics on the personality scores of the full  $N = 313,699$  triad dataset. Missing data accounts for the variation in sample size between the facets. The means and standard deviations for each sample seem typical of an online internet population [36].

TABLE II  
DESCRIPTIVES FOR PERSONALITY SCORES IN THE SAMPLE

personality facet	N	minimum	maximum	mean	std
openness	313,519	0.00	5.00	3.76	0.74
conscientiousness	313,520	0.00	5.00	3.40	0.78
extraversion	313,521	0.00	5.00	3.55	0.85
agreeableness	313,521	0.00	5.00	3.47	0.76
neuroticism	280,100	0.00	5.00	2.73	0.83

Inter-correlations between facets were small, ranging from 0.01 to  $-0.36$ . The highest correlation between facet scores was between agreeableness and neuroticism. The distribution of personality scores for each facet can be found in Figure 1.

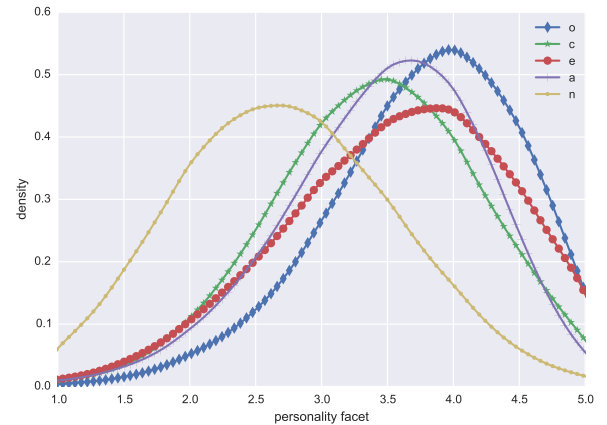


Fig. 1. Distribution of scores for each of the five personality facets.

## IV. RESULTS

We first tested the effects of age and gender on personality to uncover any biases that might arise from these two demographic variables. We then explored the effects of personality on different network characteristics in accordance with our hypotheses. We adopt a conservative alpha-value of .001 to account for the large number of tests carried out.

### A. Age

We explore the effects of age on personality, with a view to accounting for this as a possible mitigating factor in subsequent analyses. The triad dataset of  $N = 313,699$  cases was used. The Facebook sample was on average quite young, with a mean age of  $M = 24.6(7.06)$  and a median age of 22. The youngest person was 18 and the oldest 60. Skewness was 1.84(.006), while kurtosis was 3.65(.011). This could be because Facebook mainly attracts a young audience. It must also be noted that age information was absent for  $N = 130,979$  users, resulting in a reduced sample size of  $N = 182,690$ , which could further skew the results.

We tested the effect of all five personality facets on age separately first to identify significant effects. Based on the literature [37], we would expect all facets to vary with age at least somewhat. Among these, conscientiousness ( $B = 1.41(0.021), t = 68.76, p < .0001$ ) and agreeableness ( $B = 0.51(0.021), t = 23.89, p < .0001$ ) were significant. Surprisingly, openness ( $p = .78$ ), extraversion ( $p = .24$ ) and neuroticism ( $p = .67$ ) did not differ with age. Agreeableness no longer had a significant effect when controlling for conscientiousness, bearing in mind our conservative alpha value of .001; the final model can be found in Table III. Agreeableness and conscientiousness are only able to explain 2.5% of variation in age. Age increases by one year with every 1.4 increase in conscientiousness.

TABLE III  
EFFECT OF AGREEABLENESS AND CONSCIENTIOUSNESS SCORES ON AGE

dependent variables	B	SE	t	p-value
conscientiousness	1.4	0.022	64.42	< .0001
agreeableness	-0.05	0.022	2.15	.03

As age did not have an effect for most personality facets, it was not further considered as a control variable.

### B. Gender

Gender differences in social media usage and their interaction with personality have been studied in the literature [8], [44], but are not as often considered in studies on personality and social networks, with a few notable exceptions (e.g., [31]). We first explore whether any differences exist between male and female Facebook users, using the triad dataset. The non-parametric Welch  $t$ -test [45] was used to explore differences in personality between male and female users, since homogeneity of variance was violated for conscientiousness ( $p = .006$ ), extraversion ( $p < .0001$ ), agreeableness ( $p < .0001$ ), and neuroticism ( $p < .0001$ ). The Welch's  $t$ -test provides a robust test of the equality of means for large, non-normally distributed samples with unequal variances across populations with unequal sample sizes [45]. The results can be found in table IV. Openness to experience, conscientiousness, agreeableness, and neuroticism all significantly differ between males and females. It must be noted, however, that effect sizes are extremely low for these results, varying between  $r = .002$  and  $r = .003$ .

This means that, although a consistent difference between both genders was found, as expected from Hypothesis 1b, these differences are minimal, as seen in their means (Table IV).

TABLE IV  
DIFFERENCES IN PERSONALITY BETWEEN MALE AND FEMALE FACEBOOK USERS

personality facet	t	p-value	N_F	N_M	mean F (SD)	mean M (SD)
openness	17.69	< .0001	5376	3285	3.9(0.65)	4.0(0.63)
conscientiousness	23.67	< .0001	5376	3285	3.5(0.71)	3.4(0.72)
extraversion	0.70	.40	5376	3285	3.7(0.81)	3.7(0.83)
agreeableness	10.77	.001	5376	3285	3.6(0.68)	3.6(0.72)
neuroticism	238.09	< .0001	5348	3251	2.8(0.81)	2.6(0.81)

### C. Network size

To answer Hypothesis 1a on network size, the ego-networks dataset of  $N = 9,659$  cases was used. Ego networks ranged from 22 to 1997 alters, with a mean score of 338 ( $SD = 299.7$ ). Extraversion indeed had a positive effect on network size, with  $B = 76.00(3.69), t(9457) = 20.57, p < .0001$ . This means that for each point-increase on the extraversion scale, a user had on average 76 more friends. As predicted by Hypothesis 1a, neuroticism indeed had a negative effect on network size, with  $B = -37.20(3.74), t(9395) = -9.95, p < .0001$ . Thus for every point-decrease on the neuroticism scale, a user loses 37 friends on average. These results confirm previous findings in the literature concerning extraversion, neuroticism and network size.

### D. Transitivity

Finally, the ego-network dataset was used for our hypotheses on transitivity, 1c and 1d. Transitivity scores ranged from 0.0006 to 0.91, with a mean score of 0.13( $SD = 0.16$ ). We found a significant negative effect of openness to experience on transitivity,  $B = -0.011(0.003), t(9458) = -4.19, p < .00001$ , as well as for extraversion,  $B = -0.019(0.002), t(9458) = -9.72, p < .0001$ . This is in line with Hypothesis 1c. However, we also predicted that network size would mediate the effect of extraversion on transitivity, see hypothesis 1d. This was not the case in our sample, however. Extraversion remained a significant predictor of transitivity, even when controlling for network size, with  $B = -.008(.002), t(9457) = -4.27, p < .0001$ .

We further expected gender to mediate the effect of openness on transitivity (Hypothesis 1d). We once again opt for a Welch's  $t$ -test for its robustness. Missing data for gender accounted for the variations in sample size. Transitivity indeed varied between both genders, with  $t(6505) = 12.53, p < .0001$ . Men ( $M = .14, SD = .17$ ), on average, had networks with higher transitivity than women ( $M = .13, SD = .15$ ). When separated by gender, openness had a significant negative effect on transitivity for males ( $B = -.023(.005), t(5375) = -5.07, p < .0001$ ), but not for females ( $p = .06$ ). This confirms previous findings that observed a significant relationship between low transitivity and openness for males only [31] (Hypothesis 1d).

### E. Assortativity

Finally, we consider the influence of personality homophily on how inter-connected nodes in a social network are, as considered in Hypothesis 2a. Personality similarity was calculated for connected pairs in network graph  $G$  ( $N = 1,048,575$ ), using Equation 3. Pairs for which either facet score was missing were not included in the final analysis. We refer to this score as the *difference score*, since a higher value corresponds to a bigger difference between the two nodes in terms of personality.

We base our strength of connectedness measure on the edges in network graph  $G$ , created using the triad data from the MyPersonality project [15], see Section III. Strength of connectedness ranged from 0.0 to 0.45, with a mean of 0.06 ( $SD = 0.077$ ) and an  $N = 1,048,575$ . Table V details the results of the effect of personality homophily on strength of connectedness.

TABLE V  
EFFECTS OF PERSONALITY SIMILARITY ON CONNECTEDNESS

personality facet	$\beta$	$t$	$p - value$
agreeableness	-.008	-9.47	< .0001
conscientiousness	-.011	-11.34	< .0001
extraversion	-.010	-10.23	< .0001

Connectedness decreased as difference scores on agreeableness, conscientiousness and extraversion scores increased. This means that, as expected, personality similarity was higher among better connected nodes for agreeableness and extraversion (Hypothesis 2a). We also expected this effect for openness, but this could not be confirmed in our sample. On the other hand, we found an effect for conscientiousness. This effect is perhaps not as surprising if we consider that similarity in conscientiousness had a positive effect on location homophily in a previous study [5].

### F. Interaction with gender

Since we found a significant effect of gender for most personality facets, (Section IV-B), we also consider whether gender would moderate the effect of personality similarity on connectedness (Hypothesis 2d). We first checked whether there were gender differences in regard to connectedness. Since connectedness scores are computed pair-wise, we separated our data according to the pairs' gender composition. Pairs were male-male ( $N = 200,974$ ), female-female ( $N = 346,762$ ), or mixed ( $N = 466,697$ ). There was a significant difference in connectedness scores between the different pairs,  $t(1,520565) = 476.59$ ,  $p < .0001$ . Connectedness was strongest among male-male pairs ( $M = 0.062$ ,  $SD = 0.078$ ), followed by female-female pairs ( $M = 0.057$ ,  $SD = 0.076$ ), and finally mixed pairs ( $M = 0.056$ ,  $SD = 0.076$ ). Since there was no gender difference for extraversion, we only considered whether there was a difference in the observed negative effect of personality variances in agreeableness and conscientiousness on connectedness (Hypothesis 2d). The results can be found in Table VI.

TABLE VI  
EFFECTS OF PERSONALITY SIMILARITY ON CONNECTEDNESS BY GENDER PAIRS

gender pairs	$\beta$	$t$	$p - value$
<b>conscientiousness</b>			
male-male	-0.014	-6.37	< .0001
female-female	-0.011	-6.76	< .0001
male-female	-0.013	-8.83	< .0001
<b>agreeableness</b>			
male-male	-0.019	-8.50	< .0001
female-female	-0.002	-1.22	.221
male-female	-0.014	-9.85	< .0001

Of interest is the observation that for agreeableness, personality similarity only had a significant effect for male-male and mixed pairs on connectedness, but not for female-female pairs (Table VI) with male-male pairs having the stronger negative effect ( $\beta = -0.019$ ) compared to mixed pairs ( $\beta = -0.014$ ). For conscientiousness, although the strength of the effect varied slightly between pairs, all of them had a negative effect on connectedness (Table VI). In accordance with Hypothesis 2d, gender had a moderating effect on the relationship between personality similarity and strength of connectedness for agreeableness.

### G. Triangle closure

We further hypothesized that personality homophily would be stronger among closed triangles, compared to open triangles (Hypothesis 2b), see Definition 4. Personality similarity was measured in terms of variance, see Equation 4. We carried out Welch's  $t$ -test to uncover any significant differences between open and closed triangles for all five facets. We found a significant effect for extraversion, agreeableness, and for openness (Table VII). Additionally, we also found an effect for conscientiousness. For all significant facets, the personality variance scores were higher among open triangles ( $N = 989,432$ ), as opposed to closed ones ( $N = 59,143$ ) (Table VII), although actual differences in means were small.

TABLE VII  
DIFFERENCES IN PERSONALITY SIMILARITY FOR CLOSED AND OPEN TRIANGLES

personality facet	$t$	$p - value$	Mean open	Mean closed
openness	36.41	< .0001	0.34(0.36)	0.33(0.35)
conscientiousness	76.15	< .0001	0.39(0.39)	0.38(0.38)
extraversion	108.41	< .0001	0.44(0.45)	0.42(0.44)
agreeableness	79.17	< .0001	0.38(0.39)	0.37(0.38)

As expected from Hypothesis 2b, personality similarity was higher among closed triangles than among open triangles for openness, extraversion, and agreeableness, as well as for conscientiousness.

## V. CONCLUSION

Online network characteristics and personality appear to be intrinsically linked. The personality facets that emerge as

related to the shape of our social networks are extraversion, neuroticism, openness to experience and to a certain extent, conscientiousness and agreeableness. As expected, network size was positively related to extraversion, and negatively related to neuroticism. Extraversion and openness to experience were both negatively related to transitivity. We also replicated a gender-dependent effect for openness to experience on transitivity [31]: openness had a significant negative effect on transitivity for males, but not for females. This confirms that extraversion and openness are related to networks low in transitivity, in which extraverted and open nodes act as bridges between loosely connected alters. In conjunction with the effects found for network size, this translates to large, loosely connected networks for extraverts.

Furthermore, we studied connectedness in two different ways: first through strength of connection between two nodes, and secondly, through triangle closure. More strongly connected nodes had higher similarity on the extraversion facet. This effect could also be confirmed for agreeableness and conscientiousness, but not for openness to experience. This homophilous effect might not have been found for open people, because they are more open to connect with both similar and dissimilar others. This is highlighted by their transitive networks, which demonstrate a diverse and loosely connected network. Of further interest is the moderating role of gender on the relationship between personality similarity and connectedness. In general, connectedness was strongest for male-only pairs, followed by female-only, and then mixed pairs. For agreeableness, the effect of personality similarity on connectedness was strongest among male-only pairs, but did not have an effect for female-only pairs. On the other hand, personality similarity continued to have the same positive effect for conscientiousness, regardless of gender. Finally, we also identified that similarity in openness, extraversion, agreeableness, and conscientiousness was higher in closed triangles compared to open ones. This suggests that personality homophily indeed plays a role in structuring connections in ego-centric networks for most facets. It must be noted, however, that effects in regard to triangle closure and personality homophily were small, with only a few .01 differences between closed and open triangles on the personality variance scores. The large MyPersonality dataset of 313,699 nodes allowed us to tease out an effect that smaller samples could have missed. Future research will have to determine whether this personality homophily effect is worth investigating further.

An alternative explanation for the observed personality homophily effects might be that some personalities are simply more attractive than others: people enjoy the company of friendly and sociable friends. This is especially true for agreeable users, who are the most popular recipients of offline friend requests [41] and online interactions on communication boards [14]. Extraverts on the other hand, are more likely to initiate friendships [41] and reach out to people online [14].

However, we also found personality homophily effects for conscientiousness, which cannot be explained through the attractive personality hypothesis. Conscientiousness had pre-

viously been identified as homophilous in a location-based network [5], but is usually overlooked as an influential factor in online and offline social networks. Future research in online social networks should nonetheless explore the attractive personality hypothesis, potentially uncovering the most popular personality combinations among online and offline friendship pairs. These results provide further evidence that observed social network features are potentially linked to embedded human characteristics.

Whether personality drives social network structure or vice versa remains unclear, however. A potential explanation could be that personality, being an inherent characteristic, plays a role in determining one's initial network position, which, in turn, is responsible for reinforcing specific personality traits. For example, neurotic people might be in broker positions because of their personality, or become more neurotic because broker positions are stressful [30]. The fact that most facets vary with age supports the idea that personality is malleable [36], although it must also be noted that personality varies only slightly over people's lifespan.

In conclusion, personality and social network position could be reinforcing each other, and further research needs to explore the possible causal links between social network position and personality. We have, however, uncovered valuable new insights in personality homophily in online social networks. Agreeableness and conscientiousness emerge as homophilous facets, both in relation to strength of connection and triangle closure, as predicted. Additionally, we have replicated some fundamental relationships between extraversion, neuroticism and network size, as well as between extraversion, openness and transitivity. Gender also played a role: the negative relationship between openness and transitivity was only significant for males, and the positive relationship between connectedness and agreeableness was not significant for female-only pairs. This supports further work concerning the conjunction of gender and personality homophily in social network analysis.

#### ACKNOWLEDGMENT

The authors would like to thank David Stillwell and Michal Kosinski for their collaboration concerning the MyPersonality dataset.

#### REFERENCES

- [1] M. Newman and J. Park, "Why social networks are different from other types of networks," *Physical Review E*, vol. 68, no. 036122, pp. 1–8, 2003.
- [2] G. Kossinets and D. J. Watts, "Origins of homophily in an evolving social network," *American Journal of Sociology*, vol. 115, no. 2, pp. 405–450, 2009.
- [3] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual Review of Sociology*, vol. 27, pp. 415–444, 2001.
- [4] G. B. Colombo, M. J. Chorley, M. J. Williams, S. M. Allen, and R. M. Whitaker, "You are where you eat: Foursquare checkins as indicators of human mobility and behaviour," in *Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 2012 *IEEE International Conference on*. IEEE, 2012, pp. 217–222.
- [5] N. Noë, R. M. Whitaker, M. J. Chorley, and T. V. Pollet, "Birds of a feather locate together? foursquare checkins and personality homophily," *Computers in Human Behavior*, vol. 58, pp. 343 – 353, 2016.



- [6] R. A. Sherman, C. S. Nave, and D. C. Funder, "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, 2012.
- [7] M. Snyder, "The influence of individuals on situations: Implications for understanding the links between personality and social behavior," *Journal of Personality*, vol. 51, no. 3, pp. 497–516, 1983.
- [8] Y. a. Hamburger and E. Ben-Artzi, "Relationship between extraversion and neuroticism and the different uses of the internet," *Computers in Human Behavior*, vol. 16, no. 4, pp. 441–449, 2000.
- [9] R. N. Landers and J. W. Lounsbury, "An investigation of big five and narrow personality traits in relation to internet usage," *Computers in Human Behavior*, vol. 22, no. 2, pp. 283–293, 2006.
- [10] T. Correa, A. W. Hinsley, and H. G. de Zúñiga, "Who interacts on the web?: The intersection of users' personality and social media use," *Computers in Human Behavior*, vol. 26, no. 2, pp. 247–253, 2009.
- [11] Y. Amichai-Hamburger and G. Vinitzky, "Social network use and personality," *Computers in Human Behavior*, vol. 26, no. 6, pp. 1289–1295, 2010.
- [12] S. Adali and J. Golbeck, "Predicting personality with social behavior," in *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2012, pp. 302–309.
- [13] J. Shen, O. Brdiczka, and J. Liu, "Computers in human behavior a study of facebook behavior : What does it tell about your neuroticism and extraversion ?" *Computers in Human Behavior*, vol. 45, pp. 32–38, 2015.
- [14] J. M. Balmaceda, S. Schiaffino, and D. Godoy, "How do personality traits affect communication among users in online social networks?" *Online Information Review*, vol. 38, pp. 136–153, 2013.
- [15] M. Kosinski., S. M. S. Gosling, V. Popov, and D. Stillwell, "Facebook as a social science research tool: Opportunities, challenges, ethical considerations and practical guidelines," *American Psychologist*, 2015.
- [16] R. M. Whitaker, M. Chorley, and S. M. Allen, "New frontiers for crowdsourcing: The extended mind," in *System Sciences (HICSS), 2015 48th Hawaii International Conference on*. IEEE, 2015, pp. 1635–1644.
- [17] C. Ross, E. S. Orr, M. Sisc, J. M. Arseneault, M. G. Simmering, and R. R. Orr, "Personality and motivations associated with facebook use," *Computers in Human Behavior*, vol. 25, no. 2, pp. 578–586, 2009.
- [18] M. Michikyan, K. Subrahmanyam, and J. Dennis, "Can you tell who i am? neuroticism, extraversion, and online self-presentation among young adults," *Computers in Human Behavior*, vol. 33, pp. 179–183, 2014.
- [19] M. J. Chorley, R. M. Whitaker, and S. M. Allen, "Personality and location-based social networks," *Computers in Human Behavior*, vol. 46, pp. 45–56, 2015.
- [20] T. Amiel and S. L. Sargent, "Individual differences in internet usage motives," *Computers in Human Behavior*, vol. 20, no. 6, pp. 711–726, 2004.
- [21] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *Proceedings of the National Academy of Sciences*, vol. 110, no. 15, pp. 5802–5805, 2013.
- [22] M. Kosinski, Y. Bachrach, P. Kohli, D. Stillwell, and T. Graepel, "Manifestations of user personality in website choice and behaviour on online social networks," *Machine Learning*, vol. 95, no. 3, pp. 357–380, 2014.
- [23] R. McCrae and P. Costa, "Validation of the five-factor model of personality across instruments and observers," *Journal of personality and social psychology*, vol. 52, no. 1, pp. 81–90, 1987.
- [24] J. M. Digman, "Personality structure: Emergence of the five-factor model," *Annual Review of Psychology*, vol. 41, pp. 417–440, 1990.
- [25] J. Staiano, B. Lepri, N. Aharony, F. Pianesi, N. Sebe, and A. Pentland, "Friends don't lie: Inferring personality traits from social network structure," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ser. UbiComp '12. ACM, 2012, pp. 321–330.
- [26] D. Markovikj, S. Gievska, M. Kosinski, and D. Stillwell, "Mining facebook data for predictive personality modeling," in *Proceedings of the 7th international AAAI conference on Weblogs and Social Media (ICWSM 2013)*, Boston, MA, USA, 2013.
- [27] G. Farnadi, S. Zoghbi, M.-F. Moens, and M. De Cock, "Recognising personality traits using facebook status updates," in *Proceedings of the workshop on computational personality recognition (WCPRI3) at the 7th international AAAI conference on weblogs and social media (ICWSM13)*. AAAI, 2013.
- [28] B. Lepri, J. Staiano, E. Shmueli, F. Pianesi, and A. Pentland, "The role of personality in shaping social networks and mediating behavioral change," *User Modeling and User-Adapted Interaction*, pp. 1–33, 2016.
- [29] K. J. Klein, B.-c. Lim, J. L. Saltz, and D. M. Mayer, "How do they get there ? an examination of the antecedents of centrality in team networks," *Academy of Management Journal*, vol. 47, no. 6, pp. 952–963, 2010.
- [30] Y. Kalish and G. Robins, "Psychological predispositions and network structure: The relationship between individual predispositions, structural holes and network closure," *Social Networks*, vol. 28, no. 1, pp. 56–84, 2006.
- [31] J.-E. Lönnqvist, J. V. Itkonen, M. Verkasalo, and P. Poutvaara, "The five-factor model of personality and degree and transitivity of facebook social networks," *Journal of Research in Personality*, vol. 50, pp. 98–101, 2014.
- [32] A. Friggeri, R. Lambiotte, M. Kosinski, and E. Fleury, "Psychological aspects of social communities," in *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom)*. IEEE, 2012, pp. 195–202.
- [33] J. Schrammel, C. Köffel, and M. Tscheligi, "Personality traits , usage patterns and information disclosure in online communities," in *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*. HCI, 2009, pp. 169–174.
- [34] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft, "Our twitter profiles, our selves: Predicting personality with twitter," in *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*. IEEE, 2011, pp. 180–185.
- [35] R. R. McCrae and P. T. Costa Jr, "A five-factor theory of personality," *Handbook of personality: Theory and research*, vol. 2, pp. 139–153, 1999.
- [36] S. Srivastava, O. P. John, S. D. Gosling, and J. Potter, "Development of personality in early and middle adulthood: set like plaster or persistent change?" *Journal of Personality and Social Psychology*, vol. 84, no. 5, p. 1041, 2003.
- [37] B. W. Roberts, K. E. Walton, and W. Viechtbauer, "Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies." *Psychological bulletin*, vol. 132, no. 1, p. 1, 2006.
- [38] Y. Dong, Y. Yang, J. Tang, and N. V. Chawla, "Inferring User Demographics and Social Strategies in Mobile Social Networks," in *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2014, pp. 15–24.
- [39] D. Hristova, M. Musolesi, and C. Mascolo, "Keep your friends close and your facebook friends closer : A multiplex network approach to the analysis of offline and online social ties," *International Conference on Web and Social Media*, pp. 206–215, 2014.
- [40] J. Bollen, B. Goncalves, G. Ruan, and H. Mao, "Happiness is assortative in online social networks." *Artificial life*, vol. 17, pp. 237–251, 2011.
- [41] M. Selfhout, W. Burk, S. Branje, J. Denissen, M. van Aken, and W. Meeus, "Emerging late adolescent friendship networks and big five personality traits: A social network approach." *Journal of Personality*, vol. 78, no. 2, pp. 509–538, 2010.
- [42] L. R. Goldberg, J. A. Johnson, H. W. Eber, R. Hogan, M. C. Ashton, C. R. Cloninger, and H. G. Gough, "The international personality item pool and the future of public-domain personality measures," *Journal of Research in personality*, vol. 40, no. 1, pp. 84–96, 2006.
- [43] R. C. Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2015. [Online]. Available: <https://www.R-project.org/>
- [44] Y. Amichai-Hamburger, G. Wainapel, and S. Fox, "On the internet no one knows i'm an introvert: extroversion, neuroticism, and internet interaction." *Cyberpsychology & behavior*, vol. 5, no. 2, pp. 125–128, 2002.
- [45] G. D. Ruxton, "The unequal variance t-test is an underused alternative to student's t-test and the mann-whitney u test," *Behavioral Ecology*, vol. 17, no. 4, pp. 688–690, 2006.