Computers in Human Behavior 58 (2016) 343–353

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Birds of a feather locate together? Foursquare checkins and personality homophily

Nyala Noë [a], Roger M. Whitaker [a, *], Martin J. Chorley [a], Thomas V. Pollet [b]

[a] School of Computer Science and Informatics, Cardiff University, 5 The Parade, CF24 3AA Cardiff, UK
[b] Dept. of Social and Organizational Psychology, VU University Amsterdam, van de Boechorststraat, 1, 1081BT Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 16 September 2015
Received in revised form
6 January 2016
Accepted 7 January 2016
Available online 21 January 2016

Keywords:
Spatial homophily
Personality
Location-based social network
Foursquare

ABSTRACT

In this paper we consider whether people with similar personality traits have a preference for common locations. Due to the difficulty in tracking and categorising the places that individuals choose to visit, this is largely unexplored. However, the recent popularity of location-based social networks (LBSNs) provides a means to gain new insight into this question through checkins - records that are made by LBSN users of their presence at specific street level locations. A web-based participatory survey was used to collect the personality traits and checkin behaviour of 174 anonymous users, who, through their common check-ins, formed a network with 5373 edges and an approximate edge density of 35%. We assess the degree of overlap in personality traits for users visiting common locations, as detected by user checkins. We find that people with similar high levels of conscientiousness, openness or agreeableness tended to have checked-in locations in common. The findings for extraverts were unexpected in that they did not provide evidence of individuals assorting at the same locations, contrary to predictions. Individuals high in neuroticism were in line with expectations, they did not tend to have locations in common. Unanticipated results concerning disagreeableness are of particular interest and suggest that different venue types and distinctive characteristics may act as attractors for people with particularly selective tendencies. These findings have important implications for decision-making and location.

1. Introduction

It is well-recognised that homophily, the attraction of individuals with similar traits to one another, is a widely occurring human disposition (McPherson, Smith-Lovin, & Cook, 2001). With the advent of the Internet and the popularity of social networking, it has become possible to understand this concept through the electronic ties that individuals choose to make with each other, leading to a wide range of insights from large electronic data sources. Despite these recent advances, relatively little is known about the manifestation of homophily in a physical context, thus the *extent to which similar people have a preference for visiting the same places* is an important question to ask. Unfortunately, a significant barrier to answering this question has been convenient data collection on a large scale, which until recently has been challenging to accomplish without access to dedicated location tracking equipment. However, the recent advent of smartphones and location-based social networks (LBSNs) allows new progress to be made. Location-based social networks run on a smartphone as a location-aware application, enabling a user to log their presence at a physical location (referred to as a checkin), which is shared across an online social network in real time. The analysis of checkins thus provides insight into the places that individuals publicly associate with.

Many socio-demographic, behavioural and intra-personal factors (McPherson et al., 2001) can potentially characterise aspects of similarity between individuals. For decisions related to human spatial activity, the most fundamental characteristics are arguably the personality traits, given that these are relatively persistent dispositions, thereby broadly framing an individual's outlook and potential approach to activity, interaction and behaviour. Trait-theorists argue that this is supported by evidence of personality trait correlation with wide-ranging human activities, ranging from consumer marketing (e.g., Kassarjian (1971)) through to organisational behaviour (e.g., Hough and Oswald (2008)) and individual tastes (e.g., Rawlings and Ciancarelli (1997)). The boundaries of scenarios where personal activities are congruent to personality

* Corresponding author.
E-mail address: WhitakerRM@cardiff.ac.uk (R.M. Whitaker).

http://dx.doi.org/10.1016/j.chb.2016.01.009

traits have been explored in Sherman, Nave, and Funder (2012), with findings that effectively characterise individual freedom consistent with choice in social, consumer-related and location-based decisions. Consequently, we focus on individual preferences regarding assortment.

To explore user similarity in location-based activity, we use data collected by a recently introduced experimental platform (Chorley, Whitaker, & Allen, 2015), which has been designed to allow users of the Foursquare[1] location-based social network to participate in anonymous collection of their checkins and personality profile in return for visualisation of their own personality relative to others at locations where common checkins are made. This novel approach naturally incentivises participation and has allowed viral participant recruitment "in-the-wild" to be accomplished, resulting in data from 174 anonymous participants who have collectively checked in 487,398 times at 119,746 venues. Taking the volume, diversity and broad categorisation of venues visited as variables, the first examination of human mobility behaviour at street level, in relation to human personality (Chorley et al., 2015) identified a number of interesting correlations. In particular, conscientiousness positively correlated with the number of venues visited, openness positively correlated with checkins at both sociable and popular venues, and neuroticism negatively correlated with the number of sociable venues visited.

In this paper we focus on the extent of overlap in personality for common place-based visits, using checkins as the observed signal. As far as we are aware this is the first investigation of personality homophily based on spatial activity.

1.1. Location-based social networks

LBSNs are an interesting hybrid technology that extends online social networking into the physical "real" world. Facebook, Foursquare, and Google + are, to date, the most commonly used LBSNs, with Foursquare recently reorganising its business to provide the checkin facility through a complementary application called *Swarm*. Users of LBSNs require location-aware smartphones and internet connectivity in order to record their presence at a location, referred to as a *checkin*. This activity triggers a notification to friends within the associated online social network. Rather than a checkin being recorded solely as a geographical reference (e.g., longitude and latitude or street address), it is usually delivered with a meaningful semantic representation, such as a named place at street level (e.g., the name of a coffee shop and its approximate location). Places that are explicitly registered through the LBSN in this way are called *venues*. Many LBSNs operate extensible taxonomies of venues that are populated by users, and these have become widespread for cities and popular areas on a global basis.

Checkins give particular insight into the venues that an individual chooses to record as important, interesting or relevant. However in some LBSNs such as Facebook and Google+, the checkin functionality has been introduced as a secondary function, built on top of other online social networking functionality. The Foursquare LBSN is different in this regard, originating with checkins as its primary function, and with limited secondary content provision. These factors, combined with a rich API[2] on which third party applications can be developed, have led to Foursquare being a popular basis for academic insight to a range of human behaviours. Primarily these have concerned physical activity, such as relating to patterns made by users (e.g., Noulas, Scellato, Mascolo, and Pontil (2011)) and with a high degree of location data aggregation. This has led to insights into the effect of social relationships and routine on spatial behaviour for example (Cho, Myers, & Leskovec, 2011).

1.2. User motivation

A LBSN users' checkin behaviour may be motivated by several factors, such as establishing a social connection with friends, discovering new places to visit, keeping track of already visited places, fighting boredom and gamification (Lindqvist, Cranshaw, Wiese, Hong, & Zimmerman, 2011). LBSNs allow users to select certain locations as a means of self-presentation, referred to as the *spatial self* (Schwartz & Halegoua, 2014). This is frequently consistent with other forms of online self-presentation and can involve venue avoidance to counter associations with perceived negative places (Lindqvist et al., 2011). Users have been found to control the volume of checkins in different ways, avoiding spamming their social networks with too many checkins and giving thought to self-presentation (Schwartz & Halegoua, 2014). Different levels of consistency (i.e., venue selection) have been reported. Some users consistently check in to any place they visit, while others select their checked in locations more carefully, based on how interesting or deserving they deem the place to be (Lindqvist et al., 2011). Audience management is a further aspect of user behaviour in LBSNs, with users sharing different checkins with different groups of friends and acquaintances. In some cases, interesting checkins, meaning checkins at unusual or new venues, were reserved for Twitter and Facebook, while more general checkins were shared with friends (Cramer, Rost, & Holmquist, 2011).

These factors mean that the checkin is a potentially noisy signal with varying purposes between individuals. To some degree, checkins represent a unique footprint which is characteristic of the individual user, and are worthy of investigation as a means to understand human behaviour. However, limited existing studies have addressed the role of checkins in relation to individual differences such as personality. Wang, Pedreschi, Song, Giannotti, and Barabasi (2011) have considered the personality characteristics that correlate with individuals sharing checkins in Facebook, and in Chorley et al. (2015), the personality traits of individual users have been correlated with observed checkins.

1.3. Personality

In psychology, trait theory (Allport, 1966) suggests that humans have underlying stable characteristics of biological origin, framing how situations are individually considered and approached. These traits, broadly referred to as *personality facets*, can influence subconscious human behaviour. As such, there has been considerable research exploring the relationships between diverse human activity and personality. Situations where personality facets are particularly influential to human behaviour have been considered by Sherman et al. (2012). These behaviours have been broadly categorised as freedom of self-expression, social interaction, lack of a-priori structure and an opportunity to engage in competencies. Aspects of both online and offline human activity fall into these categories, including checkins and spatial behaviour.

From lexical origins, dimensions capturing personality have progressively emerged since the 1930's, with the NEO Personality Inventory being developed by Costa and McCrae (1985) and validated by McCrae and Costa (1987) in the 1980's. The concept of the Big Five and the NEO Personality Inventory has been updated and revised throughout the years (Digman, 1990), with the revised NEO-PI-3 published by McCrae, Costa, and Martin (2005). Although

[1] Foursquare have recently reorganised their business model and checkins are now made through a dedicated application called Swarm: http://www.swarmapp.com.

[2] API stands for Application Program Interface.

not without considerable debate (e.g., Block (2001)), the five factor model has become a widespread model of personality (Costa & McCrae, 1985; Goldberg, 1990), with its dimensions capturing Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism. An alternative model to the Big 5 is the HEXACO model (Lee & Ashton, 2004); which of these models captures personality dimensions more accurately and more universally is still an ongoing debate (Ashton & Lee, 2008; Lee, Ogunfowora, & Ashton, 2005). In terms of correlation with online activity, research in this area has addressed relationships between personality, Internet and social network usage, primarily concerning Facebook (e.g., Amichai-Hamburger and Vinitzky (2010); Ross et al. (2009)), Twitter (e.g., Quercia, Kosinski, Stillwell, and Crowcroft (2011)), and to a lesser extent concerning LBSNs, such as Foursquare (e.g., Chorley et al., 2015).

1.4. Homophily and personality

Homophily, the attraction of similar nodes in a network, is a fundamental organizing principle in social networks. Homophily can predict interests and characteristics of users in a network, based on the characteristics and interests of their neighbours (Kossinets & Watts, 2009). This is of value to many Internet services: for example websites such as Amazon and Netflix apply similarity of buying and watching patterns to predict and recommend future consumption (Ziegler & Golbeck, 2007).

Homophily has important structural implications for social networks. Strongly connected users tend to be more similar than weaker connected users (McPherson et al., 2001), while nodes in small communities may be more prone to assort than larger ones (Launay & Dunbar, 2015), in line with the small world effect Milgram (1967) and the prevalence of hubs (high degree nodes) and a low mean shortest path (Newman, 2000). In the online domain, homophily has also been observed in directed networks such as Twitter (Bollen, Gonçalves, Ruan, & Mao, 2011), where psychological dispositions have been investigated as the basis for homophily. Loneliness has for example been shown to be assortative (McPherson et al., 2001). Furthermore, positive Twitter users were most likely to follow and be followed by other positive users. Negative users assorted in the Twitter network, and also tended to follow and be followed by fellow negative users (Bollen et al., 2011).

Considerably less attention has yet been paid to personality and homophily and as compared to other measures of similarity (e.g., political affiliation or friendship), assessing personality requires a greater levels of participant interaction. However personality-based homophily has been found to be a predictor for connections in a social network formed among first-year university students (Selfhout et al., 2010). Students tended to befriend others with similar levels of extraversion, agreeableness and openness to experience.

1.5. The emergence of spatial homophily

Given that personality is a potential predictor for behaviour and attitudes in a range of situations (Goldberg, 1990), it is possible that personality-based homophily may support the attraction of like individuals for a wide range of scenarios (Sherman et al., 2012). One conceivable scenario where personality may have a homophilic effect relates to the type of location that individuals choose to visit. So-called *spatial homophily* has only recently been considered (Pelechrinis & Krishnamurthy, 2015; Zhang & Pelechrinis, 2014), and captures the attraction of individuals, who are in some sense similar, to common locations (Colombo et al., 2012; Williams et al., 2012).

Recent work (Schwartz & Halegoua, 2014) has proposed that people may use the places that they visit to build an online representation of themselves. Hence, potentially the characteristics of people can be derived from the locations that they choose to affiliate with through checkins. Graham and Gosling (2011) demonstrated that impressions of a place and its visitors could systematically be derived from the Foursquare user profiles of its visitors. Participants were able to accurately predict the personality of a typical visitor of a specific location, based on the Foursquare profiles of actual visitors ($ICC = .69$). Ambiance ($ICC = .32$) and typical activities of visitors ($ICC = .33$) of a specific place had far lower agreement. On a larger scale, Cranshaw, Schwartz, Hong, and Sadeh (2012) demonstrated that a city's character could be derived from the mobility patterns of its residents. Similar people tended to visit a network of venues within a neighbourhood or region of a city that form a comprehensive whole, rather than individual locations (Cranshaw et al., 2012).

Personality has also been related to spatial location and to spatial homophily. For example different neighbourhoods in London have different personality profiles (Jokela, Bleidorn, Lamb, Gosling, & Rentfrow, 2015). Here it was identified that the centre of London has a higher prevalence of high openness to experience and low agreeableness, while neighbourhoods further away from the city centre are low in neuroticism and high in conscientiousness. Jokela et al. (2015) also showed that personality mitigated the effect of neighbourhood on life satisfaction. More specifically, open individuals were the happiest in neighbourhoods with a high number of fellow open people. This suggests that personality-homophily can have important implications for life satisfaction in specific London neighbourhoods (Jokela et al., 2015). Personality not only characterizes specific neighbourhoods, but evidence has been presented that it may characterize entire countries. For example, Rentfrow, Jokela, and Lamb (2015) indicated that within the United Kingdom, Scotland was agreeable and emotionally stable while Wales was, on average, introverted and neurotic.

The places considered through spatial homophily need not be restricted by one's residential neighbourhood or region, however. For example, Joseph, Tan, and Carley (2012) identified clusters of individuals, such as gym enthusiasts or art enthusiasts, who had similar interests in venues consistent with their Foursquare checkins. Interestingly, the venues visited by individuals within the same cluster were spread throughout the city, rather than being confined to a particular neighbourhood. Specific types of locations, rather than general geographic areas, can therefore be places where people with similar personality traits assort. This contributes to the motivation for our investigation.

2. Research objective and hypotheses

Our focus concerns observing signals of homophily through common LBSN checkins and similarity of personality. The extent of the effect of individual differences in personality on the similarity of locations visited remains unknown. Developing further understanding of this issue is our objective, while acknowledging that checkin activity represents only a subset of human physical behaviour and a conscious but noisy signal, with different motivations for its use (see Section 1.2).

Based on previous findings (e.g., Zhang and Pelechrinis (2014)) it is possible some venues may play a greater role in facilitating spatial homophily than others, such as leisure venues (e.g. sports centre) and sociable avenues, (e.g. nightlife spots), as compared to venues people only pass through as a necessity and with little option for choice or self-expression (e.g., transport hubs). Furthermore, each checkin may serve as a signal to social network followers concerning personal affiliations with places that they feel are important.

Given this context, we consider the implications of personality facets on spatial homophily in the following sections. As the literature on spatial homophily and location-based social networks is limited, we additionally consider the usage of online social networks and user personality.

2.1. Openness and spatial homophily

Individuals that score highly with reference to Openness to experience tend to be curious, creative and open to new experiences. People who score low on this facet tend to be conservative and unimaginative in their proposed solutions (Goldberg, 1990). Recent research from spatial homophily (Jokela et al., 2015) suggests that openness to experience might be the strongest predictor of homophilous connections in an LSBN such as Foursquare. Openness to experience was also positively correlated with visiting sociable and popular venues (Chorley et al., 2015). In terms of online social networks, open people tend to enjoy a diverse network of friends (Wehrli, 2009) and are frequent users (Ross et al., 2009; Schrammel, Köffel, & Tscheligi, 2009; Wehrli, 2009). The motivation for use of online social networks by highly open users is most likely tied to their novelty (Amichai-Hamburger & Vinitzky, 2010). Therefore, one could infer that in a LBSN setting, open users might seek popular venues, because such locations appeal to them through their novelty and originality. Sociable venues might be attractive because open people tend to enjoy socializing with and meeting new people. Additionally, by virtue of their curiosity, open people might have a tendency to assort at common venues that are new and interesting to them. However, this could lead to widespread dispersion of checkins, reducing scope for spatial overlap and common checkins, thus resulting in lower spatial homophily. In terms of low openness scoring, such individuals may have a tendency to congregate at a more limited range of familiar places, affecting likelihood of common checkins being detected.

2.2. Extraversion and spatial homophily

Highly extraverted individuals are generally social, talkative and energetic. They tend to engage in many social activities and have a large number of friends. In contrast, introverts tend to be less inclined to engage in social activities, preferring a smaller number of friends and also enjoying doing activities in isolation (Goldberg, 1990). In terms of LBSNs, extraversion has not been found to correlate with any particular checkin behaviours (Chorley et al., 2015), but their high sociability characteristics might make them likely to assort at sociable venues nonetheless (Shen, Brdiczka, & Liu, 2015). When using Facebook, extraverts post and share updates about their social life through photos and events more often than introverts; and have, unsurprisingly, a bigger network of friends in online communities (Amichai-Hamburger & Vinitzky, 2010; Quercia et al., 2011; Schrammel et al., 2009). Therefore LBSNs might be especially suited to extraverts who like to readily share the events and offline activities they take part in through online means (Amichai-Hamburger & Vinitzky, 2010). However we could equally find that extraverts are attracted by a diverse range of venues, and therefore do not display the predicted homophilous behaviour. Furthermore, in terms of online behaviour, extraverts have been found to refrain from using the Internet as a substitute for social interactions (Amiel & Sargent, 2004). This means that extraverts could use LBSNs consistent with meeting friends or partaking in social activities. For online social networks it has also been argued that extraverts, although enjoying a vast number of friends and being less prone to loneliness, tend to have less well connected neighbours, while introverts are embedded in strongly connected networks, albeit with fewer neighbours (Hamburger & Ben-Artzi, 2000; Shen et al., 2015). Introverts post and share less on social media, however, when they do, they gain more likes and comments than their extraverted counterparts (Amichai-Hamburger, Wainapel, & Fox, 2002), providing support for the idea that introverts are embedded in small, but tight-knit social networks. Homophily has been shown to be stronger in smaller communities (Launay & Dunbar, 2015), we could therefore find introverts to be more homophilous than extroverts, including in a location-based social network such as Foursquare.

2.3. Conscientiousness and spatial homophily

Conscientious individuals tend to be well organized and disciplined, while unconscientious people tend to be disorganized and inconsistent (Goldberg, 1990). For online activity, Conscientiousness was found to be negatively correlated with leisure-related Internet use and positively with academic Internet use among adolescents (Landers & Lounsbury, 2006). It has been argued that conscientious users tend to stay focused on their tasks, which makes them less likely to engage in distracting behaviours, such as going on Facebook (Ross et al., 2009). Conscientious users have more friends on Facebook than unconscientious users, but also use some Facebook features less (Amichai-Hamburger & Vinitzky, 2010). Conscientiousness has been linked to the use of LBSN through Foursquare (Chorley et al., 2015), being positively correlated with the number of venues visited. The nature of the Foursquare application might be especially suitable for conscientious users: they consistently remember to checkin at the venues they visit, unlike their more disorganized counterparts. There is no indication that being a consistent LBSN user increases their likelihood to checkin to the same venues, however. Previous social network and communication studies have not identified conscientiousness as playing a role in homophilous processes of other social networks (Amichai-Hamburger & Vinitzky, 2010; Balmaceda, Schiaffino, & Godoy, 2013; Ross et al., 2009). Therefore, in terms of spatial homophily the basis for specific expectations for the conscientiousness facet to be assortative are limited. However, a conscientious user's consistent checkin behaviour might increase the likelihood of detecting homophilic effects.

2.4. Agreeableness and spatial homophily

Highly agreeable people are friendly and likeable. Highly disagreeable people, however, are unpleasant to be around and tend to come across as unfriendly to others (Goldberg, 1990). Highly agreeable people are popular communication partners for extraverted and emotionally stable users (Balmaceda et al., 2013). They also tend to preferentially communicate between themselves, while disagreeable users were not as likely to communicate amongst themselves (Balmaceda et al., 2013). Agreeableness has not been found to be related to number of friends on Facebook or other online communities (Amichai-Hamburger & Vinitzky, 2010; Schrammel et al., 2009; Wehrli, 2009) nor to time spent on Facebook or online in general (Amichai-Hamburger & Vinitzky, 2010; Schrammel et al., 2009). Overall, agreeableness appears assortative in a communication setting, but does not seem to be specifically correlated to online behaviour or social networking site use (Amichai-Hamburger & Vinitzky, 2010; Ross et al., 2009; Schrammel et al., 2009). It was also uncorrelated with venue checkins in Foursquare (Chorley et al., 2015). Other than a friendly atmosphere, it is difficult to speculate on what aspects of a venue attract agreeable individuals. Agreeableness is a personality facet that is most related to social interactions between acquainted individuals, which might be difficult to capture from LBSN data when the relations between users are not known. Communication

between users, the only aspect that agreeable individuals have proven homophilous on (Balmaceda et al., 2013), cannot be assessed. We therefore expect that agreeable LBSN users would not necessarily increase likelihood of attraction to similar venues.

2.5. Neuroticism and spatial homophily

Highly neurotic people are sensitive and nervous, and generally susceptible to negative emotions while emotionally stable people tend to be in control of their emotions (Goldberg, 1990). Neuroticism, which has been associated with a lack of perceived social support, also has a negative relationship with Internet use (Swickert, Hittner, Harris, & Herring, 2002), in particular with leisure usage such as instant messaging and social gaming (Amiel & Sargent, 2004). Neurotic people have been found to avoid discussion boards, showing little interest in participating in them online (Amiel & Sargent, 2004). Unsurprisingly, neurotics are avoided as online interaction partners on discussion boards, even by other neurotics (Balmaceda et al., 2013). Whether these avoidance patterns are reflected in their spatial behaviour is unclear. Emotionally stable users preferred to communicate with agreeable users, but not with each other (Balmaceda et al., 2013). It seems that neurotic people tend to have difficulties forming and maintaining social relationships online and offline (Wehrli, 2009). However, neurotic individuals are speculated to be more comfortable in some online settings, as they are more likely to construe their online persona as their 'real-me' (Amichai-Hamburger et al., 2002), which they create in LBSN by regulating their checkins (Schwartz & Halegoua, 2014). This 'altered' version of their profile might therefore be an inaccurate reflection of their 'true', offline personality. Despite this, neuroticism was found to be negatively correlated with number of checkins to sociable venues (Chorley et al., 2015). However, no relation to spatial homophily was identified in Jokela et al. (2015) and therefore we expect to detect no spatial homophily effect for neuroticism, but one might expect highly neurotic users to be disassortative.

2.6. Overall personality profile

Analysing each of the five personality traits separately gives us valuable insight into homophily processes. However, given the spatial context in which homophily is being investigated, we further consider the overall personality profile. Concerning LBSNs, Graham and Gosling (2011) found that participants were able to accurately predict the personality of typical visitors of a venue, solely based on images from Foursquare. Additionally, previous studies on ties in social networks found that similarity in three of the five facets (extraversion, agreeableness, openness to experience) promotes tie formation (Selfhout et al., 2010). It remains unclear from this study whether tie formation is especially strong among people who score similarly on all three facets at once. According to McPherson et al. (2001), the stronger the connection between two people, the higher their similarity. In line with this assertion, the homophily effect appears to be especially strong among spouses and close friends (McPherson et al., 2001).

However, in the present study, connection between people represents the extent of commonality (i.e., number of checkins) at a location in a LBSN, rather than a direct human relationship. To the extent of our knowledge, this is the first time connection strength has been assessed in this way. But based on previous work on close ties and personality (McPherson et al., 2001; Selfhout et al., 2010) and predictions based on Foursquare activity Graham and Gosling (2011), there is some basis to hypothesise that increased commonality at which checkins are made positively influences overall personality similarity.

2.7. Hypotheses

Based on Sections 2.1–2.6 we summarise the hypotheses as follows:

H1. Open users have a greater tendency to checkin at common venues;

H2. Spatial homophily and conscientiousness are not correlated;

H3. Extraverted users have a greater tendency to checkin at common venues;

H4. Spatial homophily and agreeableness are not correlated;

H5. Neurotic users have a lesser tendency to checkin at common venues;

H6. Greater similarity in overall personality profile implies a greater tendency to checkin at common venues.

3. Methodology

To model spatial homophily we use a graph-based representation, defined as follows.

Definition 1. *For a graph $G=(V,E)$ let node $v\in V$ represent a unique LBSN user, and edge $\{u,v\}\in E$ represent the common checkin of u and v at 1 or more locations. For an edge $e\in E$, let the weight of e, denoted e_w, indicate the number of common venues at which u and v checkin.*

Definition 2. *$G_w=(V_w,E_w)$ is a subgraph of G such that $E_w\subseteq E$, where $e\in E_w$ if and only if e has edge weight of at least w and $V_w\subseteq V$ such that $v\in V_w$ if and only if v has degree of at least 1.*

Graph G allows commonality between individuals, based on checkins, to be assessed. To model the relative ranking of an individual's personality score we label the nodes as described in Definition 3.

Definition 3. *For graph $G=(V,E)$ each node $v\in V$ is labelled with a five-dimensional vector $(v_1,\ldots,v_5)$. v_i indicates the facet value for the i^{th} personality facet, which collectively represent openness, conscientiousness, extraversion, agreeableness and neuroticism.*

The facet value v_i can either represent the actual raw personality rating deduced from personality questionnaire or it can represent the tercile (first, second or third in ascending rank) in which v's personality score is categorised, relative to all nodes within V for the i^{th} facet. We opt to use terciles in our analysis for two main reasons. Firstly, individuals tend to shy away from extreme values in surveys that use midpoints (Weijters, Cabooter, & Schillewaert, 2010). Creating terciles helps disentangle low, middle and high scorers in view of this natural tendency. Secondly, terciles allow a clearer demarkation between the stronger and weaker scores, through which the first and third terciles can be used to test hypotheses that concern extreme values (e.g., extraverts and introverts). For similar reasons, this approach has been successfully adopted for the analysis for personality in a number of settings (e.g., Amichai-Hamburger and Vinitzky (2010); Ross et al. (2009); Schrammel et al. (2009)).

To test for significance relating to the structure of G_w, we benchmark G_w against a set of random graphs R_w^*, where each graph $R_w\in R_w^*$ has the same dimensions as G_w (i.e., same number of nodes and edges). Therefore each node v in R_w corresponds to a node v in G_w, and the corresponding five dimensional facet value vector for v (Definition 3) is fixed for each v in R_w. Thus the personality profile associated with nodes in R_w remains fixed with edges randomised. We use $|R_w^*| = 1000$ and $\overline{R}$ indicates the hypothetical average graph in R^*. This approach is commonly used in

social network analysis (e.g., Zhang and Pelechrinis (2014), Croft et al. (2005)).

3.1. Data collection

The data for the study was collected from an open web-based participatory study (Chorley et al., 2015) that was created to examine checkin behaviour and personality of volunteer users of the Foursquare LBSN. Based on substantial software engineering, this was open to all Foursquare users, and referred to as the 'Foursquare Personality Experiment', which allowed an individual's checkin history to be assessed while undertaking a questionnaire-based assessment of the user's personality.

The Foursquare Personality Experiment was powered by a bespoke web-based system created for the study. The Foursquare location-based social network was adopted because it has comprehensive API that allows application developers access selected checkin information based on users permission, and subject to Foursquare terms and conditions. Participants were recruited worldwide, using an online social media campaign that was promoted through online social networks. Participants were able to access the Foursquare Personality Experiment through a single webpage that initially required a participant to login using their own Foursquare account. The webpage adopted the "OAuth" protocol to ensure the security and privacy of Foursquare login details, from the user's perspective. From this login the software was able to analyse a participant's checkins, using the "venuehistory" API function provided by Foursquare.

Subsequently, on completion of a personality questionnaire tailored for the project, the participants were able to view a map of their checkins. For each venue on the map, a comparison of the participant's own personality as compared to all other participants who checked in at that location was derived and presented. This visualization was used to incentivize participation in uncontrolled conditions. Further details of the system, including visualization, are presented in (Chorley et al., 2015). Participation using this approach allows new forms of exploration to take place but it is important to also understand the related limitations of in-the-wild studies of this nature (Chorley et al., 2015).

Concerning the personality questionnaire, it is recognised that the higher the number of items, the more accurate the personality assessment (Gosling, Rentfrow, & Swann, 2003) and the most recent and standard version of the Big 5 personality questionnaire, the NEO-PI-3, is comprised of 240 items (McCrae et al., 2005). However to maximise completion rates, the 44-item 'Big Five Inventory' (BFI) was used (Benet-Martínez & John, 1998), with answers represented on the Likert scale 1 to 5.

The survey data from the Foursquare Personality Experiment was conducted in the four-month period up to January 2014 and the data collection involved participation from 218 Foursquare users. Personality data was given by 183 of these users. Of these 9 users did not have any checkins, leaving a total of 174 users for analysis. In terms of internal consistency, within the BFI questionnaire the extraversion facet was comprised of 8 items (α=.87), the agreeableness facet of 9 items (α=.81), conscientiousness of 9 items (α=.82), neuroticism of 8 items (α=.83) and openness to experience of 10 items (α=.83). Checkin variables from this data set were assessed in detail (Chorley et al., 2015), addressing correlations concerning number of checkins, number of distinct venues visited, number of checkins at sociable venues, number of sociable venues visited and the average popularity of venues visited.

3.2. Characteristics of G

G has $|V|$=174 and $|E|$=5373, representing an edge density of approximately 35%. Edge weights reach a maximum of 319, with a mean of 2.92 and standard deviation of 10.85. In total 8075 unique venues are represented from 347 Foursquare venue categories. The Foursquare users in the sample considered scored around the midpoint of 3 on the Likert scale for most personality facets as shown in Table 5, with the highest score for openness to experience and the lowest score for neuroticism.

We compare the aggregate personality profile from G (Table 1) with results obtained for the general Internet population (Srivastava, John, Gosling, & Potter, 2003), assuming a sample aged 30 years old. The mean and standard deviation rather than raw scores were available for each facet, and the comparison sample was larger ($N = 3007$). Foursquare users in our sample scored similarly on openness to experience ($t(3180), p = .12$) and marginally lower on extraversion ($t(3180)$=1.86,$p = .06$). However, Foursquare users in our sample scored significantly lower on the conscientiousness facet (mean = 3.43,std = 0.65) compared to the general Internet population (mean = 3.63,std = 0.72), $t(3180)$=3.09,$p = .002$. The Foursquare users in our sample also scored significantly lower on the agreeableness facet (mean = 3.56,std = 0.64) compared to the general Internet population (mean = 3.67,std = 3.69), $t(3180)$=2.25,$p = .02$. Finally, Foursquare users scored significantly lower on neuroticism as well (mean = 2.91,std = 0.73), compared to the general internet population (mean = 3.22,std = 0.84), $t(3180) = 4.78, p<.0001$. However, it must be noted that effect sizes for these differences were small (conscientiousness: $d = .11$; agreeableness: $d = .08$; neuroticism: $d = .17$). In conclusion, our Foursquare sample exhibited some small, albeit significant, differences with a general internet population in terms of personality traits. Generalizability of our subsequent findings to other populations, especially non-internet ones, might therefore be limited.

In Table 2 we present the correlation between facets for graph G. Ideally absolute correlations should be no more than around r=|.30| for facets to be tested without confounding each other. All inter-facet correlations are within or around this threshold with the greatest being neuroticism and agreeableness ($r = -.32$) which is overall weak and deemed acceptable for independent analysis.

Finally we check that representing personality facets by tercile, as commonly adopted in other work (e.g., Ross et al. (2009); Amichai-Hamburger and Vinitzky (2010)), retains strong correlation with raw average personality scores from the completed questionnaires. Let u_i denote the i^{th} personality facet for node u. For a pair of users u,v such that $u,v \in G$, we define the sum of absolute difference between personality profiles as $SAD_{u,v} = \sum_{i=1}^{5} |u_i - v_i|$. When facet values represent terciles (i.e., 1, 2 or 3), this metric is denoted by $SAD^T_{u,v}$. When facet values represent raw personality scores (i.e., a Likert scale rating in the range 1–5), the metric is denoted by $SAD^R_{u,v}$. For all $u,v \in G$, the correlation between $SAD^T_{u,v}$ and $SAD^R_{u,v}$ is significant and strong for all personality facets (openness: $r = .88, p = .0001$; conscientiousness: $r = .89, p = .0001$; extraversion: $r = .91, p = .0001$; agreeableness: $r = .92, p = .0001$;

Table 1
Descriptives for personality scores.

Personality facet	N	Minimum	Maximum	Mean	Std
openness to experience	174	1.20	5.00	3.87	.61
conscientiousness	174	2.00	5.00	3.43	.65
extraversion	174	1.13	5.00	3.15	.84
agreeableness	174	1.89	5.00	3.56	.64
neuroticism	174	1.00	4.50	2.91	.73

Table 2
Pearson correlations across all personality facets of graph G, *significant at $p<.05$,**significant at.$p<.001$

	Openness	Conscientiousness	Extraversion	Agreeableness
conscientiousness	.03			
extraversion	.29**	.26**		
agreeableness	.18*	.17*	.14	
neuroticism	-.11	−.18*	−.24*	−.32**

Table 3
Terciles cut-offs for edge weights.e_w

Tercile group	Lower cut-off	Upper cut-off	N
low	1	1	1592
middle	2	5	1871
top	6	319	1910

Table 4
Descriptives for node clustering in graphs G_w and $\overline{R}_w$

Dependent variable	Minimum	Maximum	Mean	Std
observed (G_1)	0	1	.73	.18
random ($\overline{R_1}$)	.36	.36	.36	.00035
observed (G_2)	0	1	.70	.22
random ($\overline{R_2}$)	.25	.25	.25	.00045
observed (G_6)	0	1	.61	.29
random ($\overline{R_6}$)	.13	.13	.13	.0007

Table 5
Descriptives for personality scores in.G_6

Personality facet	N	Minimum	Maximum	Mean	Std
openness to experience	164	1.20	4.90	3.89	.60
conscientiousness	164	2.11	5.00	3.45	.65
extraversion	164	1.13	5.00	3.15	.84
agreeableness	164	1.89	5.00	3.57	.64
neuroticism	164	1.00	4.38	2.89	.72

neuroticism: $r=.90$,$p=.0001$). This provides confidence that terciles are representative of the raw personality scores.

4. Results

From the checkin data and personality data collected in Section 3.1 a graph G is constructed consistent with Definitions 1 and 3. Three subgraphs of G, G_1, G_2, and G_6 were generated according to Definition 2. $w=1$, $w=2$, and $w=6$ represent meaningful cut-offs for edge weight, when these are distributed according to terciles (Table 3). As mentioned in Section 2.6, homophily effects might increase as connections between nodes grow stronger. Consequently, G_1 represents the graph with the weakest connections (1 common check-in to create an edge), G_2 represents a subgraph with moderate connections (at least 2 common check-ins to create an edge) and G_6 represents a subgraph with strong connections (at least 6 common check-ins to create an edge). We present the results of the analyses in subsequent sections for G_1, G_2 and G_6.

4.1. Characteristics of G_w

G_1 has $|V_1|$=173 and $|E_1|$=5373, representing an edge density of approximately 36%. Mean node degree in G_1 is 62.12 (std = 35.65), with a range from 1 to 137. G_2 has $|V_2|$=170 and $|E_2|$=3781, which represents an edge density of approximately 26%. Mean node degree in G_2 is 44.48 (std = 30.27) with a range from 1 to 122. Finally, G_6 has $|V_6|$=164 and $|E_6|$=1910, with an edge density slightly above 14%. In G_6, mean node degree was 23.29 (std = 20.34) with a range from 1 to 85. Degree differed significantly between G_1, G_2, and G_6. Median node degree in G_2 (median = 44) is significantly lower than in G_1 (median = 62), $U=10{,}451$,$Z=-4.63$,$p=.0001$. Median node degree in G_6 (median = 16.5) is, in turn, significantly lower than in G_2, $U=8{,}122$,$Z=-6.60$,$p=.0001$.

Degree for G_1 is not normally distributed ($W(173)=0.97$,$p=.002$). A skewness value of ($S=0.009$) indicates that the distribution is close to being symmetrical around the mean, suggesting that the right skew of the distribution is limited. Kurtosis values of $K=-0.93$ suggest a platykurtotic distribution, which is qualified by less extreme values at either tails and a flattening of the values around the mean, when compared to a normal distribution (Dancey & Reidy, 2014). Degree for G_2 and G_6 follow a similar distribution with kurtosis values of $K=-0.72$ and $K=0.19$ respectively. Skewness values were $S=0.38$ for G_2 and $S=0.96$ for G_6 (Fig. 1).

Clustering statistics are presented in Table 4. Significantly higher clustering is seen in G_1 (mean = .73, std = 0.18) as compared to R_1 (mean = .36, std = 0.00036), with $U=870$,$Z=-15.21$,$p=.0001$. Clustering was also higher for G_2 compared to R_2 ($U=1566$,$Z=-14.46$,$p=.0001$) and for G_6 compared to R_6 ($U=3828$,$Z=-12.06$,$p=.0001$). This suggests that checkins indeed have a tendency to cluster at particular locations and are not randomly distributed.

4.2. Personality scores for G_w

Personality scores from Foursquare users of G_1, G_2 and G_6 were similar to the users considered in G (Table 5).

Personality scores remained consistent across all subgraphs G_1, G_2 and G_6 even though each subgraph had fewer nodes than the parent graph, G, see Fig. 2. This gives confidence that despite reductions in sample size, subgraphs G_1, G_2 and G_6 are comparable in

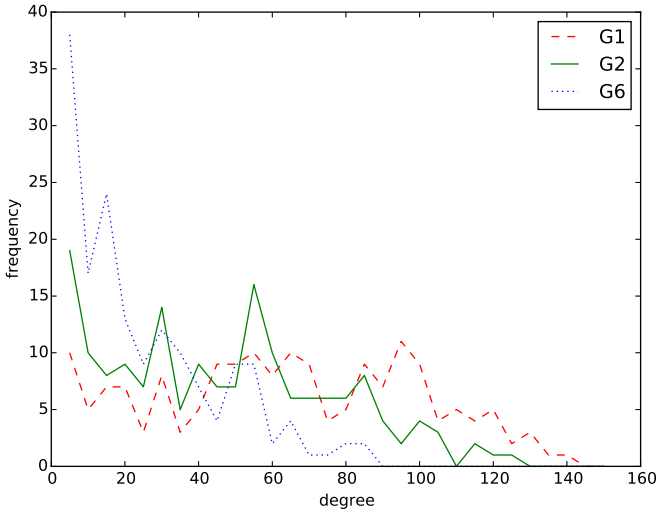

Fig. 1. Degree distribution for G_1, G_2, and.G_6

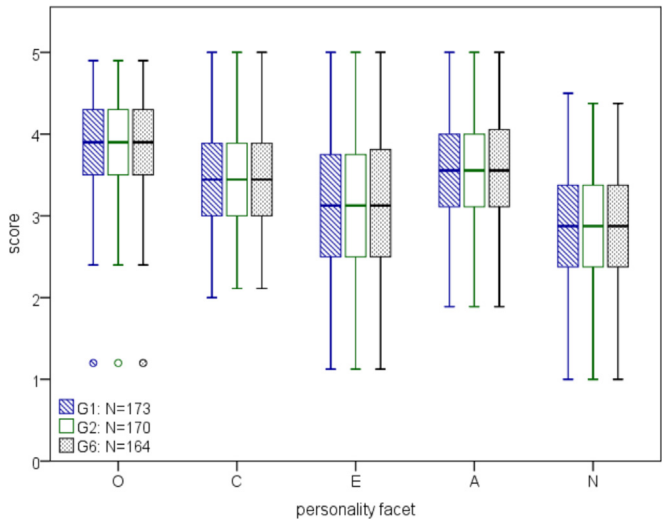

Fig. 2. Scores of each personality facet for G_1, G_2, and.G_6

terms of personality.

4.3. Assessing personality co-occurence

By considering the co-occurrence of similar personality facets at connected nodes in G_w, we are able to assess personality homophily in the context of common checkin locations. Significance is determined by comparison of G_w against $\overline{R}_w$. We firstly assess each facet in isolation, using tercile values. Only personality scores attaining the first and third terciles are considered in our analysis. This avoids ambiguity of mid-scale personality characteristics and focuses on the polar opposite strengths. Thus for graph G_w and personality facet i, all node pairs u,v where $\{u,v\}\in E_w$ and either $u_i = v_i = 1$, $u_i = v_i = 3$ or $u_i = 1$ and $v_i = 3$ are considered. The frequency of the same low facet value connections (both users scored in the 1st tercile), the same high facet value connections (both users scored in the 3rd tercile) and dissimilar facet value connections (one user scored in the 1st tercile and the other in the 3rd tercile) are assessed by comparison with $\overline{R}_w$. The results of the chi-square test on the observed frequencies (from graph G_w) and expected frequencies (from graph $\overline{R}_w$) of each combination, and for each personality facet separately, are presented in Table 6. This approach allows us to directly address hypotheses **H1**–**H5**.

For an individual facet, it is feasible for multiple co-occurrence relationships to be simultaneously significant. For example, given the fixed number of users in tercile 3, a significantly higher number of high facet value connections (i.e., both users in tercile 3) necessitates potentially fewer connections from such nodes to those in tercile 1, which may result in significantly lower dissimilar facet value connections (one user scored in the 1st tercile and the other in the 3rd tercile). Given these dependencies our primary focus concerns low to low or high to high facet interactions.

Hypothesis **H1** is equivalent to the observed frequency of high facet value connections for openness occurring significantly more often than otherwise expected by chance. This was supported by the data for G_1 ($p = .0001$), G_2 ($p = .0001$) and G_6 ($p = .005$). This complements previous findings (Chorley et al., 2015), where Openness to experience was found to be correlated with checkins to popular and sociable venues. Combining these observations, it is feasible that popular and sociable venues could be an underlying feature attracting open people to common locations.

On the other hand, observed frequency of low facet value connections were significantly below expectations for G_1 ($p = .0001$) and G_2 ($p = .016$), but not for G_6 ($p = .71$). This is consistent with the observation that people *low on Openness* tend to be conservative in their choices and this may manifest itself with preference for checkins at familiar locations, instead of exposure to new locations that reflect additional diversity. As a result, individuals with low Openness scores might co-locate with similar others less often, due to reduced opportunities to do so, with this reflected in checkin behaviour.

Hypothesis **H2** is equivalent to the observed frequency of high facet value connections for conscientiousness being not significantly different from chance. Contrary to expectations, conscientious users follow a similar pattern of homophily as open individuals. Observed frequency of high facet value connections was significantly above expectations for G_1, G_2, and G_6 (all $p = .0001$), while low facet value connections were significantly below expectations for G_1, G_2, and G_6 (all $p = .0001$). The observed frequency of dissimilar facet value connections is significantly above expectations for conscientiousness for G_1, G_2, and G_6 (all $p = .0001$).

These results extend the observation in (Chorley et al., 2015) that Conscientiousness and number of checkins in Foursquare correlate, indicating that venue selection has an important role to play for this personality facet. It is possible that Conscientiousness in conducting checkins may well lead to increases in volume which in turn increase the likelihood of common checkins. However, certain characteristics of locations might be especially attractive to conscientious people, such as a well-organized, distraction-free environments, which increases the likelihood of visiting locations that have these characteristics in common, and instigating a checkin.

Hypothesis **H3** is equivalent to the observed frequency of high facet value connections for extraversion being significantly above expectation. Evidence does not support this hypothesis and interestingly it is further observed that the low facet value connections for extraversion are significantly above expectation for G_1 ($p = .005$), but not for G_2 ($p = .17$) and G_6 ($p = .60$). Dissimilar facet value connections are significantly above expectations for G_2 only ($p = .031$).

This indicates that extraverts might not be commonly attracted to specific characteristics of a location, or may not be consistent in displaying checkins based on the location's characteristics. From existing literature, extraverts are known to use social media as a means to portray their social activities but this does not replace their social interactions (Amichai-Hamburger & Vinitzky, 2010), nor do they construe their online self-representation as part of their identity (Amichai-Hamburger et al., 2002). Consequently it is possible that these features of Extraversion are dominant in spatial homophily.

Introverts could also be considered to pursue checkins at locations with common characteristics, which are aligned with the facet (e.g., quietness). However, it is notable that this homophilic effect disappears with increased commonality of checkins (i.e., $w = 2,6$) and so we discount this for further consideration.

Hypothesis **H4** is equivalent to the observed frequency of high facet value connections for agreeableness being insignificant as compared to expectation. This is indeed the case for G_2 ($p = .26$) and G_6 ($p = .22$), but evidence suggests that high facet value connections for agreeableness are significantly above expectation for G_1 ($p = .006$). Surprisingly, low facet value connections are significantly above expectations for G_2 ($p = .009$) and G_6 ($p = .0001$); this is, however, not the case for G_1 ($p = .61$). Dissimilar facet value connections are significantly above expectations for G_1 ($p = .007$), G_2 ($p = .0001$), and G_6 ($p = .0001$).

These unexpected results are of interest given that across the existing literature, of all the personality facets explored, findings concerning Agreeableness have generally featured the least.

Table 6
Parameter Estimates for the effect of pairwise association type on observed and expected frequencies per personality facet for G_1, G_2, and G_6.

Combination	G_1				G_2				G_6			
	NG_1	NR_1	χ^2	*p-value*	NG_2	NR_2	χ^2	*p−value*	NG_6	NR_6	χ^2	*p-value*
openness to experience												
not open- not open	530	617	13.86	**.0001**	360	406	5.84	**.016**	200	205	.14	.71
not open−open	1336	1384	2.24	.13	962	959	.013	.91	507	478	2.35	.13
open−open	866	751	20.47	**.0001**	626	547	13.34	**.0001**	313	270	7.98	**.005**
conscientiousness												
unconscientious−unconscientious	325	517	78.90	**.0001**	217	349	55.00	**.0001**	98	168	31.98	**.0001**
unconscientious-conscientious	1168	1306	19.26	**.0001**	781	916	26.26	**.0001**	381	449	13.46	**.0001**
conscientious−conscientious	981	797	49.88	**.0001**	723	583	39.75	**.0001**	408	287	60.04	**.0001**
extraversion												
introverted−introverted	617	555	7.72	**.005**	417	391	1.93	.17	211	204	.27	.60
introverted-extraverted	1332	1273	3.58	.058	940	884	4.63	**.031**	468	464	.05	.83
extraverted−extraverted	680	706	1.10	.29	489	482	.12	.73	247	253	.16	.69
agreeableness												
disagreeable−disagreeable	417	407	.27	.61	302	261	6.92	**.009**	186	141	15.51	**.0001**
disagreeable-agreeable	1135	1057	7.17	**.007**	809	723	12.65	**.0001**	449	380	15.64	**.0001**
agreeable−agreeable	718	660	5.81	**.02**	505	482	1.26	.26	262	244	1.52	.22
neuroticism												
emotionally stable-emotionally stable	519	479	3.67	.055	370	349	1.39	.24	190	175	1.42	.23
emotionally stable-neurotic	1080	1106	.77	.38	745	795	3.98	**.046**	332	393	11.92	**.001**
neurotic−neurotic	520	617	17.23	**.0001**	375	435	9.35	**.002**	198	213	1.19	.28

Bold text indicates significance with $p < 0.05$.

However this facet may have more significance for spatial homophily because disagreeableness is consistent with the inclination to be critical of others (Goldberg, 1990; Meier & Robinson, 2004). This may manifest itself in specific and stringent standards for the locations they visit. As a result, disagreeable people are more inclined to visit common locations from a much smaller subset of venue types, in contrast to their agreeable counterparts.

Hypothesis **H5** is equivalent to the observed frequency of high facet value connections for neuroticism being significantly below expectation. This is supported by the data for G_1 ($p = .0001$) and G_2 ($p = .002$), but not for G_6 ($p = .28$). Dissimilar facet value connections are, on the other hand, significantly above expectation for G_2 ($p = .046$) and G_6 ($p = .001$).

By virtue of their personality, individuals high in Neuroticism are much more likely to use electronic media to present themselves favourably online (Ross et al., 2009), although they also tend to provide accurate personal information (Amichai-Hamburger et al., 2002; Ross et al., 2009). Furthermore, neurotic individuals might be less inclined to visit locations in the first place, resulting in fewer opportunities to gain common checkins with others. This makes spatial homophily effects less likely to exist for neurotic personalities, which is in line with our findings. It is interesting to note, however, that the spatial behaviour of neurotics offline mirrors the communication behaviour of neurotics online, in the sense that they seem to be less likely to be co-located and communicate, respectively, with one another.

Lastly, it was hypothesized in **H6** that overall personality profiles correlate with a greater tendency to checkin at common venues. This can be assessed using the *SAD* measure as a similarity metric, applying the raw personality scores as defined in Section 3.2. Contrary to our hypothesis, *SAD* scores were similar between graph G_1 (mean = 3.95, std = 1.72) and graph $\overline{R}_1$ (mean = 3.97, std = 1.71), $F(1,10744)=0.62$, $p = .43$. Similarly, there was no significant difference in *SAD* scores for G_2 ($p = .84$) and G_6 ($p = .77$) as compared to $\overline{R}_2$ and $\overline{R}_6$.

5. Discussion

Previous work on personality homophily has focused on the direct attraction between people with similar personality profiles, such as through evidence of particular relationships (e.g., friendships) or interactions between people (e.g., communication). In contrast, the current study addresses personality homophily in the spatial dimension, with connections being defined through commonality of location, as indicated by checkins. Each individual effectively filters whether a visit to a location is recorded by a checkin, and the personality traits themselves could affect the emphasis an individual places on this action (Chorley et al., 2015). These issues are consistent with the new role that LBSNs play in augmenting human behaviour, which has to date received relatively little attention, and results should be interpreted in this context. We note that as compared with other scenarios in which homophily has been addressed, assortative individuals in spatial homophily may be strangers, with limited or implicit awareness of the other individuals with which they assort. Existing literature has very limited coverage of this scenario, meaning that the characteristics of common locations are the indirect attractors driving personality homophily, rather than the characteristics of other LBSN users.

Overall, the hypotheses were not fully supported, which is in part reflective of the basis on which they were formulated, being informed by the dominant literature concerning online social networks rather than homophily in the context of location-based social networks. When considering all personality facets simultaneously (**H6**), personality profile similarity did not correlate with common checkins. Of the individual personality facets considered, only the hypothesis on openness was strongly supported (**H1**). Partial support was found concerning agreeableness (**H4**) and neuroticism (**H5**). No support was found concerning extraversion (**H3**) and the conscientiousness facet proved to be assortative, which was not anticipated (**H2**) and is of particular interest. Results for all hypotheses, including those that are unsupported in the current study, present interesting avenues for future research. While we identified which personality facets might play a role in spatial homophily, we can only speculate on the ways these facets contribute to the observed homophily effect. For example, open individuals could be attracted to venues because they are popular or new, while introverts are attracted to quiet places. Open individuals might also value different characteristics than introverts. Atmosphere might be an important characteristic for them, while introverts value the location of the venue more, for example. Future research will have to determine whether

personality indeed predicts a preference for distinct characteristics of visited venues, and identify what these characteristics precisely are.

We also hypothesised that connection strength could have an effect on overall personality similarity, taking into account all factors simultaneously. However, there was no significant difference between either weakly, moderately or strongly connected users, suggesting that the existence of a connection, rather than its strength, had an effect on personality similarity. In other words, even if users had only visited one common venue, they were already more likely to be similar in terms of personality, compared to users who had never been to the same venue. However, there was no difference in overall personality differences between co-located users and users who had never been to the same venues. This was assessed using the sum of absolute differences (*SAD*) applied to the raw score on the five factor personality profiles. A limitation of this particular analysis is that it is much harder to capture similarity as the number of dimensions increases, and the five personality facets are only weakly correlated, making it less likely that effects based on their aggregated scores are present. A further potential issue of using *SAD* to measure overall personality is the loss of information. Measuring personality scores of users results in loss of information as they are the average of the aggregate scores from the 44 questionnaire items.

It could be argued that these findings, in particularly for disagreeableness, may occur as a consequence of the underlying LBSN database which could skew the availability of pre-existing checkin opportunities around particular locations. We feel this is unlikely given the extent of coverage of Foursquare in the developed world, and the user-generated phenomenon of venue creation leads to multiple checkins sometimes representing the same location, which diminishes the detection of spatial homophily. A further consideration is that users refrain from making checkins, resulting in a loss of information and skewed results. The nodes in the graph-based representation of spatial homophily might therefore appear more clustered than they actually are, and clustering in G_1 is indeed high with a mean of .72. However, this is in line with the small world effect often found in networks with a limited amount of nodes (Milgram, 1967). There is also a notable absence of hubs in our graph-based representation of spatial homophily, which the small world effect also predicts (Milgram, 1967). Degree decreases significantly as commonality increases, while clustering stays relatively constant. A possible explanation is that increased commonality reduces the number of connected individuals in the homophily network, but does not drastically alter the interconnectedness of those same individuals.

5.1. Limitations

It is important to understand the constraints that are inherent in the study, as compared to lab-based experimentation. The open participatory nature of this survey means that conventional controls are relaxed with a view to obtaining data that cannot be conveniently accessed by any other means. Selection by this mechanism is a necessary compromise that allows us to gain new insights, but these need to be interpreted with caution. The broad characteristics of Foursquare usage is consistent with early adopters of technology, who are motivated by new forms of knowledge sharing (e.g., Jadin, Gnambs, and Batinic (2013)). As discussed in Chorley et al. (2015), this means that robust generalisation cannot be made to a wider population, but new insights are provided within a restricted context and it is noted that Foursquare users are not necessarily representative of the general population. In Section 4.2 personality results from the collected data are compared with those of a general Internet population (Srivastava et al., 2003). Results show that subject to the assumptions made in Srivastava et al. (2003), Foursquare users in our study were significantly lower on their conscientiousness, agreeableness and neuroticism, but with a small effect size.

6. Conclusion

Valuable insights have been gained into the co-location patterns of people with similar personality profiles through this study. Our findings further consolidate the importance of individual differences in homophilic processes of social networks. Considering the results overall, Openness and Conscientiousness persist as the most dominant personality traits that are present in spatial homophily, which is consistent with the role that LBSNs fulfil. These findings reflects the indirect nature of spatial homophily where the attraction between participants is a function of location and checkins. We conclude that personality seems to influence spatial and non-spatial homophily quite differently. Both for social (e.g., friendship (Selfhout et al., 2010) or communication (Balmaceda et al., 2013)) and spatial contexts, openness to experience appears to have a positive impact on homophily. Similarly, neuroticism appears to negatively affect homophily in both spatial and social contexts (Balmaceda et al., 2013). However, while extraversion is homophilous in social contexts (Balmaceda et al., 2013; Selfhout et al., 2010), it does not appear to have any particular effect on spatial homophily. On the other hand, conscientiousness appears to play a role in spatial homophily, but not in social homophily. Finally, Agreeableness, which appears to be homophilous among friends (Selfhout et al., 2010) but not among online communication partners (Balmaceda et al., 2013), does not have a significant influence on spatial homophily, as predicted. However, an interesting trend emerged with disagreeable people, who seemed to assort at common locations, while nothing in the literature seems to indicate that disagreeable people associate in social settings (Balmaceda et al., 2013; Selfhout et al., 2010). Future research needs to address current shortcomings in the explanations given for the observed spatial homophily effects. In particular identifying the characteristics in venues that drive the observed effects would be of considerable value.

In summary, we consider that there is a basis for spatial homophily as a consequence of personality, and through the checkin, LBSNs provide a new form of data for its assessment, while also noting caution around the limitations inherent in this approach (Section 5.1). Unanticipated results concerning disagreeableness are of particular interest and signal possible effects concerning decision-making and location. This indicates that different venue types and distinctive characteristics may act as attractors for people with particular selective tendencies. For example, brand associations and the local extent of alternative choice could well be influential factors in driving personality based spatial homophily. The results serve to reaffirm the value and power of new forms of data obtained from mobile and social technology. In particular, the nature of spatial homophily differs considerably as compared to homophily that captures direct attraction between individuals.

Acknowledgements

Roger Whitaker was supported by the RECOGNITION project, Grant No. 257756, funded under the European Commission FP7 Future Emerging Technologies programme. Martin Chorley was supported by a United Kingdom EPSRC Doctoral Award Fellowship, Grant EP/L504749/1. Thomas Pollet was supported by a NIAS Fellowship. Due to the use of a third party API, participant data has been analyzed under terms and conditions that do not permit redistribution of the data.

References

Allport, G. W. (1966). Traits revisited. *American psychologist, 21*(1), 1–10.

Amichai-Hamburger, Y., & Vinitzky, G. (2010). Social network use and personality. *Computers in Human Behavior, 26*(6), 1289–1295.

Amichai-Hamburger, Y., Wainapel, G., & Fox, S. (2002). "on the internet no one knows i'm an introvert": extroversion, neuroticism, and internet interaction. *Cyberpsychology & behavior, 5*(2), 125–128.

Amiel, T., & Sargent, S. L. (2004). Individual differences in internet usage motives. *Computers in Human Behavior, 20*(6), 711–726.

Ashton, M. C., & Lee, K. (2008). The prediction of honestyhumility-related criteria by the HEXACO and five-factor models of personality. *Journal of Research in Personality, 42*(5), 1216–1228.

Balmaceda, J. M., Schiaffino, S., & Godoy, D. (2013). How do personality traits affect communication among users in online social networks? *Online Information Review, 38*, 136–153.

Benet-Martínez, V., & John, O. P. (1998). Los cinco grandes across cultures and ethnic groups: multitrait-multimethod analyses of the big five in spanish and english. *Journal of Personality and Social Psychology, 75*(3), 729.

Block, J. (2001). Millennial contrarianism: the five-factor approach to personality description 5 years later. *Journal of Research in Personality, 35*(1), 98–107.

Bollen, J., Gonçalves, B., Ruan, G., & Mao, H. (2011). Happiness is assortative in online social networks. *Artificial life, 17*, 237–251.

Cho, E., Myers, S. A., & Leskovec, J. (2011). Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1082–1090). ACM.

Chorley, M. J., Whitaker, R. M., & Allen, S. M. (2015). Personality and location-based social networks. *Computers in Human Behavior, 46*, 45–56.

Colombo, G. B., Chorley, M. J., Williams, M. J., Allen, S. M., & Whitaker, R. M. (2012, March). 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* (pp. 217–222). IEEE.

Costa, P. T., & McCrae, R. R. (1985). *The NEO personality inventory: Manual, form S and form R.* Psychological Assessment Resources.

Cramer, H., Rost, M., & Holmquist, L. E. (2011). Performing a check-in: emerging practices, norms and'conflicts' in location-sharing using foursquare. In *Proceedings of the 13th international conference on human computer interaction with mobile devices and services* (pp. 57–66). ACM.

Cranshaw, J., Schwartz, R., Hong, J. I., & Sadeh, N. M. (2012). The livehoods project: utilizing social media to understand the dynamics of a city. In *Proceedings of the sixth international AAAI conference on weblogs and social media* (pp. 58–65). AAAI.

Croft, D. P., James, R., Ward, A. J. W., Botham, M. S., Mawdsley, D., & Krause, J. (2005). Assortative interactions and social networks in fish. *Oecologia, 143*(2), 211–219.

Dancey, C., & Reidy, J. (2014). *Statistics without maths for psychology* (6th ed.). Pearson.

Digman, J. M. (1990). Personality structure: emergence of the five factor model. *Annual Review of Psychology, 41*(1), 417–440.

Goldberg, L. R. (1990). An alternative description of personality : the big-five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216–1229.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality, 37*(6), 504–528.

Graham, L. T., & Gosling, S. D. (2011). Can the ambiance of a place be determined by the user profiles of the people who visit it?. In *Proceedings of the fifth international AAAI conference on weblogs and social media* (pp. 145–152). AAAI.

Hamburger, Y. a., & Ben-Artzi, E. (2000). Relationship between extraversion and neuroticism and the different uses of the internet. *Computers in Human Behavior, 16*(4), 441–449.

Hough, L. M., & Oswald, F. L. (2008). Personality testing and industrial–organizational psychology: reflections, progress, and prospects. *Industrial and Organizational Psychology, 1*(3), 272–290.

Jadin, T., Gnambs, T., & Batinic, B. (2013). Personality traits and knowledge sharing in online communities. *Computers in Human Behavior, 29*(1), 210–216.

Jokela, M., Bleidorn, W., Lamb, M. E., Gosling, S. D., & Rentfrow, P. J. (2015). Geographically varying associations between personality and life satisfaction in the london metropolitan area. In *Proceedings of the National Academy of Sciences* (pp. 725–730).

Joseph, K., Tan, C. H., & Carley, K. M. (2012). Beyond local, categories and friends: clustering foursquare users with latent topics. In *Proceedings of the 2012 ACM conference on ubiquitous computing* (pp. 919–926). ACM.

Kassarjian, H. H. (1971). Personality and consumer behavior: a review. *Journal of Marketing Research, 8*(4), 409–418.

Kossinets, G., & Watts, D. J. (2009). Origins of homophily in an evolving social network. *American Journal of Sociology, 115*(2), 405–450.

Landers, R. N., & Lounsbury, J. W. (2006). An investigation of big five and narrow personality traits in relation to internet usage. *Computers in Human Behavior, 22*(2), 283–293.

Launay, J., & Dunbar, R. I. (2015). Does implied community size predict likeability of a similar stranger? *Evolution and Human Behavior, 36*(1), 32–37.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the hexaco personality inventory. *Multivariate Behavioral Research, 39*(2), 329–358.

Lee, K., Ogunfowora, B., & Ashton, M. C. (2005). Personality traits beyond the big five: are they within the hexaco space? *Journal of Personality, 73*(5), 1437–1463.

Lindqvist, J., Cranshaw, J., Wiese, J., Hong, J., & Zimmerman, J. (2011). I'm the mayor of my house: examining why people use foursquare-a social-driven location sharing application. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2409–2418). ACM.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*(1), 81–90.

McCrae, R., Costa, P., Jr., & Martin, T. (2005). The neo pi 3: a more readable revised neo personality inventory. *Journal of Personality Assessment, 84*(3), 37–41.

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

Meier, B. P., & Robinson, M. D. (2004). Does quick to blame mean quick to anger? The role of agreeableness in dissociating blame and anger. *Personality and Social Psychology Bulletin, 30*(7), 856–867.

Milgram, S. (1967). The small world problem. *Psychology Today, 1*(1), 61–67.

Newman, M. E. J. (2000). Models of the small world. *Journal of Statistical Physics, 101*(3), 819–841.

Noulas, A., Scellato, S., Mascolo, C., & Pontil, M. (2011). An empirical study of geographic user activity patterns in foursquare. *ICwSM, 11*, 570–573.

Pelechrinis, K., & Krishnamurthy, P. (2015). *Socio-spatial affiliation networks*. Computer Communications.

Quercia, D., Kosinski, M., Stillwell, D., & Crowcroft, J. (2011). Our twitter profiles, our selves: predicting personality with twitter. In *Privacy, security, risk and trust (PASSAT) and 2011 IEEE third inernational conference on social computing (SocialCom), 2011 IEEE third international conference on. IEE* (pp. 180–185). E.

Rawlings, D., & Ciancarelli, V. (1997). Music preference and the five-factor model of the neo personality inventory. *Psychology of Music, 25*(2), 120–132.

Rentfrow, P. J., Jokela, M., & Lamb, M. E. (2015). Regional personality differences in great britain. *Plos One, 10*(3), 1–20.

Ross, C., Orr, E. S., Sisic, M., Arseneault, J. M., Simmering, M. G., & Orr, R. R. (2009). Personality and motivations associated with facebook use. *Computers in Human Behavior, 25*(2), 578–586.

Schrammel, J., Köffel, C., & Tscheligi, M. (2009). 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* (pp. 169–174).

Schwartz, R., & Halegoua, G. R. (2014). *The spatial self: Location-based identity performance on social media* (pp. 1–18). New Media & Society.

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.

Shen, J., Brdiczka, O., & Liu, J. (2015). Computers in human behavior a study of facebook behavior : what does it tell about your neuroticism and extraversion ? *Computers in Human Behavior, 45*, 32–38.

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.

Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: set like plaster or persistent change? *Journal of Personality and Social Psychology, 84*(5), 1041.

Swickert, R. J., Hittner, J. B., Harris, J. L., & Herring, J. A. (2002). Relationships among internet use, personality, and social support. *Computers in Human Behavior, 18*(4), 437–451.

Wang, D., Pedreschi, D., Song, C., Giannotti, F., & Barabasi, A.-L. (2011). Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1100–1108). ACM.

Wehrli, S. (2009). *Personality on social network sites: An application of the five factor model.* ETH Zurich Sociology.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: the number of response categories and response category labels. *International Journal of Research in Marketing, 27*(3), 236–247.

Williams, M. J., Whitaker, R. M., & Allen, S. M. (2012). Decentralised detection of periodic encounter communities in opportunistic networks. *Ad Hoc Networks, 10*(8), 1544–1556.

Zhang, K., & Pelechrinis, K. (2014). Understanding spatial homophily: the case of peer influence and social selection. In *Proceedings of the 23rd international conference on world wide web* (pp. 271–282). ACM.

Ziegler, C. N., & Golbeck, J. (2007). Investigating interactions of trust and interest similarity. *Decision Support Systems, 43*(2), 460–475.