City Research Online

City St George's, University of London

Citation: Alessandretti, L., Lehmann, S. & Baronchelli, A. (2018). Individual mobility and social behaviour: Two sides of the same coin. .

This is the draft version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/18994/

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

arXiv:1801.03962v1 [physics.soc-ph] 11 Jan 2018

Individual mobility and social behaviour: Two sides of the same coin

Laura Alessandretti[a,*], Sune Lehmann[b,c], and Andrea Baronchelli[a]

[a]City, University of London, London EC1V 0HB, United Kingdom
[b]Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark
[c]Niels Bohr Institute, University of Copenhagen, DK-2100 København Ø, Denmark
*l.alessandretti@gmail.com

Abstract

According to personality psychology, personality traits determine many aspects of human behaviour. However, validating this insight in large groups has been challenging so far, due to the scarcity of multi-channel data. Here, we focus on the relationship between mobility and social behaviour by analysing two high-resolution longitudinal datasets collecting trajectories and mobile phone interactions of ~ 1000 individuals. We show that there is a connection between the way in which individuals explore new resources and exploit known assets in the social and spatial spheres. We point out that different individuals balance the exploration-exploitation trade-off in different ways and we explain part of the variability in the data by the big five personality traits. We find that, in both realms, extraversion correlates with an individual's attitude towards exploration and routine diversity, while neuroticism and openness account for the tendency to evolve routine over long time-scales. We find no evidence for the existence of classes of individuals across the spatio-social domains. Our results bridge the fields of human geography, sociology and personality psychology and can help improve current models of mobility and tie formation.

Introduction

Two forces shape our social and spatial behaviour. On the one hand, we all experience limitations [1]. Time, cognition, age, the need for food, etc. all constrain our behaviour. On the other hand, each one of us is characterised by personality traits that make us, if perhaps not unique, at least different from many. Personality psychology conjectured a long ago that a set of personality traits underlie all aspects of human behaviour [2, 3].

In the social realm, individuals cope with cognitive and temporal constraints by establishing and maintaining connections in a distinctive [4, 5] and persistent [4] manner. For example, the size of an individual's social circle is bounded under ~ 150, the so-called Dunbar number [6], but varies among individuals around this limit [7]. These differences result from an interplay between physical and extrinsic factors such as gender [8], age [9] and socio-economic status [10] as well as from stable individual dispositions underlying personality [11].

Spatially, individuals are characterised by an *activity space* of repeatedly visited locations within which they move during their daily activities [12], but this geo-spatial signature varies in size [13] and spatial shape [14]. However, unlike the social case, the conjecture that individuals' spatial behaviour is persistent in time [15] had not been verified until recently.

Here, we capitalise on the recent discovery that the size of the activity space is conserved and correlates with the social circle size [16] to test the conjecture that the same personality dispositions in part determine social and spatial behaviour. We test this theory by analysing two long-term datasets consisting of ~ 1000 individuals mobility trajectories and their phone interactions (for previous studies see section '*State of the art*' below).

First, we test the hypothesis that the strategies individuals adopt in order to choose where to go and with whom to interact are similar. Then, we identify and characterise the prevailing socio-spatial

profiles appearing in the datasets. Finally, we show that socio-spatial profiles can be partially explained by the widely adopted big-five personality trait model, often used to describe aspects of the social and emotional life [7,11,17–22]. In the section '*State of the art*', we review the relevant literature; in '*Methods*' we describe data collection and pre-processing, and we provide details of the methods implemented; in '*Results*' we present our findings.

State of the art

Individual-level variability in social and spatial behaviour has mostly been investigated in isolation so far, with few notable efforts to reconcile the two. Here, we briefly review the empirical findings in the two domains.

The social domain

Individuals deal with limited time and cognitive capacity resulting in finite social networks [6, 23] by distributing time unevenly across their social circle [4, 24–28]. While this is a shared strategy, there is clear evidence for individual-level variation. First, social circles vary in terms of diversity: they differ in size [7] - within a maximum upper-bound of ~ 150 individuals [6] - and in structure [4, 29]. Second, individuals display different attitudes towards exploration of social opportunities as they are more or less keen on creating new connections [30–33]. Finally, individuals manage social interactions over time in different ways. Some are characterised by high level of stability as they maintain a very stable social circle, while others renew their social ties at high pace [5].

These heterogeneities can be partially explained by factors including gender [8, 34], age [9, 35, 36], socio-economic status [10, 37] and physical attractiveness [38]. Moreover, as conjectured by personality psychologists [2,39], differences in personalities partially explain the variability in social circle composition [7,11,17,40–44], and the different attitudes towards forming [30,45], developing [20,46] and replacing [29] social connections. It is worth noticing that many of these findings are recent, resulting from the analysis of digital communication traces.

The spatial domain

Constraints including physical capabilities, the distribution of resources, and the need to coordinate with others limit our possibilities to move in space [1]. Individuals cope with these limitations by allocating their time within an activity space of repeatedly visited locations [47], whose size is conserved over several years according to a recent study based on high-resolution trajectories [16], and previous ones based on unevenly sampled and low spatial resolution data [48,49]. The activity space varies across individuals in terms of size [16] and shape [14]: it was shown that two distinct classes of individuals can be identified based on the spatial distribution of their locations, similarly to the social domain [5]. Heterogeneities in spatial behaviour can be explained in terms of gender [50], age [51, 52], socio-economic [35, 53] and ethnic [54] differences. There has only been sporadic efforts to include personality measures in geographic research, despite the strong connections between the two [55]. Recent works [44, 56] suggest that spatial behaviour can be partially explained from personality traits. However, in [56], this understanding is based on biased data collected from location-based social networks. In [44], the connection between spatial behaviour and personality is not investigated extensively, as it is not the main focus of the study.

Social and spatial connection

Recently, connections between the social and spatial behaviour of pairs [57–62] and groups [63] of individuals have been demonstrated, and used to design predictive models of mobility [58, 64, 65] or social ties [59,66–68]. Shifting the attention to the individual level, recent works based on online social network data [69, 70], mobile phone calls data [62] and evenly sampled high resolution mobility trajectories [16] have shown correlations between the activity space size and the ego network structure, calling for further research to more closely examine the connections between social and spatial behaviour at the individual level.

Methods

Data description and pre-processing

Our study is based on 850 high resolution trajectories and call records of participants in a 24 months longitudinal experiment, the Copenhagen Networks Study (CNS) [71]. Results on the connections between social and spatial behaviour were corroborated with data from another experiment with fixed rate temporal sampling, but lower spatial resolution and sample size: the Lausanne Data Collection Campaign (MDC) [72, 73], lasted for 19 months (see Table 1).

	N	δt	T	δx	TC
CNS	850	16 s	24 months	10 m	0.84
MDC	185	60 s	19 months	100-200 m	0.73

Table 1: **Characteristics of the mobility datasets considered.** N is the number of individuals, δt the temporal resolution, T the duration of data collection, δx the spatial resolution, TC the median weekly time coverage, defined as the fraction of time an individual's location is known.

CNS dataset

The Copenhagen Networks Study (CNS) experiment took place between September 2013 and September 2015 [71] and involved ~ 1000 Technical University of Denmark students ($\sim 22\%$ female, $\sim 78\%$ male) typically aged between 19 and 21 years old. Participants' position over time was estimated combining their smart-phones WiFi and GPS data using the method described in [16, 74]. The location estimation error is below 50 meters in 95% of the cases. Participants' calls and sms activity was also collected as part of the experiment. Individuals' background information were obtained through a 310 questions survey including the Big Five Inventory [75], which measures how individuals score on five broad domains of human personality traits: openness, conscientiousness, extraversion, agreeableness, neuroticism. Data collection was approved by the Danish Data Protection Agency. All participants provided individual informed consent. Mobility patterns of participants in the CNS experiment display statistical properties consistent with previous literature [13], as shown in [16].

MDC dataset

Data was collected by the Lausanne Data Collection Campaign between October 2009 and March 2011. The campaign involved an heterogeneous sample of ~ 185 volunteers with mixed backgrounds from the Lake Geneva region (Switzerland), who were allocated smart-phones [73]. In this work we used GSM data, that has the highest temporal sampling. Following Nokia's privacy policy, individuals participating in the study provided informed consent [73]. The Lausanne Mobile Data Challenge experiment involves 62% male and 38% female participants, where the age range 22-33 year-old accounts for roughly 2/3 of the population [76].

Metrics

In this section, we define the concepts and metrics used to quantify the social and spatial behaviour of an individual i.

Exploration behaviour is characterised by the following quantities:

Number of new locations/week: $n_{loc}(i,t)$ is the number of locations discovered by i in the week preceding t. We discard data collected in the first 20 weeks.

Number of new ties/week: $n_{tie}(i,t)$ is the number of individuals who had contact with i (by sms or call) for the first time in the week preceding t.

Exploitation behaviour can be quantified by considering:

Activity space: The set $AS(i,t) = \{\ell_1, \ell_2, ..., \ell_j, ...\ell_C\}$ of locations ℓ_j that individual i visited at least twice and where she spent a time τ_j larger than $200min$ during a time-window of $T = 20$ weeks preceding time t (see Supplementary Material for the analysis with $T = 30$ weeks). Among the locations in the activity space, i visited ℓ_j with probability $p(\ell_j) = \tau_j / \sum \tau_j$. (It is worth noting that this time-based definition of activity space includes all significant locations independently of their spatial position and it is only loosely connected with space-oriented definitions widespread in the geography literature such as the "standard deviational ellipse" and the "road network buffer" [77]).

Social circle: The set $SC(i,t) = \{u_1, u_2, ..., u_j, ...u_k\}$ of individuals u_j with whom individual i had a number of contacts $n_j > 5$ by sms or call during a time-window of $T = 20$ consecutive weeks preceding time t (see Supplementary Material for the analysis with $T = 30$ weeks). The probability that i has contact with a given member u_j of her social circle is $p(u_j) = n_j / \sum n_j$.

For these two sets $AS(i,t)$ and $SC(i,t)$, we consider their sizes $C(i,t)$ and $k(i,t)$, quantifying the number of favoured locations and social ties, respectively; their entropies $H_{AS}(i,t)$ and $H_{SC}(i,t)$, measuring how time is allocated among locations and ties; their stabilities $J_{AS}(i,t)$ and $J_{SC}(i,t)$, quantifying the fraction of conserved locations and ties, respectively, across consecutive non-overlapping windows of $T = 20$ weeks (see Supplementary Material for $T = 30$); their rank turnovers $R_{AS}(i,t)$ and $R_{SC}(i,t)$ measuring the average absolute change in rank of an element in the set between consecutive windows. The mathematical definition of these quantities is provided in Table 2

	Activity space	**Social circle**
1) Size	$C(i,t) = \lvert AS_i(t)\rvert$	$k(i,t) = \lvert SC(i,t)\rvert$
2) Entropy	$H_{AS}(i,t) = \sum_{j=1}^{C(i,t)} p(j) \log p(j)$	$H_{SC}(i,t) = \sum_{j=1}^{k(i,t)} p(j) \log p(j)$
3) Stability	$J_{SC}(i,t) = \frac{\lvert SC(i,t) \cap SC(i,t-T)\rvert}{\lvert SC(i,t) \cup SC(i,t-T)\rvert}^{*}$	$J_{AS}(i,t) = \frac{\lvert AS(i,t) \cap AS(i,t-T)\rvert}{\lvert AS(i,t) \cup AS(i,t-T)\rvert}^{*}$
4) Rank Turnover	$R_{AS}(i,t) = \sum_{j=1}^{N} \frac{\lvert r(j,t) - r(j,t-T)\rvert^{**}}{N}$	$R_{SC}(i,t) = \sum_{j=1}^{N} \frac{\lvert r(j,t) - r(j,t-T)\rvert^{**}}{N}$

* Here $T = 20$ weeks, see Supplementary Material for the analysis with $T = 30$ weeks

** $r(\ell_k, t)$ and $r(u_k, t)$ denote the rank of a location ℓ_k and individual u_k at t, respectively

Table 2: **Definition of the metrics characterising the activity space and the social circle.** 1) The size of a set is the number of elements in the set 2) We compute the entropy of a set considering the probability $p(j)$ associated to each element j of the set. 3) We measure the stability J_{AS} by computing the Jaccard similarity between the activity space at t and at $t-T$, with $T = 20$ weeks. J_{SC} is computed in the same way for the social circle. 4) We compute the rank turnover of a set by measuring for each of its elements j the absolute change in rank between two consecutive time windows of length $T = 20$ weeks. The rank is attributed based on the probability $p(j)$. The average absolute change in rank across all elements corresponds to the rank turnover.

.

Other metrics

In order to compare the difference in entropy between two different sets, we compute their Jensen-Shannon divergence (JSD). The JSD between two sets P_1 and P_2 is computed as $JSD(P_1, P_2) = H(\frac{1}{2}(P_1 + P_2)) - \frac{1}{2}[H(P_1) + H(P_2)]$ (see also [4]).

Results

Both in their spatial and social behaviour, individuals are constantly balancing a trade-off between the exploitation of familiar options (such as returning to a favourite restaurant or spending time with an old friend) and the exploration of new opportunities (such as visiting a new bar or going on a first date) [78]. We adopt this exploration-exploitation perspective to analyse the relationship between social and spatial strategies in our dataset [16].

We quantify the propensity for exploration and exploitation within each individual, i, using the metrics reported in Table 3, Fig. 1 and described in section '*Methods*'. We focus on two aspects of exploitation, (i) *diversity*, characterising how diverse individuals' routine are, and (ii) *evolution*, characterising the tendency to change exploited locations and friends over time.

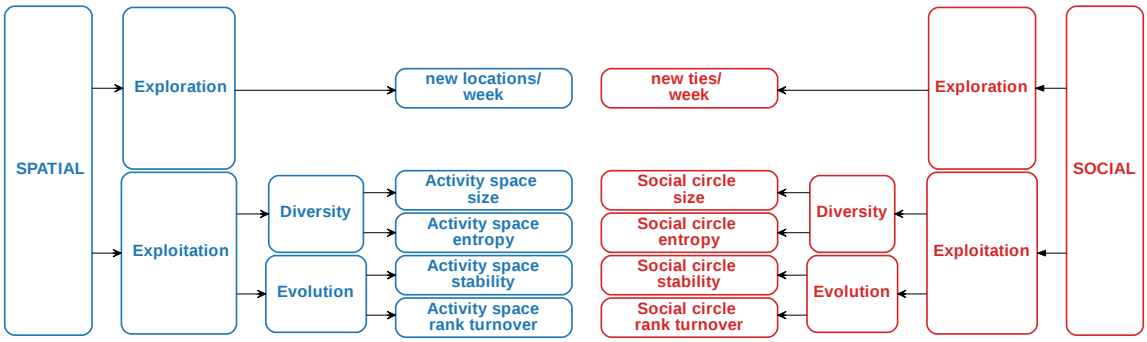

Figure 1: **Schematic description of our framework.**

	Exploration	**Exploitation: Diversity**	**Exploitation: Evolution**
Spatial	New loc./week, n_{loc}	Activity space size, C	Activity space stability, J_{AS}
		Activity space entropy, H_{AS}	Activity space rank turnover, R_{AS}
Social	New ties/week, n_{tie}	Social circle size, k	Social circle stability, J_{SC}
		Social circle entropy, H_{SC}	Social circle rank turnover, R_{SC}

Table 3: **Metrics characterising social and spatial behaviour.** The metrics are defined in section *Methods*

.

Exploration and exploitation are persistent in time. First, we verify that individual behaviour is persistent in time. For all the aformentioned measures, we compare the individual self-variation across time $d_{self}(i)$ with a reference difference $d_{ref}(i,j)$ between individuals i and j. In the case of the activity space size, for example, self-variation is measured as $d_{self} = \langle |C(i,t) - C(i,t-T)| \rangle$, where $\langle \cdot \rangle$ is the average across time and $T = 20$ weeks (see Supplementary Material for $T = 30$) ; the reference difference is computed as $d_{ref}(i,j) = |\langle C(i,t) \rangle - \langle C(j,t) \rangle|$. If $d_{self}(i) < d_{ref}(i,j)$ for most j, we can conclude that for individual i, fluctuations of the activity space size are negligible compared to the difference with other individuals. The same procedure is followed for all metrics with an adjustment in the case of entropies: The persistence of the entropy H_{AS} is verified by comparing the Janson-Shannon divergences $d_{self} = JSD(AS(i,t), AS(i,t-T))$ and $d_{ref} = JSD(AS(i,t), AS(j,t))$. The same method was used for H_{SC} (see Methods and [4]).

Results from the CNS dataset reported in Table 4 show that for all metrics $d_{self}(i) < d_{ref}(i,j)$ holds in more than 99% of cases on average (MDC: 97%, see Supplementary Material Table S1). Moreover, the average self-variation across the population $\overline{d_{self}}$ is consistent with $\overline{d_{self}} = 0$ within errors, and $\overline{d_{self}}$

significantly smaller than the average reference difference $\overline{d_{ref}}$ (see Tables 4 and S1 in Supplementary Material).

	$\overline{d_{self}}$	$\overline{d_{ref}}$	$\overline{d_{self}(i) < d_{ref}(i,j)}$
Social circle size, k	0.04 ± 0.09	12 ± 5	99%
Activity space size, C	0.04 ± 0.07	7 ± 3	99%
New locations/week, n_{loc}	0.05 ± 0.10	0.9 ± 0.5	96%
New ties/week, n_{tie}	0.10 ± 0.17	1 ± 1	95%
Social circle entropy, H_{SC}	0.002 ± 0.007	0.7 ± 0.2	99%
Activity space entropy, H_{AS}	0.002 ± 0.005	0.4 ± 0.1	99%
Social circle stability, J_{SC}	$(9 \pm 22) \cdot 10^{-4}$	0.13 ± 0.05	100%
Activity space stability, J_{AS}	$(9 \pm 26) \cdot 10^{-4}$	0.10 ± 0.04	99%
Social circle rank turnover, R_{SC}	0.05 ± 0.39	2 ± 1	99%
Activity space rank turnover, R_{AS}	0.04 ± 0.10	2 ± 1	99%

Table 4: **CNS dataset: Persistence of social and spatial behaviour.** For each of the social and spatial metrics, $\overline{d_{self}}$ is the average self-distance and $\overline{d_{ref}}$ is the reference distance between an individual and all others, averaged across individuals. The third column reports the fraction of cases where $\overline{d_{self}(i) < d_{ref}(i,j)}$, averaged across the population.

These results extend previous findings [4, 16] and suggest that each individual is characterised by a distinctive socio-spatial behaviour captured by the ensemble of these metrics averaged across time. In fact, these averages are heterogeneously distributed across the samples considered (see Fig. 2).

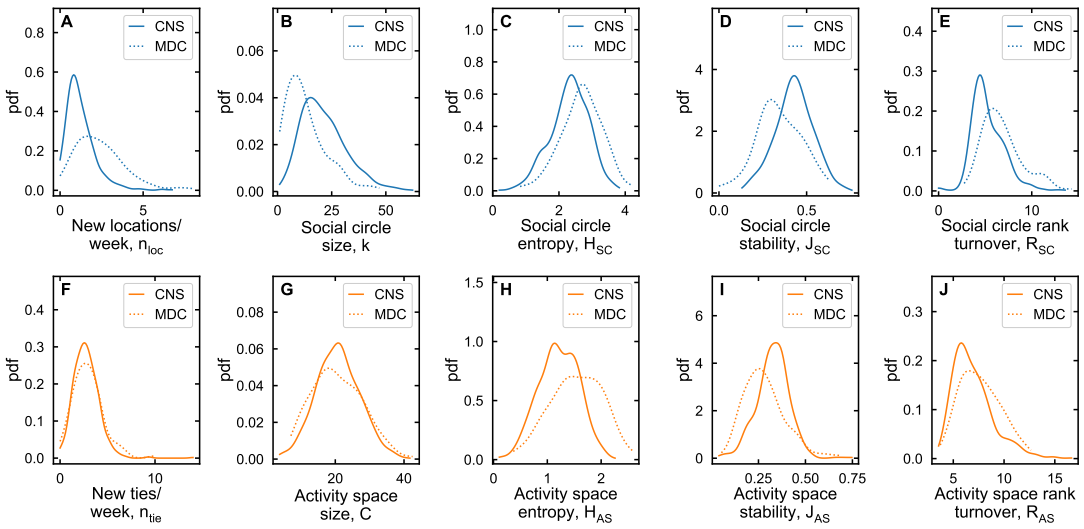

Figure 2: **Distribution of social (above line) and spatial (bottom line) metrics for the CNS and MDC datasets.**

Exploration and exploitation are correlated in the social and spatial domain. A natural way to test the interdependency between social and spatial behaviours is measuring the correlation between a given social metric and a corresponding spatial one. We find positive and significant correlations for all metrics and datasets (see Figs. 3 and S1 in Supplementary Material).

We find that individuals with high propensity to explore new locations are also more keen on exploring social opportunities (see Fig. 3A). Those with diverse mobility routine are also likely to have a

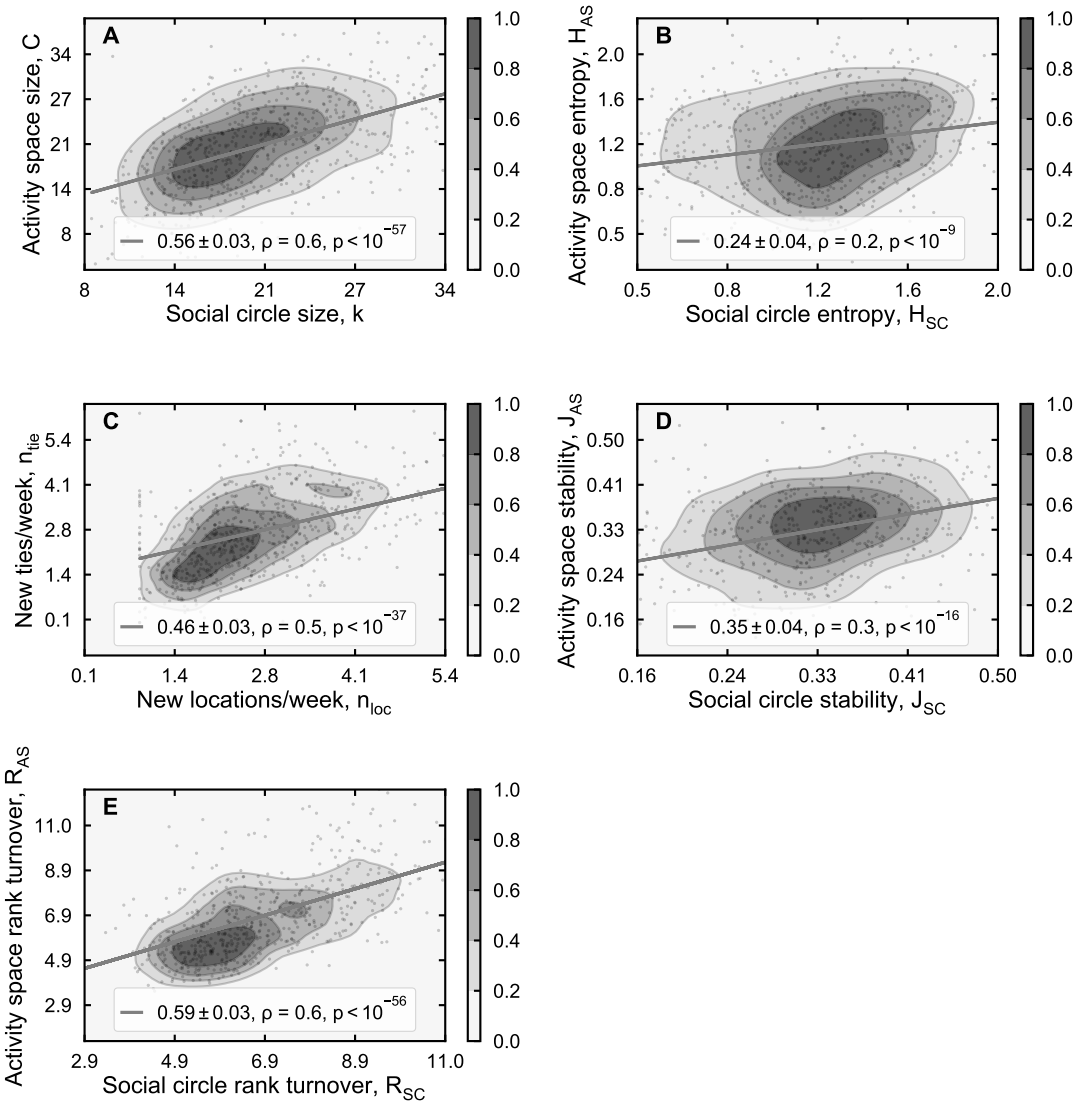

Figure 3: **CNS dataset: correlation between the four dimensions of social and spatial behaviour.** (A) Activity space vs social circle size. (B) Activity space vs social circle composition measured as their entropy. (C) Average number of new locations vs new ties per week. (D) Stability of the activity space vs the stability of the social circle measured as the Jaccard similarity between their composition in consecutive time-windows. (E) Rank turnover of the activity space vs the rank turnover of the social circle. Coloured filled areas correspond to cumulative probabilities estimated via Gaussian Kernel Density estimations. Grey lines correspond to linear fit with angular coefficient b reported in the legend. The Pearson correlation coefficient, with corresponding p-value, is reported in the legend.

correspondingly large social circle (see Fig. 3B), and those that often replace social ties, have also an unstable set of favourite locations (see Fig. 3C and D).

We verify that the observed correlations are not spurious by performing multiple regression analyses that control for other possible sources of variation: gender, age, and time coverage (the average time an individual position is known). We implement five multiple linear regression models M1, M2, M3, M4 and M5. Each regression model predicts a given spatial metric (the activity space size C,the activity space entropy H_{AS}, the number of new locations/week n_{loc}, the activity space stability J_{AS} and the rank turnover R_{AS}) using the corresponding social metric and the control variables (age, gender and time coverage) as regressors. The relative importance of each regressor is assessed using the LMG [79] method.

Results obtained via weighted least square regression (see Tables 5 and S2 in Supplementary Material) reveal that the social metrics are significant predictors for spatial metrics (p value> 0.01 in all cases except for M4 in the MDC dataset), and they typically have more importance than factors such as gender, time, coverage and age group (see Fig. 4).

Among the control variables, gender is a significant predictor of spatial behaviour in the CNS dataset: Females display higher level of routine diversity and propensity towards exploration, in accordance with [80]. Time coverage, measuring the fraction of time an individual position is known, plays a significant role in explaining spatial entropy and activity space stability, since individuals who spend long time in the same place (or leave their phone in the same place) are more easily geo-localised. Age differences are not present within the sample of students participating in the CNS study, and they are not estimated to be relevant with respect to spatial behaviour in the MDC study.

We do not identify distinct classes of individuals. A natural question is whether or not, in the samples considered, there is evidence for distinct classes of individuals based on their socio-spatial behaviour [5, 14]. We approach this problem by reducing the set of metrics to a smaller number of uncorrelated variables by applying Principal Component Analysis [81, 82].

In both datasets, we find that a single predominant component explains $\sim 40\%$ of the differences between individuals (see Table 6). This dimension is dominated by the metrics quantifying exploration and routine diversity (see Table 7). This suggests that individuals with higher exploration propensity tend to have larger social circle and more diverse spatial routine, while those who explore less also have less diverse routines.

The second principal component, which accounts for $\sim 15\%$ of the total variation, is dominated by the effects of evolving routines over long time scales (see Table 7). We consider the two predominant components to reduce the effects of noise and we test the hypothesis that there exists different classes of individuals using the gap statistic method [83]. The test does not support the existence of more than one class of individuals.

The big-five personality traits partly explain spatial and social behaviour. We verify if the differences between individuals can be explained by the Big five personality traits model [75], typically used to describe social and emotional life (see Table 8). We build two multiple linear regression models that use the Big five personality traits as regressors and one of the principal components describing socio-spatial behaviour as target. Results, shown in Table 9, show that three personality traits, neuroticism, openness and extraversion, are relevant predictors for socio-spatial behaviour. In particular, extraversion is the most important predictor of the first principal component: it characterises the tendency to diversify routine and to explore opportunities. Neuroticism and openness explain instead the second principal component, which characterises the tendency to change routine over time (see also Fig. 5). We verify that the same analysis performed considering only the spatial metrics leads to similar results (see Tables 10, 11, 12 and Fig. 6). All the result presented above hold when choosing a time-window with length $T = 30$ weeks (see Supplementary Material, section 2).

Discussion

Using high resolution data from two large scale studies, we have investigated the connection between social and spatial behaviour for the first time. We have shown that, in both domains, individuals balance the trade-off between exploring new opportunities and exploiting known options in a distinctive and persistent manner. We have found that, to a significant extent, individuals adopt a similar strategy in the social and spatial sphere. These strategies are heterogeneous across the two samples considered, and there is no evidence suggesting that there exist distinct classes of individuals. Finally, we have shown that the big five personality traits explain related aspects of both social and spatial behaviour. In particular, we have found that extraverted individuals are more explorative and have diverse routines in both the social and the spatial sphere while neuroticism and openness associate with high level of routine instability in the social and spatial domain.

Our findings confirm the usefulness of mobile phone data to study the connections between behaviour and personality [29, 40, 44, 84–86]. The results are in line with previous findings on the relation between personality and social behaviour: extraversion correlates with social network size [18, 41, 43], openness to experience to social network turnover [29] and neuroticism does not correlate with social network size [11]. Finally, our findings establish a relation between personality and spatial behaviour, validating

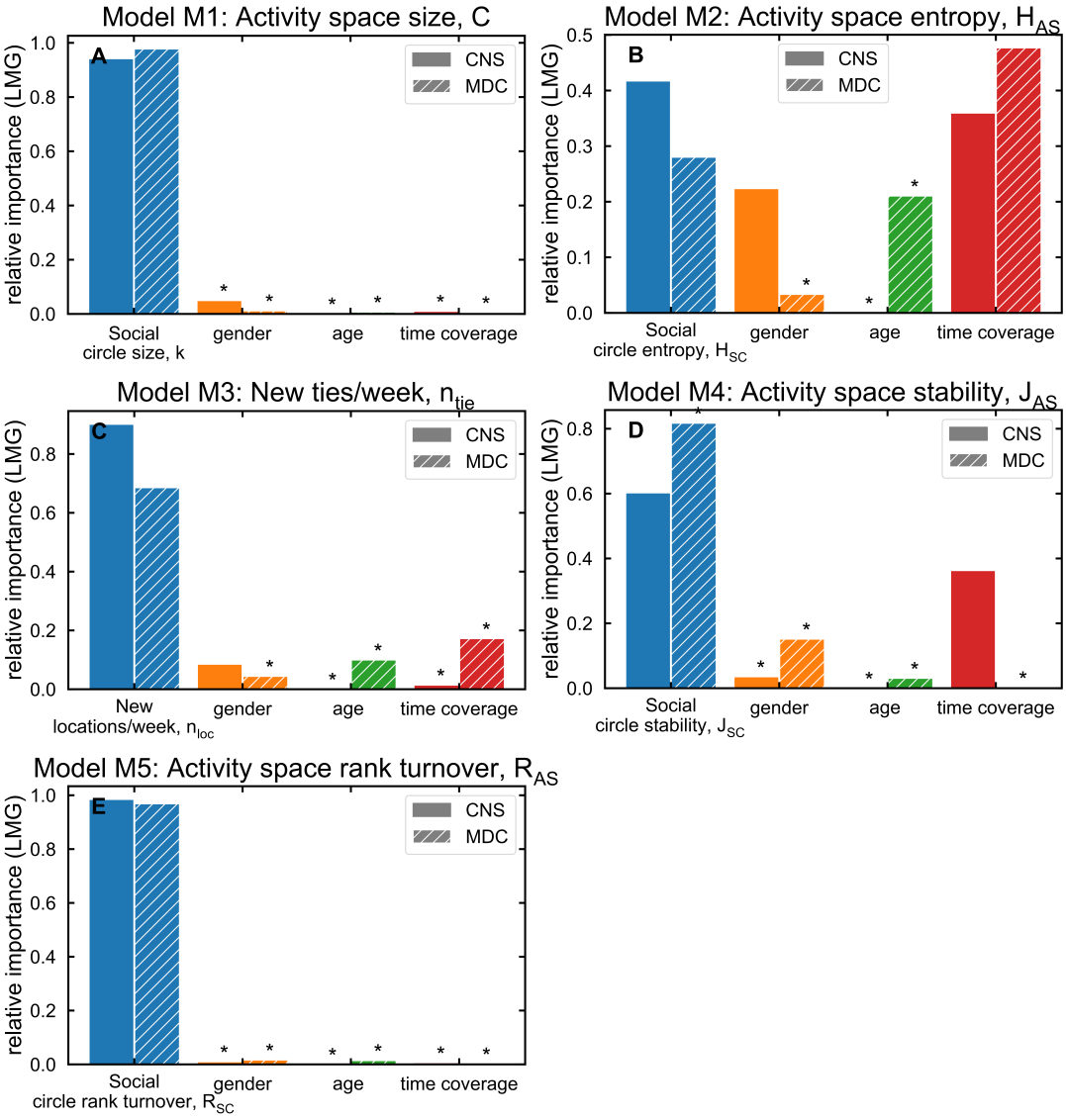

Figure 4: **Relative importance of regressors** LMG of each regressor computed using the Lindeman, Merenda and Gold method [79] for models M1 (A), M2 (B), M3 (C), M4 (D) and M5 (D). Plain bars show results for the CNS dataset, dashed bars for the MDC dataset. Variables that are not significant in the regression model are marked with *.

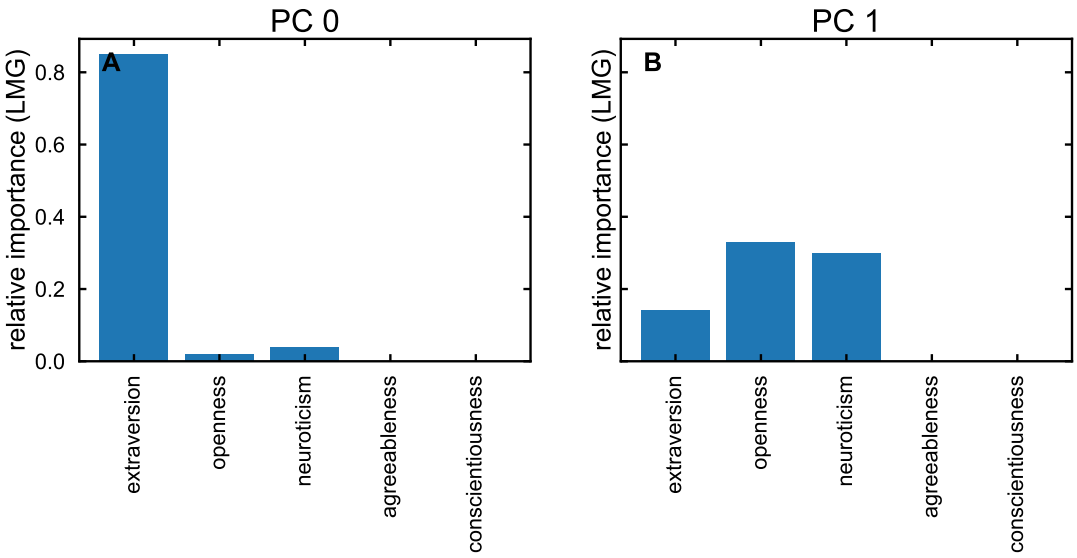

Figure 5: **Relative importance of personality traits for socio-spatial behaviour** Relative importance of each personality trait (computed using the Lindeman, Merenda and Gold method [79]) in explaining the first (A) and the second (B) principal components of socio-spatial behaviour (see also Table 9).

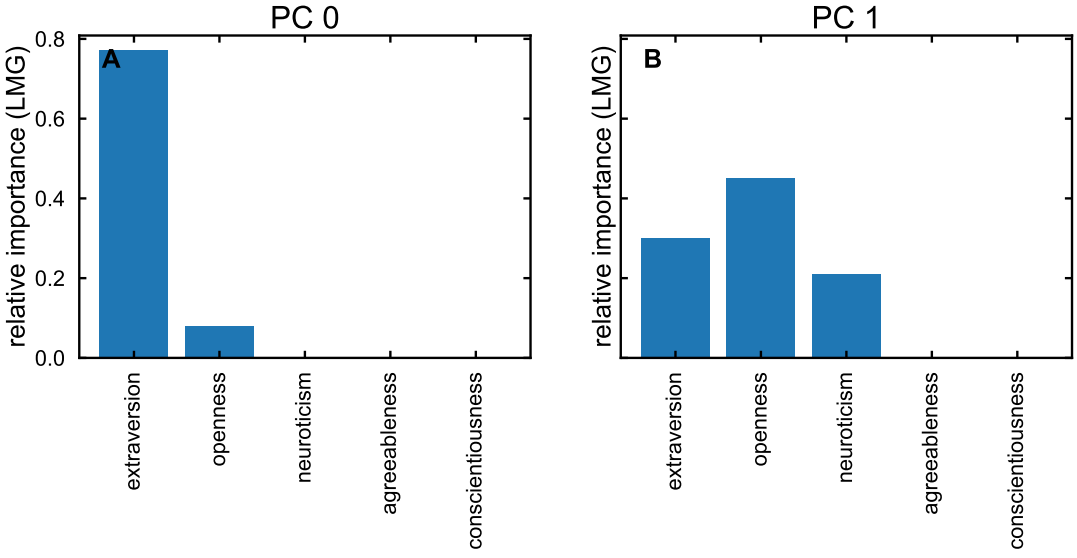

Figure 6: **Relative importance of personalitry traits for spatial behaviour** Relative importance of each personality trait (computed using the Lindeman, Merenda and Gold method [79]) in explaining the first (A) and the second (B) principal components of spatial behaviour (see also Table 12).

Model M1: Activity space size, C	coeff	p val	LMG
Social circle size, k	4 ± 0	$< 10^{-50}$	0.94
gender	-0.4 ± 0.2	0.05	0.05
time coverage	0.4 ± 0.2	0.06	0.01
$[R^2 = 0.32,\ F = 100.44,\ p_F = 0.0\ ]$			
Model M2: Activity space entropy, H_{AS}			
Social circle entropy, H_{SC}	0.07 ± 0.01	$< 10^{-6}$	0.42
gender	-0.06 ± 0.01	$< 10^{-4}$	0.22
time coverage	-0.07 ± 0.01	$< 10^{-5}$	0.36
$[R^2 = 0.11,\ F = 27.30,\ p_F = 0.0]$			
Model M3: New ties/week, n_{tie}			
New locations/week, n_{loc}	0.60 ± 0.05	$< 10^{-32}$	0.9
gender	-0.16 ± 0.05	$< 10^{-3}$	0.08
time coverage	0.001 ± 0.047	1.0	0.01
$[R^2 = 0.22,\ F = 61.99,\ p_F = 0.0]$			
Model M4: Activity space stability, J_{AS}			
Social circle stability, J_{SC}	0.024 ± 0.004	$< 10^{-10}$	0.6
gender	0.007 ± 0.003	0.05	0.04
time coverage	0.017 ± 0.004	$< 10^{-5}$	0.36
$[R^2 = 0.16,\ F = 33.36,\ p_F = 0.0]$			
Model M5: Activity space rank turnover, R_{AS}			
Social circle rank turnover, R_{SC}	1 ± 0	$< 10^{-56}$	0.98
gender	0.12 ± 0.07	0.06	0.01
time coverage	-0.12 ± 0.07	0.07	0.01
$[R^2 = 0.36,\ F = 108.31,\ p_F = 0.0]$			

Table 5: **Linear regression models for the CNS dataset.** For each model, we show the coefficients (coeff) calculated by the regression model, the probability (p val) that the variable is not relevant, and the relative importance (LMG) of each regressor computed using the Lindeman, Merenda and Gold method [79]. Gender is a binary variable taking value 1 for females and 2 for males. For this dataset, age is not relevant as all participants have similar age. For each model, we report the R^2 goodness of fit, the $F-test$ statistics with the corresponfing p-value p_F. W

the theories suggesting that spatial choices are partially dictated by personality dispositions [15] and that a single set of personality traits underlies all aspect of a person's behaviour [2, 3]. The individual

	PC 0	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9
CNS	0.39	0.17	0.12	0.08	0.07	0.06	0.04	0.03	0.03	0.01
MDC	0.43	0.14	0.13	0.08	0.07	0.06	0.04	0.03	0.02	0.01

Table 6: **Variance explained by principal components.** The fraction of variance explained by each principal component for the CNS and MDC dataset.

	CNS		MDC	
	PC 0	PC 1	PC 0	PC 1
Social circle size, k	0.41	0.16	0.37	-0.15
Activity space size, C	0.42	-0.24	0.42	-0.08
New locations/week, n_{loc}	0.33	0.28	0.27	0.33
New ties/week, n_{tie}	0.38	-0.05	0.37	0.19
Social circle entropy, H_{SC}	0.31	0.30	0.34	0.09
Activity space entropy, H_{AS}	0.38	-0.16	0.30	-0.07
Social circle stability, J_{SC}	-0.16	-0.46	0.07	-0.72
Activity space stability, J_{AS}	-0.10	-0.49	-0.12	-0.51
Social circle rank turnover, R_{SC}	-0.20	0.28	-0.33	0.10
Activity space rank turnover, R_{AS}	-0.30	0.44	-0.38	0.17

Table 7: **Principal Components.** The weight of each metric in the first two principal components, for both datasets.

Trait	Related Adjectives
Extraversion	Active, Assertive, Energetic, Enthusiastic, Outgoing, Talkative
Agreeableness	Appreciative, Forgiving, Generous, Kind, Sympathetic
Conscientiousness	Efficient, Organised, Planful, Reliable, Responsible, Thorough
Neuroticism	Anxious, Self-pitying, Tense, Touchy, Unstable, Worrying
Openness to Experience	Artistic, Curious, Imaginative, Insightful, Original, Wide Interests

Table 8: The Big-Five traits and examples of adjectives describing them [88]

characterisation of spatial behaviour is also fundamental to develop conceptual [55] and predictive [87] models of travel behaviour accounting for individual-level differences.

	PC 0 $R^2 = 0.17$, $F = 21.40$, $p_F = 0.0$			PC 1 $R^2 = 0.03$, $F = 3.64$, $p_F = 0.0$		
	coeff	**p val**	**LMG**	**coeff**	**p val**	**LMG**
extraversion	0.85 ± 0.09	$< 10^{-19}$	0.85	0.12 ± 0.06	0.05	0.14
openness	-0.17 ± 0.08	0.03	0.02	0.13 ± 0.06	0.02	0.33
neuroticism	0.25 ± 0.09	0.004	0.04	0.15 ± 0.06	0.02	0.3
agreeableness	0.11 ± 0.08	0.2	0.04	-0.07 ± 0.06	0.2	0.12
conscientiousness	0.06 ± 0.08	0.4	0.04	-0.07 ± 0.06	0.2	0.11

Table 9: **Extraversion, openness, and neuroticism explain socio-spatial behaviour.** The result of a multiple linear regression explaining principal components of socio-spatial data (see Table 7). The value of each coefficient (coeff) is reported together with the probability (p val) that the coefficient is not relevant for the model. The relative importance of each coefficient (LMG) is computed using the LMG method [79].

	PC 0	**PC 1**	**PC 2**	**PC 3**	**PC 4**
CNS	0.53	0.21	0.13	0.10	0.04
MDC	0.56	0.19	0.13	0.07	0.04

Table 10: **Variance explained by principal components (only spatial data).** The fraction of variance explained by each principal component for the CNS and MDC dataset.

	CNS		**MDC**	
	PC 0	**PC 1**	**PC 0**	**PC 1**
Activity space size, C	-0.58	0.02	0.55	0.12
New ties/week, n_{tie}	-0.48	-0.19	0.51	-0.09
Activity space entropy, H_{AS}	-0.50	-0.08	0.43	0.20
Activity space stability, J_{AS}	-0.02	0.94	-0.19	0.95
Activity space rank turnover, R_{AS}	0.43	-0.25	-0.47	-0.16

Table 11: **Principal Components (only spatial data).** The weight of each metric in the first two principal components, for both datasets.

	PC 0 $R^2 = 0.10$, $F = 12.83$, $p_F = 0.0$			**PC 1** $R^2 = 0.03$, $F = 3.50$, $p_F = 0.0$		
	coeff	**p val**	**LMG**	**coeff**	**p val**	**LMG**
extraversion	-0.50 ± 0.07	$< 10^{-10}$	0.77	-0.11 ± 0.05	0.02	0.3
openness	0.19 ± 0.07	0.004	0.08	-0.11 ± 0.04	0.009	0.45
neuroticism	-0.07 ± 0.07	0.4	0.03	-0.10 ± 0.05	0.03	0.21
agreeableness	-0.10 ± 0.07	0.2	0.07	0.01 ± 0.05	0.8	0.01
conscientiousness	-0.05 ± 0.07	0.5	0.05	0.03 ± 0.04	0.4	0.03

Table 12: **Extraversion, openness, and neuroticism explain spatial behaviour.** The result of a multiple linear regression explaining principal components of spatial data (see Table 7). The value of each coefficient (coeff) is reported together with the probability (p val) that the coefficient is not relevant for the model. The relative importance of each coefficient (LMG) is computed using the LMG method [79].

Competing interests

The authors declare that they have no competing interests.

Author's contributions

LA, SL and AB designed the study; LA performed the analysis; LA, SL and AB analysed the results; LA, SL and AB wrote the paper.

Acknowledgements

This work was supported by the Danish Council for Independent Research ("Microdynamics of influence in social systems", grant id. 4184-00556, SL is PI). Portions of the research in this paper used the MDC Database made available by Idiap Research Institute, Switzerland and owned by Nokia.

List of Abbreviations

CNS: Copenhagen Networks Study
MDC: (Lausanne) Mobile Data Challenge
GSM: Global System for Mobile Communications
JSD: Jensen-Shannon divergence
LMG: Lindeman, Merenda and Gold

References

[1] Hägerstraand, T.: What about people in regional science? Papers in regional science **24**(1), 7–24 (1970)

[2] Mischel, W.: Toward a cognitive social learning reconceptualization of personality. Psychological review **80**(4), 252 (1973)

[3] Allport, G.W.: Personality: A psychological interpretation. (1937)

[4] Saramäki, J., Leicht, E.A., López, E., Roberts, S.G., Reed-Tsochas, F., Dunbar, R.I.: Persistence of social signatures in human communication. Proceedings of the National Academy of Sciences **111**(3), 942–947 (2014)

[5] Miritello, G., Lara, R., Cebrian, M., Moro, E.: Limited communication capacity unveils strategies for human interaction. Scientific reports **3** (2013)

[6] Dunbar, R.I.: Coevolution of neocortical size, group size and language in humans. Behavioral and brain sciences **16**(04), 681–694 (1993)

[7] Roberts, S.G., Dunbar, R.I., Pollet, T.V., Kuppens, T.: Exploring variation in active network size: Constraints and ego characteristics. Social Networks **31**(2), 138–146 (2009)

[8] Igarashi, T., Takai, J., Yoshida, T.: Gender differences in social network development via mobile phone text messages: A longitudinal study. Journal of Social and Personal Relationships **22**(5), 691–713 (2005)

[9] Wrzus, C., Hänel, M., Wagner, J., Neyer, F.J.: Social network changes and life events across the life span: A meta-analysis. Psychological bulletin **139**(1), 53 (2013)

[10] McPherson, M., Smith-Lovin, L., Brashears, M.E.: Social isolation in america: Changes in core discussion networks over two decades. American sociological review **71**(3), 353–375 (2006)

[11] Roberts, S.G., Wilson, R., Fedurek, P., Dunbar, R.: Individual differences and personal social network size and structure. Personality and individual differences **44**(4), 954–964 (2008)

[12] Perchoux, C., Chaix, B., Cummins, S., Kestens, Y.: Conceptualization and measurement of environmental exposure in epidemiology: accounting for activity space related to daily mobility. Health & place **21**, 86–93 (2013)

[13] Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.-L.: Understanding individual human mobility patterns. Nature **453**(7196), 779–782 (2008)

[14] Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., Barabási, A.-L.: Returners and explorers dichotomy in human mobility. Nature communications **6** (2015)

[15] Aitken, S.C.: Person-environment theories in contemporary perceptual and behavioural geography i: personality, attitudinal and spatial choice theories. Progress in Human Geography **15**(2), 179–193 (1991)

[16] Alessandretti, L., Sapiezynski, P., Lehmann, S., Baronchelli, A.: Evidence for a conserved quantity in human mobility. arXiv preprint arXiv:1609.03526 (2016)

[17] Lu, Y.-E., Roberts, S., Lio, P., Dunbar, R., Crowcroft, J.: Size matters: variation in personal network size, personality and effect on information transmission. In: Computational Science and Engineering, 2009. CSE'09. International Conference On, vol. 4, pp. 188–193 (2009). IEEE

[18] Casciaro, T.: Seeing things clearly: Social structure, personality, and accuracy in social network perception. Social Networks **20**(4), 331–351 (1998)

[19] Hojat, M.: Loneliness as a function of selected personality variables. Journal of Clinical Psychology **38**(1), 137–141 (1982)

[20] Branje, S.J., van Lieshout, C.F., van Aken, M.A.: Relations between big five personality characteristics and perceived support in adolescents' families. Journal of personality and social psychology **86**(4), 615 (2004)

[21] Amato, P.R.: Personality and social network involvement as predictors of helping behavior in everyday life. Social Psychology Quarterly, 31–43 (1990)

[22] Hamburger, Y.A., Ben-Artzi, E.: The relationship between extraversion and neuroticism and the different uses of the internet. Computers in human behavior **16**(4), 441–449 (2000)

[23] Gonçalves, B., Perra, N., Vespignani, A.: Modeling users' activity on twitter networks: Validation of dunbar's number. PloS one **6**(8), 22656 (2011)

[24] Sutcliffe, A., Dunbar, R., Binder, J., Arrow, H.: Relationships and the social brain: integrating psychological and evolutionary perspectives. British journal of psychology **103**(2), 149–168 (2012)

[25] Arnaboldi, V., Conti, M., Passarella, A., Pezzoni, F.: Analysis of ego network structure in online social networks. In: Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom), pp. 31–40 (2012). IEEE

[26] Miritello, G., Moro, E., Lara, R., Martínez-López, R., Belchamber, J., Roberts, S.G., Dunbar, R.I.: Time as a limited resource: Communication strategy in mobile phone networks. Social Networks **35**(1), 89–95 (2013)

[27] Zhou, W.-X., Sornette, D., Hill, R.A., Dunbar, R.I.: Discrete hierarchical organization of social group sizes. Proceedings of the Royal Society of London B: Biological Sciences **272**(1561), 439–444 (2005)

[28] Marlow, C., Byron, L., Lento, T., Rosenn, I.: Maintained relationships on facebook. Retrieved February **15**(8) (2009)

[29] Centellegher, S., López, E., Saramäki, J., Lepri, B.: Personality traits and ego-network dynamics. PloS one **12**(3), 0173110 (2017)

[30] Wehrli, S., et al.: Personality on social network sites: An application of the five factor model. Zurich: ETH Sociology (Working Paper No. 7) (2008)

[31] Kumar, R., Novak, J., Tomkins, A.: Structure and evolution of online social networks. In: Link Mining: Models, Algorithms, and Applications, pp. 337–357. Springer, ??? (2010)

[32] Newman, M.E.: Clustering and preferential attachment in growing networks. Physical review E **64**(2), 025102 (2001)

[33] Mislove, A., Koppula, H.S., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Growth of the flickr social network. In: Proceedings of the First Workshop on Online Social Networks, pp. 25–30 (2008). ACM

[34] Dunbar, R.I., Spoors, M.: Social networks, support cliques, and kinship. Human nature **6**(3), 273–290 (1995)

[35] Carrasco, J., Miller, E., Wellman, B.: How far and with whom do people socialize?: Empirical evidence about distance between social network members. Transportation Research Record: Journal of the Transportation Research Board (2076), 114–122 (2008)

[36] van den Berg, P., Arentze, T., Timmermans, H.: Size and composition of ego-centered social networks and their effect on geographic distance and contact frequency. Transportation Research Record: Journal of the Transportation Research Board (2135), 1–9 (2009)

[37] Campbell, K.E., Marsden, P.V., Hurlbert, J.S.: Social resources and socioeconomic status. Social networks **8**(1), 97–117 (1986)

[38] Reis, H.T., Nezlek, J., Wheeler, L.: Physical attractiveness in social interaction. Journal of Personality and Social Psychology **38**(4), 604 (1980)

[39] Jaccard, J.J.: Predicting social behavior from personality traits. Journal of Research in Personality **7**(4), 358–367 (1974)

[40] Staiano, J., Lepri, B., Aharony, N., Pianesi, F., Sebe, N., Pentland, A.: Friends don't lie: inferring personality traits from social network structure. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, pp. 321–330 (2012). ACM

[41] Asendorpf, J.B., Van Aken, M.A.: Personality–relationship transaction in adolescence: Core versus surface personality characteristics. Journal of personality **71**(4), 629–666 (2003)

[42] Kalish, Y., Robins, G.: Psychological predispositions and network structure: The relationship between individual predispositions, structural holes and network closure. Social Networks **28**(1), 56–84 (2006)

[43] Pollet, T.V., Roberts, S.G., Dunbar, R.I.: Extraverts have larger social network layers. Journal of Individual Differences (2011)

[44] de Montjoye, Y.-A., Quoidbach, J., Robic, F., Pentland, A.: Predicting personality using novel mobile phone-based metrics. In: SBP, pp. 48–55 (2013). Springer

[45] Selfhout, M., Burk, W., Branje, S., Denissen, J., Van Aken, M., Meeus, W.: Emerging late adolescent friendship networks and big five personality traits: A social network approach. Journal of personality **78**(2), 509–538 (2010)

[46] Asendorpf, J., Denissen, J.J.: Predictive validity of personality types versus personality dimensions from early childhood to adulthood: Implications for the distinction between core and surface traits. Merrill-Palmer Quarterly **52**(3), 486–513 (2006)

[47] Golledge, R.G.: Spatial Behavior: A Geographic Perspective. Guilford Press, ??? (1997)

[48] Järv, O., Ahas, R., Witlox, F.: Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. Transportation Research Part C: Emerging Technologies **38**, 122–135 (2014)

[49] Schönfelder, S., Axhausen, K.W.: Urban Rhythms and Travel Behaviour: Spatial and Temporal Phenomena of Daily Travel. Ashgate Publishing, Ltd., ??? (2010)

[50] Kwan, M.-P.: Gender differences in space-time constraints. Area **32**(2), 145–156 (2000)

[51] Vazquez-Prokopec, G.M., Bisanzio, D., Stoddard, S.T., Paz-Soldan, V., Morrison, A.C., Elder, J.P., Ramirez-Paredes, J., Halsey, E.S., Kochel, T.J., Scott, T.W., *et al.*: Using gps technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment. PloS one **8**(4), 58802 (2013)

[52] Kang, C., Gao, S., Lin, X., Xiao, Y., Yuan, Y., Liu, Y., Ma, X.: Analyzing and geo-visualizing individual human mobility patterns using mobile call records. In: Geoinformatics, 2010 18th International Conference On, pp. 1–7 (2010). IEEE

[53] Zenk, S.N., Schulz, A.J., Matthews, S.A., Odoms-Young, A., Wilbur, J., Wegrzyn, L., Gibbs, K., Braunschweig, C., Stokes, C.: Activity space environment and dietary and physical activity behaviors: a pilot study. Health & place **17**(5), 1150–1161 (2011)

[54] Kwan, M.-P., Lee, J.: Geovisualization of human activity patterns using 3d gis: a time-geographic approach. Spatially integrated social science **27** (2004)

[55] Van Acker, V., Van Wee, B., Witlox, F.: When transport geography meets social psychology: toward a conceptual model of travel behaviour. Transport Reviews **30**(2), 219–240 (2010)

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

[57] Backstrom, L., Sun, E., Marlow, C.: Find me if you can: improving geographical prediction with social and spatial proximity. In: Proceedings of the 19th International Conference on World Wide Web, pp. 61–70 (2010). ACM

[58] McGee, J., Caverlee, J., Cheng, Z.: Location prediction in social media based on tie strength. In: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, pp. 459–468 (2013). ACM

[59] Sadilek, A., Kautz, H., Bigham, J.P.: Finding your friends and following them to where you are. In: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, pp. 723–732 (2012). ACM

[60] Crandall, D.J., Backstrom, L., Cosley, D., Suri, S., Huttenlocher, D., Kleinberg, J.: Inferring social ties from geographic coincidences. Proceedings of the National Academy of Sciences **107**(52), 22436–22441 (2010)

[61] Grabowicz, P.A., Ramasco, J.J., Gonçalves, B., Eguíluz, V.M.: Entangling mobility and interactions in social media. PloS one **9**(3), 92196 (2014)

[62] Toole, J.L., Herrera-Yaqüe, C., Schneider, C.M., González, M.C.: Coupling human mobility and social ties. Journal of The Royal Society Interface **12**(105), 20141128 (2015)

[63] Onnela, J.-P., Arbesman, S., González, M.C., Barabási, A.-L., Christakis, N.A.: Geographic constraints on social network groups. PLoS one **6**(4), 16939 (2011)

[64] Jurgens, D.: That's what friends are for: Inferring location in online social media platforms based on social relationships. ICWSM **13**(13), 273–282 (2013)

[65] Cho, E., Myers, S.A., Leskovec, J.: 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 (2011). ACM

[66] Wang, D., Pedreschi, D., Song, C., Giannotti, F., Barabasi, A.-L.: 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 (2011). ACM

[67] Scellato, S., Noulas, A., Mascolo, C.: Exploiting place features in link prediction on location-based social networks. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1046–1054 (2011). ACM

[68] Pham, H., Shahabi, C., Liu, Y.: Ebm: an entropy-based model to infer social strength from spatiotemporal data. In: Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data, pp. 265–276 (2013). ACM

[69] Cheng, Z., Caverlee, J., Lee, K., Sui, D.Z.: Exploring millions of footprints in location sharing services. ICWSM **2011**, 81–88 (2011)

[70] Cranshaw, J., Toch, E., Hong, J., Kittur, A., Sadeh, N.: Bridging the gap between physical location and online social networks. In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing, pp. 119–128 (2010). ACM

[71] Stopczynski, A., Sekara, V., Sapiezynski, P., Cuttone, A., Madsen, M.M., Larsen, J.E., Lehmann, S.: Measuring large-scale social networks with high resolution. PloS one **9**(4), 95978 (2014)

[72] Kiukkonen, N., Blom, J., Dousse, O., Gatica-Perez, D., Laurila, J.: Towards rich mobile phone datasets: Lausanne data collection campaign. Proc. ICPS, Berlin (2010)

[73] Laurila, J.K., Gatica-Perez, D., Aad, I., Bornet, O., Do, T.-M.-T., Dousse, O., Eberle, J., Miettinen, M., *et al.*: The mobile data challenge: Big data for mobile computing research. In: Pervasive Computing (2012)

[74] Sapiezynski, P., Gatej, R., Mislove, A., Lehmann, S.: Opportunities and challenges in crowdsourced wardriving. In: Proceedings of the 2015 ACM Conference on Internet Measurement Conference, pp. 267–273 (2015). ACM

[75] John, O.P., Srivastava, S.: The big five trait taxonomy: History, measurement, and theoretical perspectives. Handbook of personality: Theory and research **2**(1999), 102–138 (1999)

[76] Laurila, J.K., Gatica-Perez, D., Aad, I., Blom, J., Bornet, O., Do, T.M.T., Dousse, O., Eberle, J., Miettinen, M.: From big smartphone data to worldwide research: The mobile data challenge. Pervasive and Mobile Computing **9**(6), 752–771 (2013)

[77] Sherman, J.E., Spencer, J., Preisser, J.S., Gesler, W.M., Arcury, T.A.: A suite of methods for representing activity space in a healthcare accessibility study. International journal of health geographics **4**(1), 24 (2005)

[78] Hills, T.T., Todd, P.M., Lazer, D., Redish, A.D., Couzin, I.D., Group, C.S.R., *et al.*: Exploration versus exploitation in space, mind, and society. Trends in cognitive sciences **19**(1), 46–54 (2015)

[79] Grömping, U., *et al.*: Relative importance for linear regression in r: the package relaimpo. Journal of statistical software **17**(1), 1–27 (2006)

[80] Mollgaard, A., Lehmann, S., Mathiesen, J.: Correlations between human mobility and social interaction reveal general activity patterns. PloS one **12**(12), 0188973 (2017)

[81] Wold, S., Esbensen, K., Geladi, P.: Principal component analysis. Chemometrics and intelligent laboratory systems **2**(1-3), 37–52 (1987)

[82] Mika, S., Schölkopf, B., Smola, A.J., Müller, K.-R., Scholz, M., Rätsch, G.: Kernel pca and denoising in feature spaces. In: Advances in Neural Information Processing Systems, pp. 536–542 (1999)

[83] Tibshirani, R., Walther, G., Hastie, T.: Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **63**(2), 411–423 (2001)

[84] Lambiotte, R., Kosinski, M.: Tracking the digital footprints of personality. Proceedings of the IEEE **102**(12), 1934–1939 (2014)

[85] Bogomolov, A., Lepri, B., Ferron, M., Pianesi, F., Pentland, A.S.: Daily stress recognition from mobile phone data, weather conditions and individual traits. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 477–486 (2014). ACM

[86] Chittaranjan, G., Blom, J., Gatica-Perez, D.: Mining large-scale smartphone data for personality studies. Personal and Ubiquitous Computing **17**(3), 433–450 (2013)

[87] Jiang, S., Yang, Y., Gupta, S., Veneziano, D., Athavale, S., González, M.C.: The timegeo modeling framework for urban motility without travel surveys. Proceedings of the National Academy of Sciences, 201524261 (2016)

[88] McCrae, R.R., John, O.P.: An introduction to the five-factor model and its applications. Journal of personality **60**(2), 175–215 (1992)

Supplementary Material for Individual mobility and social behaviour: Two sides of the same coin

1 Results obtained with the MDC dataset

Tables S1, S2 and Fig. S1 report the results of the persistence analysis, the multiple regression analysis, and the correlation analysis for the MDC dataset.

	$\overline{d_{self}}$	$\overline{d_{ref}}$	$\overline{d_{self}(i) < d_{ref}(i,j)}$
Social circle size, k	0.05 ± 0.13	10 ± 5	97%
Activity space size, C	0.07 ± 0.12	8 ± 3	97%
New locations/week, n_{loc}	0.2 ± 0.3	2 ± 1	91%
New ties/week, n_{tie}	0.2 ± 0.6	2 ± 1	90%
Social circle entropy, H_{SC}	0.006 ± 0.014	0.7 ± 0.3	97%
Activity space entropy, H_{AS}	0.004 ± 0.008	0.5 ± 0.2	97%
Social circle stability, J_{SC}	0.002 ± 0.005	0.15 ± 0.05	99%
Activity space stability, J_{AS}	0.002 ± 0.004	0.12 ± 0.05	99%
Social circle rank turnover, R_{SC}	0.07 ± 0.15	2 ± 1	98%
Activity space rank turnover, R_{AS}	0.2 ± 0.6	2 ± 1	97%

Table S1: **MDC dataset: Persistence of social and spatial behaviour.** For each of the social and spatial metrics, $\overline{d_{self}}$ is the average self-distance and $\overline{d_{ref}}$ is the reference distance between an individual and all others, averaged across individuals. The third column reports the fraction of cases where $\overline{d_{self}(i) < d_{ref}(i,j)}$, averaged across the population.

Model M1: Activity space size, C	coeff	p val	LMG
Social circle size, k	5 ± 1	$< 10^{-11}$	0.98
gender	0.1 ± 0.6	0.8	0.01
age group	0.6 ± 0.6	0.3	0.01
time coverage	-0.4 ± 0.6	0.4	0.0
$[R^2 = 0.40,\ F = 16.80,\ p_F = 0.0]$			
Model M2: Activity space entropy, H_{AS}			
Social circle entropy, H_{SC}	0.11 ± 0.04	0.009	0.28
gender	0.04 ± 0.04	0.3	0.03
age group	-0.08 ± 0.04	0.06	0.21
time coverage	-0.14 ± 0.04	0.002	0.48
$[R^2 = 0.20,\ F = 6.50,\ p_F = 0.0]$			
Model M3: New ties/week, n_{tie}			
New locations/week, n_{loc}	0.5 ± 0.1	0.002	0.69
gender	0.01 ± 0.15	0.9	0.04
age group	0.2 ± 0.1	0.2	0.1
time coverage	-0.3 ± 0.1	0.06	0.17
$[R^2 = 0.13,\ F = 3.78,\ p_F = 0.0]$			
Model M4: Activity space stability, J_{AS}			
Social circle stability, J_{SC}	0.02 ± 0.01	0.1	0.82
gender	-0.006 ± 0.012	0.6	0.15
age group	-0.003 ± 0.012	0.8	0.03
time coverage	$(-10 \pm 1213) \cdot 10^{-5}$	1.0	0.0
$[R^2 = 0.04,\ F = 0.80,\ p_F = 0.5]$			
Model M5: Activity space rank turnover, R_{AS}			
Social circle rank turnover, R_{SC}	1 ± 0	$< 10^{-15}$	0.97
gender	0.04 ± 0.15	0.8	0.02
age group	-0.2 ± 0.1	0.1	0.01
time coverage	-0.06 ± 0.15	0.7	0.0
$[R^2 = 0.55,\ F = 27.24,\ p_F = 0.0]$			

Table S2: **Linear regression models for the MDC dataset.** For each model, we report the R^2 goodness of fit, the $F-test$ statistics with the corresponfing p-value p_F. We show the coefficients (coeff) calculated by the regression model, the probability (p val) that the variable is not relevant, and the relative importance (LMG) of each regressor computed using the Lindeman, Merenda and Gold method. Gender is a binary variable taking value 1 for females and 2 for males.

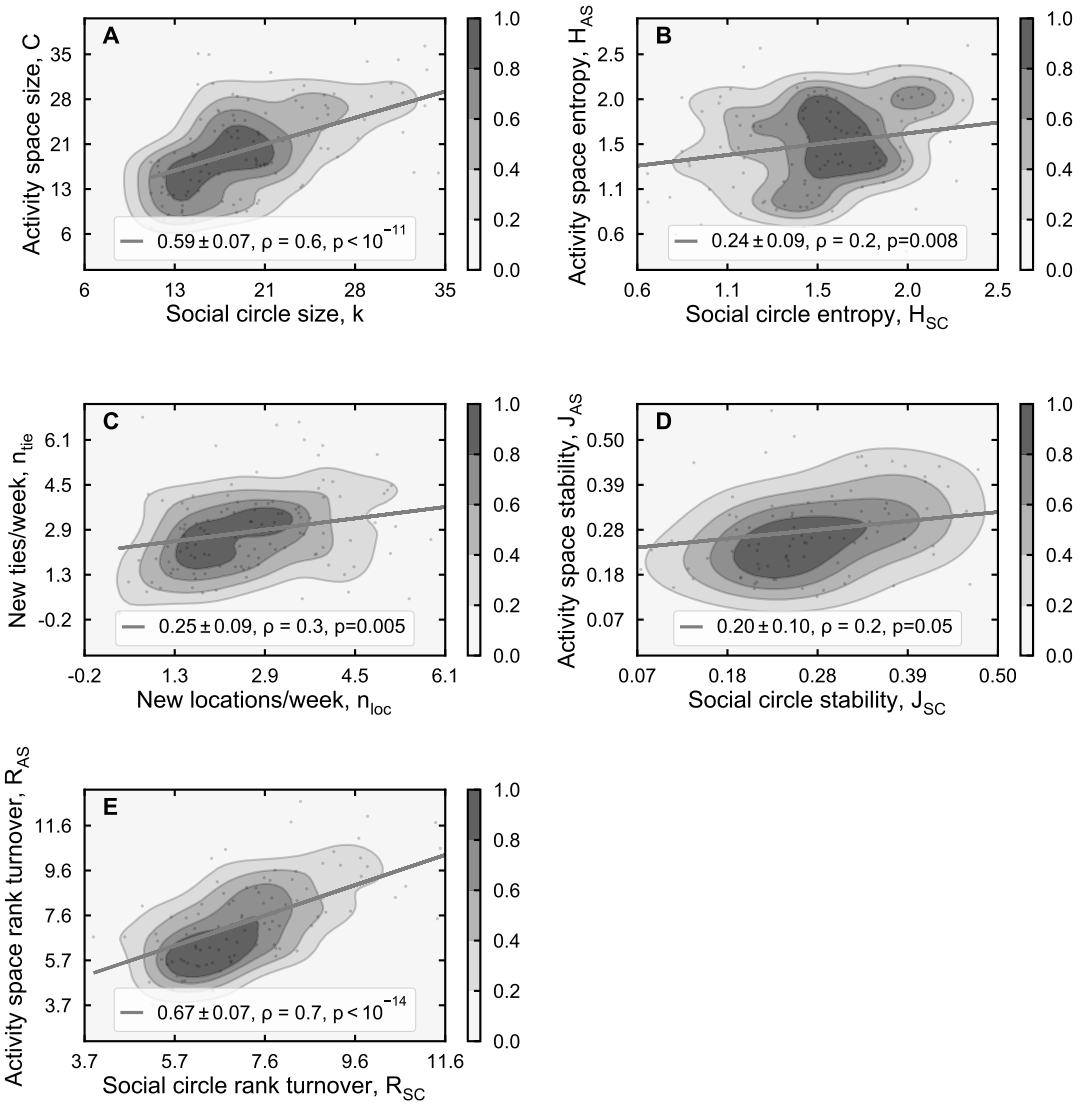

Figure S1: **MDC dataset: correlation between the four dimensions of social and spatial behaviour.** (A) Activity space vs social circle size. (B) Activity space vs social circle composition measured as their entropy. (C) Average number of new locations vs new ties per week. (D) Stability of the activity space vs the stability of the social circle measured as the Jaccard similarity between their composition in consecutive time-windows. (E) Rank turnover of the activity space vs the rank turnover of the social circle. Coloured filled areas correspond to cumulative probabilities estimated via Gaussian Kernel Density estimations. Grey lines correspond to linear fit with angular coefficient b reported in the legend. The Pearson correlation coefficient, with corresponding p-value, is reported in the legend.

2 Results obtained with other windows

Figs. S3, S5, S5, S6 and Tables S3, S4, S5, S6, S7, S8, S9, S10 report the results obtained choosing a time-window with length $T = 30$ weeks (see main manuscript, section 'Methods').

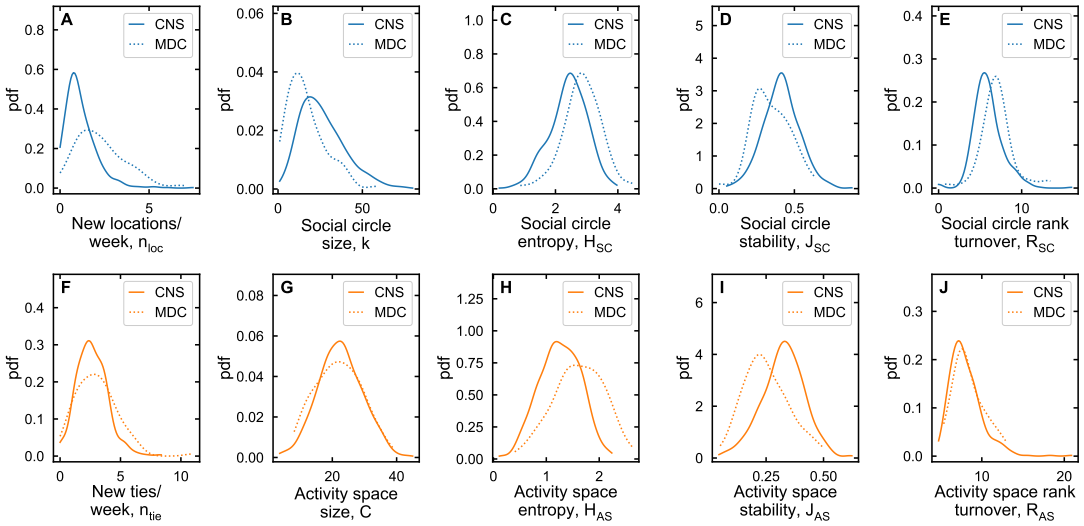

Figure S2: **T=30, Distribution of social (above line) and spatial (bottom line) metrics for the CNS and MDC datasets.**

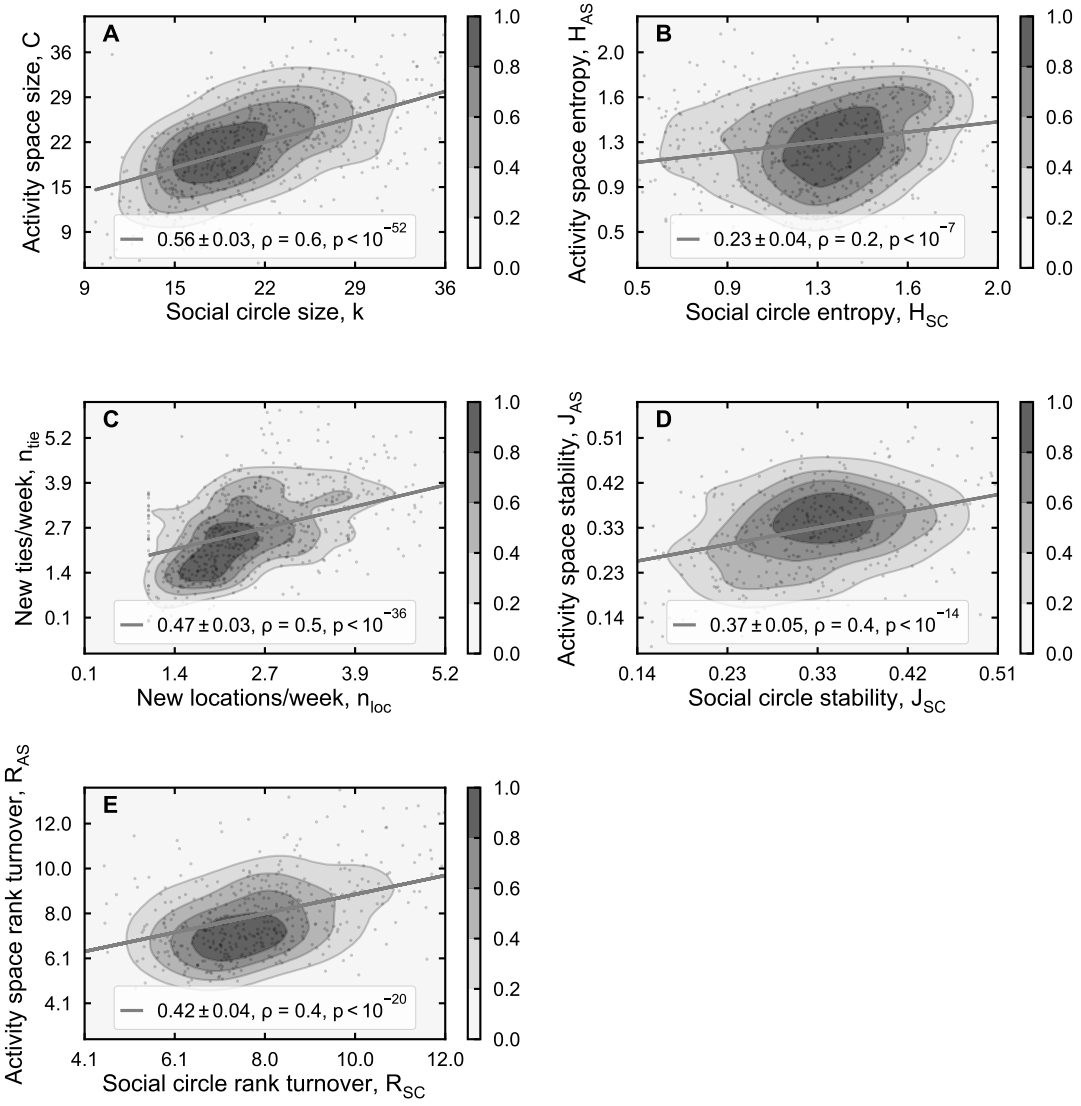

Figure S3: **T=30, CNS dataset: correlation between the four dimensions of social and spatial behaviour.** (A) Activity space vs social circle size. (B) Activity space vs social circle composition measured as their entropy. (C) Average number of new locations vs new ties per week. (D) Stability of the activity space vs the stability of the social circle measured as the Jaccard similarity between their composition in consecutive time-windows. (E) Rank turnover of the activity space vs the rank turnover of the social circle. Coloured filled areas correspond to cumulative probabilities estimated via Gaussian Kernel Density estimations. Grey lines correspond to linear fit with angular coefficient b reported in the legend. The Pearson correlation coefficient, with corresponding p-value, is reported in the legend.

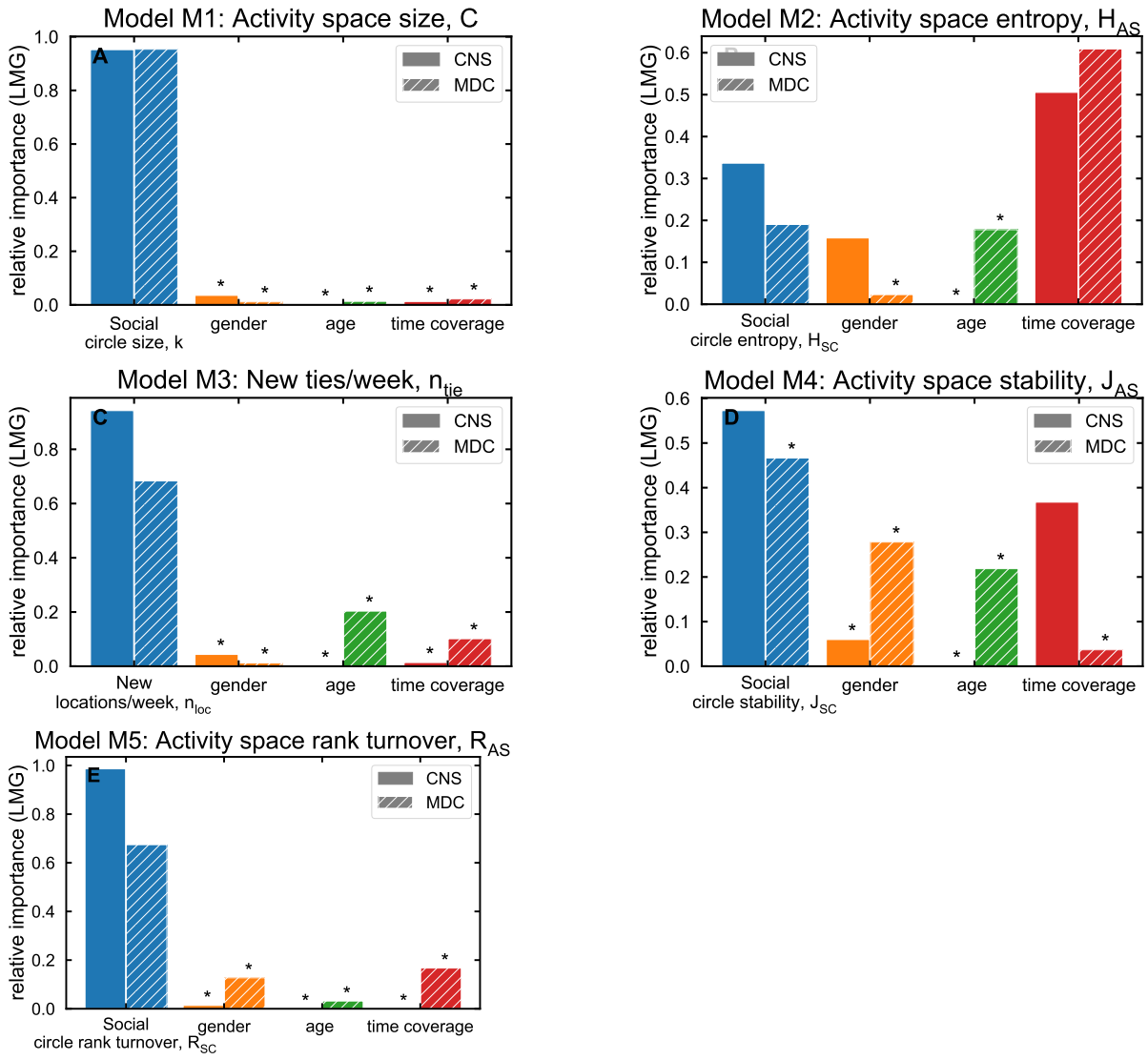

Figure S4: **T=30, Relative importance of regressors** LMG of each regressor computed using the Lindeman, Merenda and Gold method for models M1 (A), M2 (B), M3 (C), M4 (D) and M5 (E). Plain bars show results for the CNS dataset, dashed bars for the MDC dataset. Variables that are not significant in the regression model are marked with *.

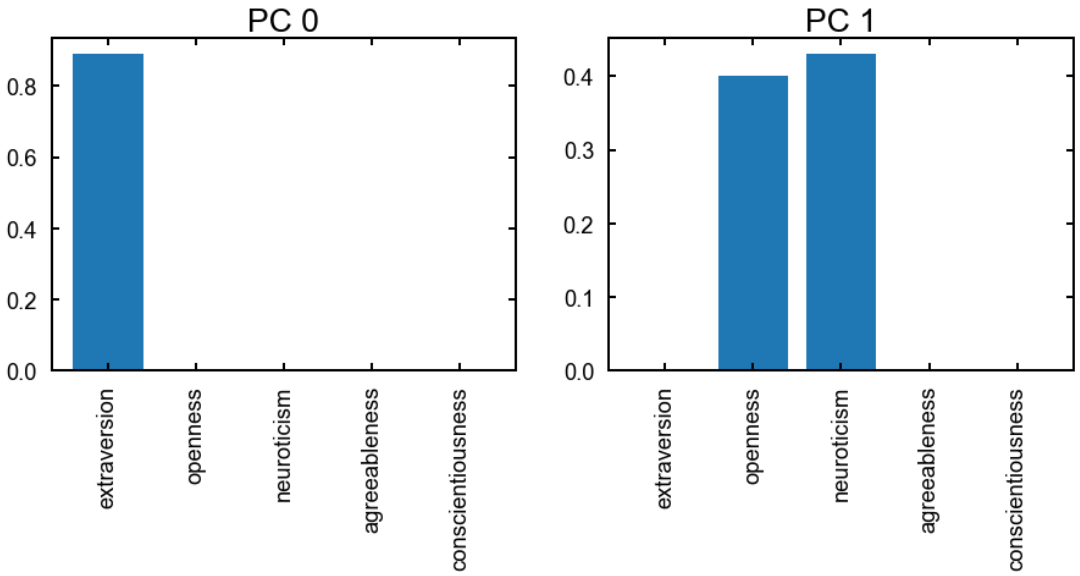

Figure S5: **T=30, Relative importance of personality traits for socio-spatial behaviour** LMG of each regressor computed using the Lindeman, Merenda and Gold method for the multiple regression model of the principal components (table S7).

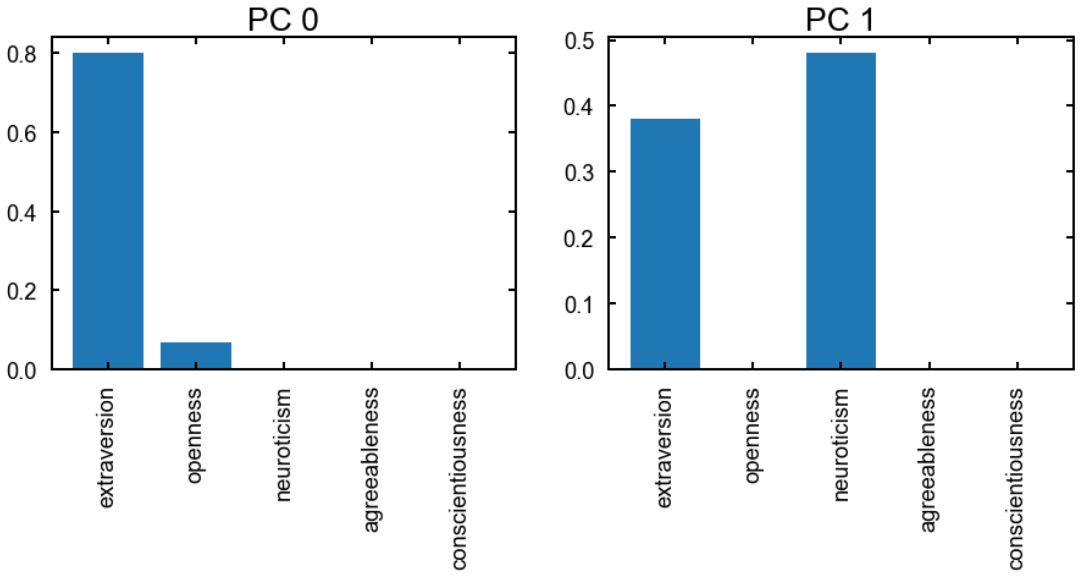

Figure S6: **T=30, Relative importance of personalitry traits for spatial behaviour** LMG of each regressor computed using the Lindeman, Merenda and Gold method for the multiple regression model of the principal components (table S10).

	$\overline{d_{self}}$	$\overline{d_{ref}}$	$\overline{d_{self}(i) < d_{ref}(i,j)}$
Social circle size, k	0.04 ± 0.13	15 ± 6	100%
Activity space size, C	0.04 ± 0.07	8 ± 3	99%
New locations/week, n_{loc}	0.06 ± 0.12	0.9 ± 0.5	96%
New ties/week, n_{tie}	0.1 ± 0.2	1 ± 1	95%
Social circle entropy, H_{SC}	0.002 ± 0.005	0.7 ± 0.3	99%
Activity space entropy, H_{AS}	0.002 ± 0.006	0.4 ± 0.1	99%
Social circle stability, J_{SC}	$(6 \pm 15) \cdot 10^{-4}$	0.14 ± 0.05	100%
Activity space stability, J_{AS}	$(6 \pm 11) \cdot 10^{-4}$	0.10 ± 0.04	100%
Social circle rank turnover, R_{SC}	0.04 ± 0.11	2 ± 1	99%
Activity space rank turnover, R_{AS}	0.04 ± 0.20	2 ± 1	99%

Table S3: **T=30, CNS dataset: Persistence of social and spatial behaviour.**For each of the social and spatial metrics, $\overline{d_{self}}$ is the average self-distance and $\overline{d_{ref}}$ is the reference distance between an individual and all others, averaged across individuals. The third column reports the fraction of cases where $\overline{d_{self}(i) < d_{ref}(i,j)}$, averaged across the population.

Model M1: Activity space size, C	coeff	p val	LMG
Social circle size, k	4 ± 0	$< 10^{-46}$	0.95
gender	-0.3 ± 0.2	0.2	0.04
time coverage	0.5 ± 0.2	0.05	0.01
$[R^2 = 0.32,\ F = 91.23,\ p_F = 0.0]$			
Model M2: Activity space entropy, H_{AS}			
Social circle entropy, H_{SC}	0.07 ± 0.02	$< 10^{-4}$	0.34
gender	-0.05 ± 0.02	$< 10^{-3}$	0.16
time coverage	-0.09 ± 0.02	$< 10^{-8}$	0.51
$[R^2 = 0.12,\ F = 26.82,\ p_F = 0.0]$			
Model M3: New ties/week, n_{tie}			
New locations/week, n_{loc}	0.58 ± 0.05	$< 10^{-30}$	0.94
gender	-0.09 ± 0.05	0.04	0.04
time coverage	0.03 ± 0.05	0.5	0.01
$[R^2 = 0.22,\ F = 55.56,\ p_F = 0.0]$			
Model M4: Activity space stability, J_{AS}			
Social circle stability, J_{SC}	0.027 ± 0.004	$< 10^{-9}$	0.57
gender	0.009 ± 0.004	0.02	0.06
time coverage	0.020 ± 0.004	$< 10^{-5}$	0.37
$[R^2 = 0.18,\ F = 30.32,\ p_F = 0.0]$			
Model M5: Activity space rank turnover, R_{AS}			
Social circle rank turnover, R_{SC}	0.81 ± 0.08	$< 10^{-19}$	0.99
gender	0.09 ± 0.08	0.3	0.01
time coverage	-0.001 ± 0.084	1.0	0.0
$[R^2 = 0.18,\ F = 31.70,\ p_F = 0.0]$			

Table S4: T=30, Linear regression models for the CNS dataset. For each model, we report the R^2 goodness of fit, the $F-test$ statistics with the corresponfing p-value p_F. We show the coefficients (coeff) calculated by the regression model, the probability (p val) that the variable is not relevant, and the relative importance (LMG) of each regressor computed using the Lindeman, Merenda and Gold method. Gender is a binary variable taking value 1 for females and 2 for males. For this dataset, age is not relevant as all participants have similar age.

	PC 0	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9
CNS	0.40	0.18	0.10	0.07	0.07	0.06	0.04	0.03	0.03	0.02
MDC	0.39	0.19	0.12	0.10	0.06	0.05	0.05	0.03	0.02	0.01

Table S5: **T=30, Variance explained by principal components.** The fraction of variance explained by each principal component for the CNS and MDC dataset.

	CNS		MDC	
	PC 0	PC 1	PC 0	PC 1
Social circle size, k	0.41	0.18	-0.36	0.04
Activity space size, C	0.42	-0.23	-0.40	-0.05
New locations/week, n_{loc}	0.33	0.27	-0.24	-0.35
New ties/week, n_{tie}	0.39	-0.11	-0.37	-0.22
Social circle entropy, H_{SC}	0.29	0.33	-0.36	-0.23
Activity space entropy, H_{AS}	0.38	-0.10	-0.35	0.13
Social circle stability, J_{SC}	-0.12	-0.50	-0.11	0.56
Activity space stability, J_{AS}	-0.06	-0.50	-0.03	0.62
Social circle rank turnover, R_{SC}	-0.17	0.26	0.28	-0.19
Activity space rank turnover, R_{AS}	-0.35	0.37	0.42	-0.18

Table S6: **T=30, Principal Components.** The weight of each metric in the first two principal components, for both datasets.

	PC 0 $R^2 = 0.17$, $F = 17.08$, $p_F = 0.0$			PC 1 $R^2 = 0.02$, $F = 2.05$, $p_F = 0.1$		
	coeff	p val	LMG	coeff	p val	LMG
extraversion	0.9 ± 0.1	$< 10^{-15}$	0.89	0.06 ± 0.07	0.4	0.04
openness	-0.18 ± 0.09	0.06	0.02	0.13 ± 0.07	0.05	0.4
neuroticism	0.1 ± 0.1	0.1	0.03	0.15 ± 0.07	0.04	0.43
agreeableness	0.05 ± 0.10	0.6	0.02	-0.04 ± 0.07	0.5	0.09
conscientiousness	0.04 ± 0.10	0.7	0.04	-0.03 ± 0.07	0.7	0.04

Table S7: **T=30, Extraversion, openness, and neuroticism explain socio-spatial behaviour.** The result of a multiple linear regression explaining principal components of socio-spatial data (Table S6). The value of each coefficient (coeff) is reported together with the probability (p val) that the coefficient is not relevant for the model. The relative importance of each coefficient (LMG) is computed using the LMG method.

	PC 0	PC 1	PC 2	PC 3	PC 4
CNS	0.57	0.21	0.10	0.08	0.04
MDC	0.55	0.24	0.12	0.06	0.03

Table S8: **T=30, Variance explained by principal components (only spatial data).** The fraction of variance explained by each principal component for the CNS and MDC dataset.

	CNS		MDC	
	PC 0	PC 1	PC 0	PC 1
Activity space size, C	-0.55	0.01	-0.55	-0.10
New ties/week, n_{tie}	-0.48	-0.16	-0.48	-0.35
Activity space entropy, H_{AS}	-0.48	-0.14	-0.44	0.20
Activity space stability, J_{AS}	-0.06	0.96	-0.02	0.88
Activity space rank turnover, R_{AS}	0.48	-0.16	0.51	-0.22

Table S9: **T=30, Principal Components (only spatial data).** The weight of each metric in the first two principal components, for both datasets.

	PC 0 $R^2 = 0.11$, $F = 11.55$, $p_F = 0.0$			PC 1 $R^2 = 0.02$, $F = 2.02$, $p_F = 0.1$		
	coeff	p val	LMG	coeff	p val	LMG
extraversion	-0.56 ± 0.09	$< 10^{-9}$	0.8	-0.12 ± 0.05	0.03	0.38
openness	0.20 ± 0.08	0.01	0.07	-0.04 ± 0.05	0.5	0.08
neuroticism	0.03 ± 0.08	0.7	0.07	-0.14 ± 0.05	0.01	0.48
agreeableness	-0.05 ± 0.08	0.5	0.03	-0.002 ± 0.052	1.0	0.01
conscientiousness	-0.005 ± 0.081	1.0	0.03	-0.03 ± 0.05	0.6	0.04

Table S10: **T=30, Extraversion, openness, and neuroticism explain spatial behaviour.** The result of a multiple linear regression explaining principal components of spatial data (Table S6). The value of each coefficient (coeff) is reported together with the probability (p val) that the coefficient is not relevant for the model. The relative importance of each coefficient (LMG) is computed using the LMG method.