REGULAR ARTICLE

Open Access

CrossMark

Inferring social influence in transport mode choice using mobile phone data

Santi Phithakkitnukoon[1*], Titipat Sukhvibul[1], Merkebe Demissie[2], Zbigniew Smoreda[3], Juggapong Natwichai[1] and Carlos Bento[4]

*Correspondence:
santi@eng.cmu.ac.th
[1]Department of Computer
Engineering, Faculty of Engineering,
Chiang Mai University, Chiang Mai,
Thailand
Full list of author information is
available at the end of the article

Abstract

A longitudinal mobile phone data that include both location and communication logs is analyzed to infer social influence in terms of ego-network effect in the commute mode choice. The results show that person's strong ties are more important to determine if driving is the person's transport mode choice, whereas weak ties are more important to determine if public transit is the person's choice. It is also evident from the results that social ties that are geographically closer are more influential for the commute mode choice than the ones who are farther away. For public transit, access distance is also one of the influential factors. The portion of transit users decreases as the access distance becomes larger. Moreover, social network is shown to influence the commute mode choice, as the likelihood of choosing a particular mode choice rises with the portion of social ties choosing that specific mode.

Keywords: social influence; transport mode choice; mobile phone data analysis

1 Introduction

Experiencing a range of trends such as urbanization, globalization, scarcity of resources, and technological advancements has been a global phenomenon. These lines of changes influence how planners address problems and evaluate solutions. The field of transportation is also experiencing a paradigm shift and focuses on applying all-inclusive multimodal approach and demand management solutions to reduce private car dependence and increase the efficiency and sustainability of public transit systems [1].

In the developed countries, the use of private vehicle constantly grew until recently. To accommodate the increased vehicle travel resulting from private vehicle growth, public authorities followed an approach that nowadays is financially unsustainable such as expanding the urban road network. In recent years, trends related to population aging, rising fuel prices, urbanization, rising health and environmental impacts, and changing consumer preferences cause the private car travel reaching its peak [2], which stimulated an increasing demand for alternative modes [1].

A range of transportation demand management measures have been examined to understand their effect in reducing private car use. Some of these measures are intended to improve the attractiveness of alternative transport modes e.g., provision of free bus travel for a specified time period, and other measures are intended to limit private car use such

Springer

as temporary changes to infrastructural conditions e.g., road closures [3], nevertheless recent experiences show that people are willing to use alternative modes such as walking, bicycling, and public transit. To meet the increasing travel demand for alternative modes, there is a need to improve the quality of the services in terms of convenience, comfort, affordability, and integration. This cannot be done without understanding the travelers' needs and their preference of using the alternative modes. One aspect that should be improved is mode choice modeling, which is considered essential for predicting the future growth for each mode, in addition to identifying factors that influence the use of each mode and shifting from one mode to another.

Mode choice models can be aggregated if they are based on aggregated zonal (and interzonal) information. It can also have disaggregated models if they are based on household and/or individual data. Disaggregate models consider that the demand is the result of several decisions of each individual traveler. The two most common discrete-choice models, multinomial logit and nested logit, have been used to investigate factors influencing travel behavior [4, 5]. The theoretical base of the discrete-choice model is that individuals travel choice is based on the utility maximization principle or relative attractiveness of competing alternatives. A discrete-choice model predicts a probability made by an individual as a function of any number of factors that describes the alternatives [4, 5].

Discrete-choice models can be based on either observed behavior (Revealed Preferences data) or on hypothetical choice surveys (Stated Preference data) that contains datasets of three main categories, which are believed to be important to influence mode choice: (i) characteristics of the trip maker - car ownership, possession of driving license, household structure, income, etc., (ii) characteristics of the journey - trip purpose, time of the day, and if trip is taken alone or with others, and (iii) characteristics of the transport facility - components of monetary cost, in-vehicle and out-of-vehicle times, parking, comfort and convenience, safety, security, etc. [5].

Discrete-choice models are based on the economic theories under the assumption that people choose between transport modes according to rational principles. The basic assumption is that people attempt to maximize the utility derived from the alternatives' different characteristics against each other and decide rationally. However, these economic theories and methods lack some factors from social relationships that influences people's transport mode choice. Social influence in transportation is becoming a topic of increasing coverage [6]. Social network theory acknowledges the role of social contagion and spreading mechanisms in the process of decision making, and social relationships may directly influence the making of different choices, such as transport mode. For example, when more members of an ego-network use a particular mode of transport, it is more likely that a given member of that network uses that mode as well, which may be due to conformity, social pressure, or imitation mechanism. Thus, the threshold models of influence [7] describe the influence of others as a function of the proportion of other people adopting a new type of behavior. On the other hand, network homophily can be responsible of imitation process by which people may modify their behaviors to bring them more closely into alignment with the behaviors of their friends [8, 9].

In this study, we use mobile phone data for the following two purposes: (i) to infer transport mode, and (ii) to investigate the influence of social network variables on transport mode choice. Our analysis focuses on the commuting trips, i.e., travels between residence

and workplace locations. A complementary work on social network influence on leisure travels can be found in [10].

2 Related work

In recent years, there have been studies that explored the use of cellular network data for different purposes. These studies investigated various topics such as traffic estimation [11–14]; origin-destination flow estimation [15–19]; travel demand analysis [20–22]; land-use detection [23–25]; interplay between users' mobility, location, and applications they access [26]; and place-related context inference [27, 28]. There is also a great body of literature that have analyzed the use of cellular network data in a wide range of issues, such as incident and traffic management [29]; urban sensing [30]; social networks, security, and privacy issues [31]. Our main concern of the present paper is to explore the potential of mobile phone data to infer the transport mode and, in particular, to analyze the influence of social network variables on transport mode choices. The remaining part of the related work section will be devoted to survey previous studies that have used mobile phone data for social network analysis.

Social influence can be readily observed in common collective decision processes, e.g., political polls [32], panic stampedes [33], stock markets [34], cultural markets [35], aid campaigns [36], product rating [37], or answering questions [38]. Some of these collective decisions can trap a population in a suboptimal state, for example a financial bubble due to financial actors' herding behavior [39]. Alternatively, they may steer a system into positive directions, such as increased tax compliance rates [40] and improved weight loss progression [41]. However, understanding how such collective decisions are formed, evaluating their benefit for the population, and even directing their outcomes, is conditional on quantifying how people perceive and respond to social influence.

2.1 Social influence on travel behavior

The idea that participation in social activities may affect travel decisions is not new in transportation research. For example, the need for belongingness was identified as a motivator for travel by Salomon [42]. Conceptually, time geography research (e.g., [43–46]) has long emphasized the importance of social contact, in the form of coupling constraints, as a key determinant of travel. There is evidence that shows a connection between social interaction and amount of travel. Harvey and Taylor [47] conducted a study using 1992 Canadian time-use data and discovered that employees who work from home spent less time with others. Their results also showed that these employees spent as little as 17% of their time (when awake) with others as compared with 50% for individuals in the conventional workplace. Harvey and Taylor also found that people who had very little social interaction with others tended to travel more. These results suggest that working from home may not reduce travel for everyone, but just alter the purpose of travel, which is also in line with the study by Arentze and Timmermans [48]. Based on these findings, Harvey and Taylor [47] discussed the necessity of achieving a better understanding of social contact, and in particular identifying the social relationship. Previous transportation research has also investigated the role of social organization (including household), socio-demographics, and gender on travel behavior, often at an aggregate level (e.g., [49–52]) but also at the individual level (e.g., [53, 54]), however considering only the characteristics of the individual, and not his or her connections to others. Nonetheless, relatively little work

has been done to operationalize the influence of these connections, and more specifically of the decisions made by others, on individual decision making. Ben-Akiva and Lerman [55] (p.33), for example, noted that "(by) considering a group of persons as a single decision maker… it is possible to abstract partially the complex interactions within… a household or firm." Although this approach, commonly adopted in travel behavior research, is a useful first-hand approximation to complex problems, it ignores important aspects derived from interpersonal interactions. More recent research, on the other hand, has started to tackle some of the additional complexity involved in dealing with social interdependencies, as seen in the small but growing literature on social relationships and interactions in travel decision making (e.g., [56–59]). Besides a small number of examples in this particular area of research, other aspects of social influence continue to be ignored in the analysis of travel behavior, such as social tie strength and socio-geography [60].

2.2 Mobile sensing approach in behavior analysis

Today's mobile phone is not just a communication device but also a new gateway for human behavior sensing - a new sensor-networking paradigm that incorporates human as part of its sensing infrastructure. The mobile sensing approach allows researchers to collect and analyze human behavior on a large-scale. Human mobility is one of the hottest topics in recent research. Song et al. [61] studied the mobility patterns of anonymous mobile phone users and concluded that, despite the common perception that our actions are random and unpredictable, human mobility follows surprisingly regular patterns and is 93% predictable. Their result is in line with González et al. [62] who showed that while most individuals travel only short distances and a few regularly move over hundreds of miles, they all follow a simple pattern regardless of time and distance, and they have a strong tendency to return to locations they visited before. The statistical properties of human mobility have been studied [63] and used in applications such as location prediction [64] and interurban analysis [65], for example. Epidemiology also benefits from understanding human movements, for example Wang et al. [66] showed that human mobility can be used to model fundamental spreading patterns that characterize a mobile virus outbreak, and Wesolowski et al. [67] studied how human travel patterns contribute to spread of Malaria. Mobile sensing approach has also been used in studies of social network structure (e.g., [68, 69]). Phithakkitnukoon and Dantu [70] discovered the scaling ratio in social structure. Other studies show that social structure can change over time due to a number of reasons such as migrations [71] or behavior adaptation as described in a study by Eagle et al. [72] that individuals change their patterns of communication to increase the similarity with their new social environment. Eagle et al. [73] later showed that social diversity is associated with economic development. Social interaction is, therefore, important not only for improving well-being but also economic status. Phithakkitnukoon et al. [74] found that weather condition can also influence social interactions, while Onnela et al. [75] and Krings et al. [76] showed that social interactions are constrained by geographical distance. Occasionally interacting with social ties may appear indicative of face-to-face meeting travel [49, 77]. Preliminary work by De Domenico et al. [78] showed that information gathered from social ties can help improve prediction of individual locations. Phithakkitnukoon et al. [60] discovered correlations between people's travel scope and locations of their social ties. It is therefore important to understand the interplay between social network and human mobility. However, fairly little work has been done to

explore this relationship. It is the dimensions and qualities of this kind of relationship that this study seeks to explore and define.

3 Methodology

3.1 Subjects

This study has taken an opportunistic sensing approach for real-life behavior observation, as human subjects are also part of the sensing infrastructure. For billing purposes, telecom operators keep records of their mobile phone customers' usage logs. Each of these usage logs, also known as Call Detail Records (CDRs), includes a caller ID, callee ID, caller's connected cell tower ID, callee's connected cell tower ID, duration of the call, and timestamp. Each time the mobile phone user makes or receives a call, i.e., connecting to a cell tower, the nearest cell tower location is recorded. Thus, collectively CDR data provides fine-grained longitudinal information about the individual's mobility and sociality. In this study, we capitalized on the opportunity that this kind of data can provide such a detailed behavior observation concerning mobility and sociality. We used anonymized CDR data of mobile phone users in Portugal over the period of one year.
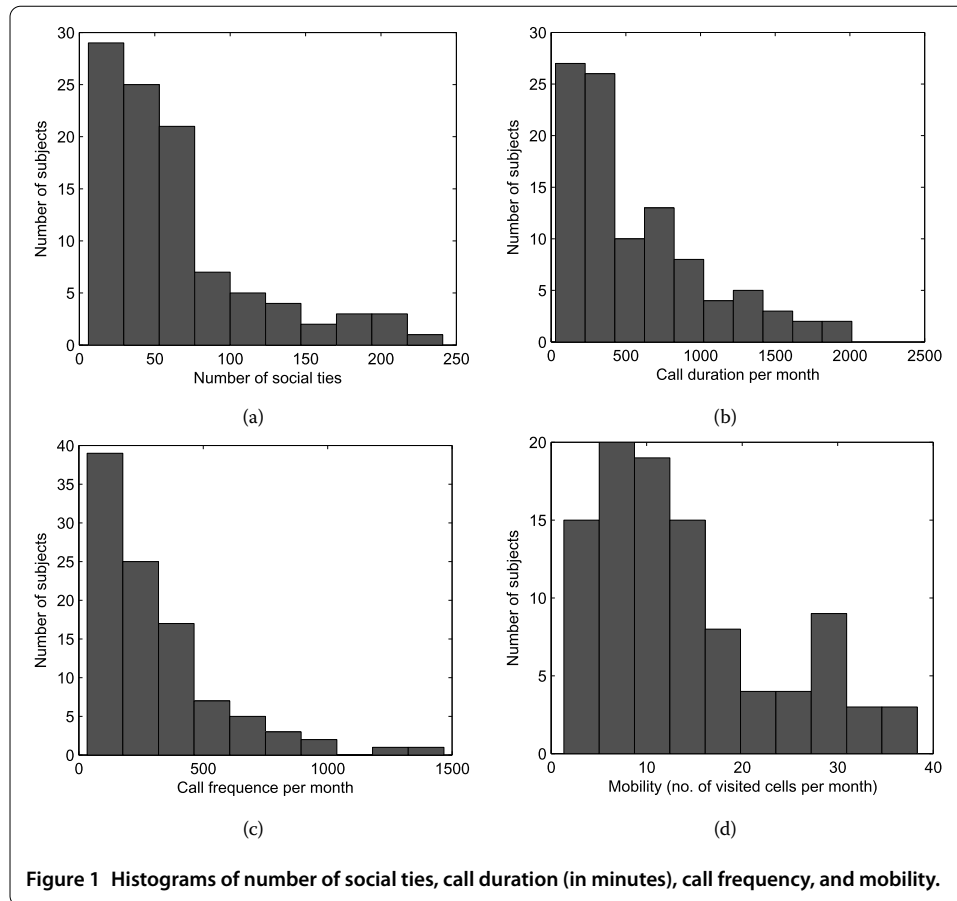
To safeguard personal privacy, individual phone numbers were anonymized by the operator before leaving their storage facilities, and were identified with a security ID (hash code). The dataset does not contain information relating to text messages (SMS) or data usage (Internet). There is a total of 6,509 cell towers, each on average serves an area of 14 km^2, which reduces to 0.13 km^2 in urban areas such as Lisbon and Porto.

For the subjects of our study, we started with an initial random sample of 100 mobile phone users (whose locations were recorded at least five times each month over the period of one year) as the *ego-subjects*, and followed the approach by Onnela et al. [69], to gather their *alters* (i.e., social ties) based on reciprocal communications. This snowball sampling-like subject recruitment gave us 5,305 alters, so there was a total of 5,405 subjects for the study. This is an egocentric network with the depth of 1.

On average, over one year of observation, there are 63.8 reciprocal links per ego-subject (i.e., number of social ties), each ego-subject spent 307.12 minutes on the mobile phone each month (approximately 18 minutes daily) and was connected to cellular network 307.12 times (calls) monthly (approximately 10 calls daily) across 168.65 different cell tower locations (14.05 cells monthly). Histograms of the number of social ties, call duration, call frequency, and mobility (number of different cell sites visited) are shown in Figure 1(a), 1(b), 1(c), and 1(d), respectively.

3.2 Residence and work location inference

In this study, we were interested in travel mode choices for commuting i.e., travel between one's place of residence and place of work (or study), which accounts for a large portion of all person trips. So, our first task was to identify the place of residence for each ego-subject. Since the location information of mobile phone users is at the level of cell tower location, so we estimated the place of residence for each ego-subject by taking our previous and successful approach in identifying residence location as the location of the most frequently used cell tower during the nighttime (10pm-7am). This approach has been proven as a reasonable approximation, as its result was validated against the actual census information in our previous study where we compared the portion of the estimated population density based on our residence location estimation with the actual population density across all 308 municipalities of Portugal [60]. Figure 2(a) shows the estimated residence locations

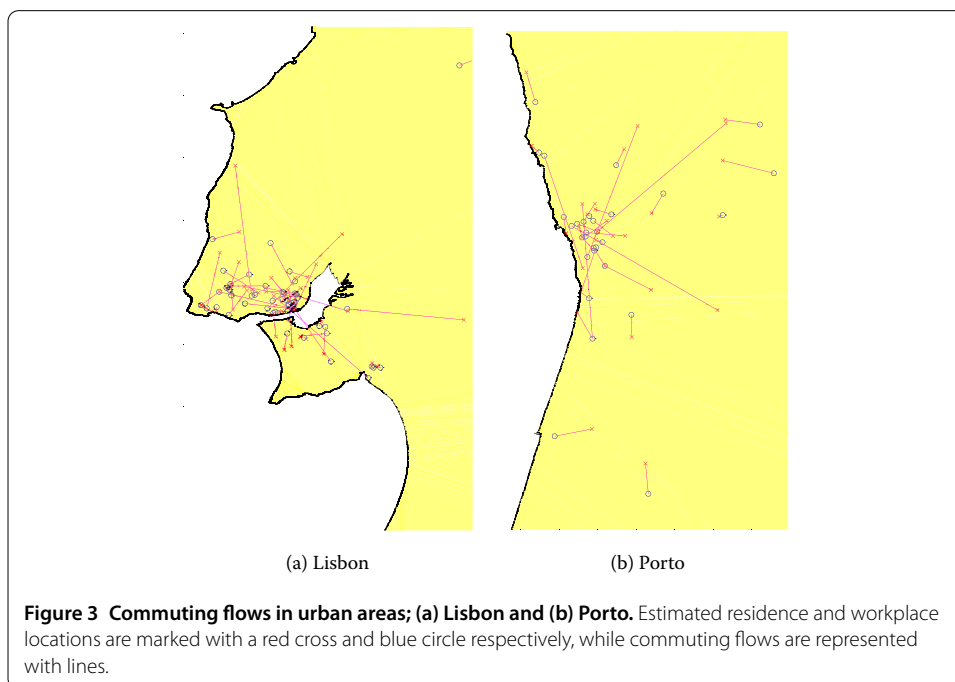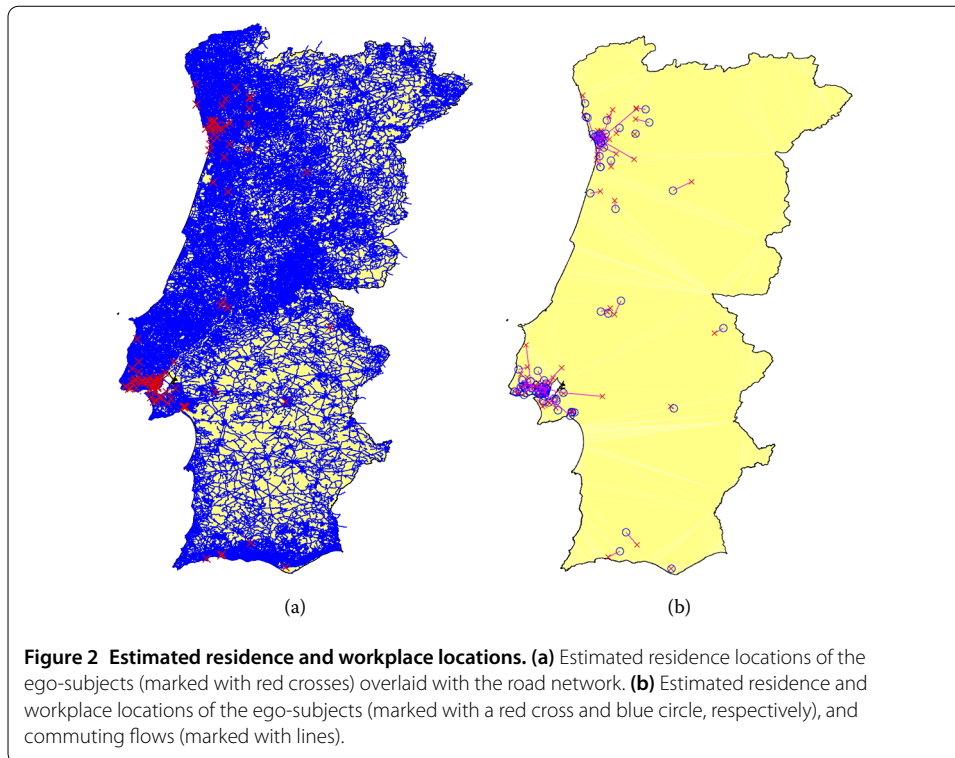**Figure 1 Histograms of number of social ties, call duration (in minutes), call frequency, and mobility.**

of our 100 ego-subjects overlaid with the road networks of Portugal. It can be seen that the residence locations of our ego-subjects are spread geographically across the country, while highly clustered in urban areas, such as Lisbon and Porto - intuitively, in coherence with the general population density distribution of the country.

By the same token, we took this approach in estimating the workplace location for each ego-subject as the location of the cell tower with the highest level of call activity during normal business hours (9am-5pm) on weekdays. Figure 2(b) shows the estimated locations of residence and workplace of our ego-subjects. The residence is marked with a red cross while the workplace is marked with a blue circle, and both places are linked with a line to represent a *commuting flow*. As commuting flows are denser in urban areas, Figure 3 shows a zoom-in version of commuting flows in Lisbon and Porto. (Note that the results are identical by considering weekdays and weekend altogether.)
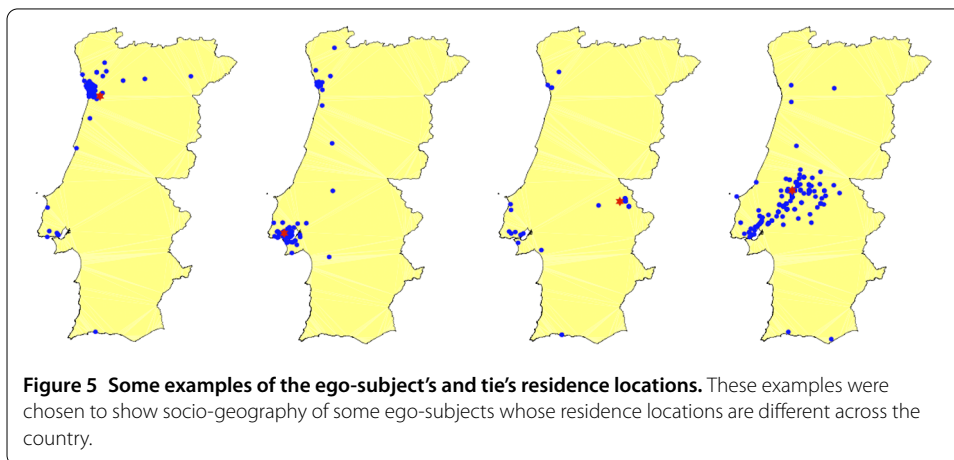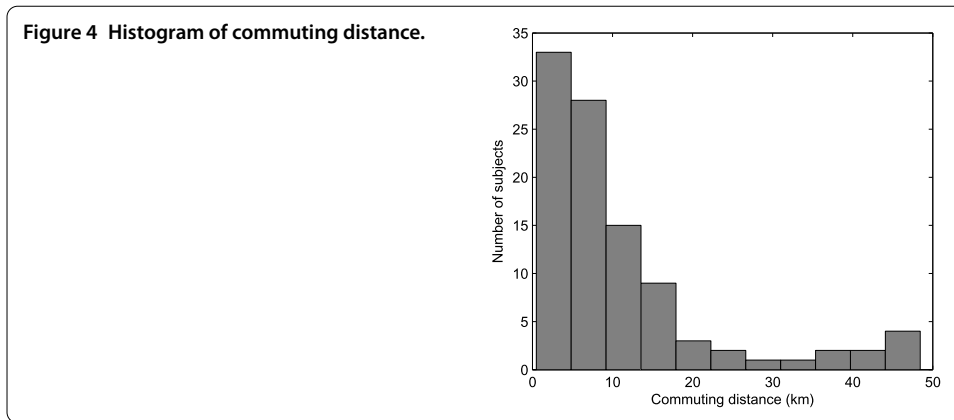
With the estimated residence and workplace locations, we further computed the commuting distance for each ego-subject. The average commuting distance was 10.84 km (median distance was 7.12 km). The longest commuting distance is 48.44 km while the shortest is 0.49 km among the ego-subjects. A histogram of the commuting distances is shown in Figure 4.

3.3 Social tie strength inference

In addition to the location information, CDRs also include call logs that carry the information about individual's social activity in the cellular network, from which social network
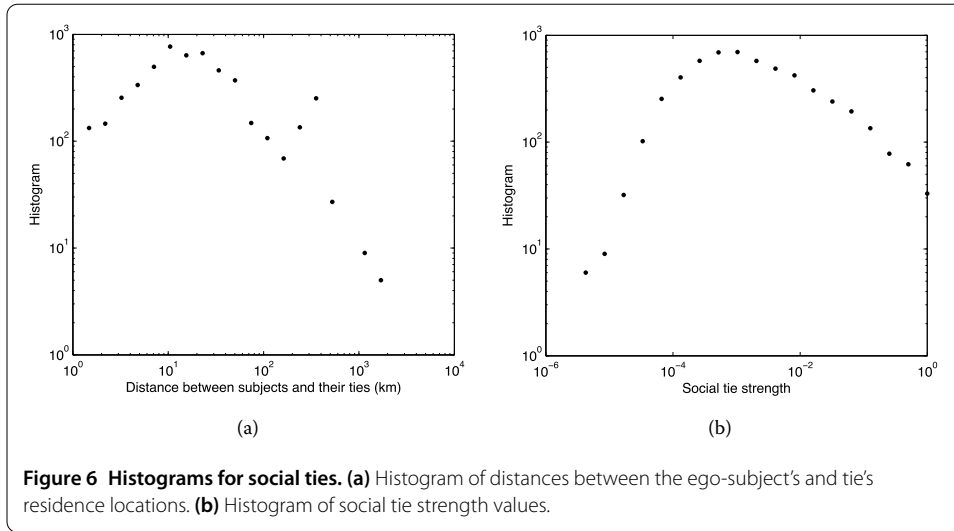
Figure 2 Estimated residence and workplace locations. (a) Estimated residence locations of the ego-subjects (marked with red crosses) overlaid with the road network. **(b)** Estimated residence and workplace locations of the ego-subjects (marked with a red cross and blue circle, respectively), and commuting flows (marked with lines).

Figure 3 Commuting flows in urban areas; (a) Lisbon and (b) Porto. Estimated residence and workplace locations are marked with a red cross and blue circle respectively, while commuting flows are represented with lines.

information can be extracted. Each ego-subject's social network was identified based on the caller IDs that were associated with the ego-subject through reciprocal calls. In other words, any individuals who had received and made calls to/from the ego-subject were considered the ego-subject's social tie (or alter). As shown previously, a histogram of number of social ties per ego-subject is illustrated in Figure 1(a). To get a geographical sense about

Figure 4 Histogram of commuting distance.

Figure 5 Some examples of the ego-subject's and tie's residence locations. These examples were chosen to show socio-geography of some ego-subjects whose residence locations are different across the country.

the ego-subjects and their ties' locations, Figure 5 shows a few examples of geographical distribution of the residence locations of the ego-subject and his/her social ties. The aforementioned approach was used to infer the social tie's residence location. Ego-subject's residence location is marked with a red star while the tie's residence location is marked with a blue dot. As reported in [35], in most cases, social ties are geographically clustered near the ego-subject's location. From all ego-subjects, the average distance between the ego-subject and social tie is 39.95 km, with the minimum of 0 km and the maximum of 1,701.74 km. A histogram of distances between the ego-subject's and tie's residence locations is shown in Figure 6(a). Due to its statistical distribution, a logarithmic scale was used to show the distribution of the histogram. It can be observed that there is a peak at around 10 km, and another peak at 300 km, which is an approximate distance between Lisbon and Porto.

In a social network, normally there are different levels of closeness in relation that defines the strength of social tie. Blondel et al. [31] surveyed different methods that have been used to set a meaningful metric to measure the importance of a link between ties. So, we further inferred the social tie strength by adopting the theory of tie strength developed by Mark Granovetter in his 1973 milestone paper [79]. He defined the strength of a tie as "a combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services". We took a similar approach to Onnela et al. [69] by using the amount of time spent in communication and reciprocity as proxies. We computed the tie strength between the ego-subject and a tie based on a total call duration

Figure 6 Histograms for social ties. (a) Histogram of distances between the ego-subject's and tie's residence locations. **(b)** Histogram of social tie strength values.

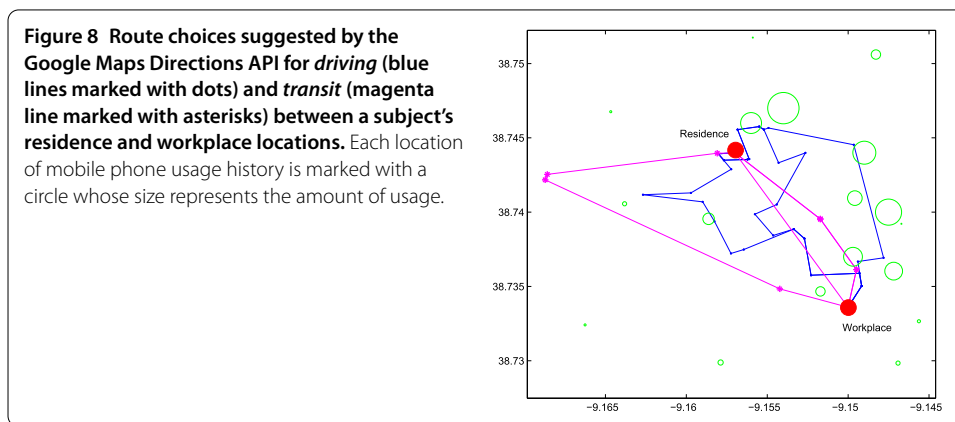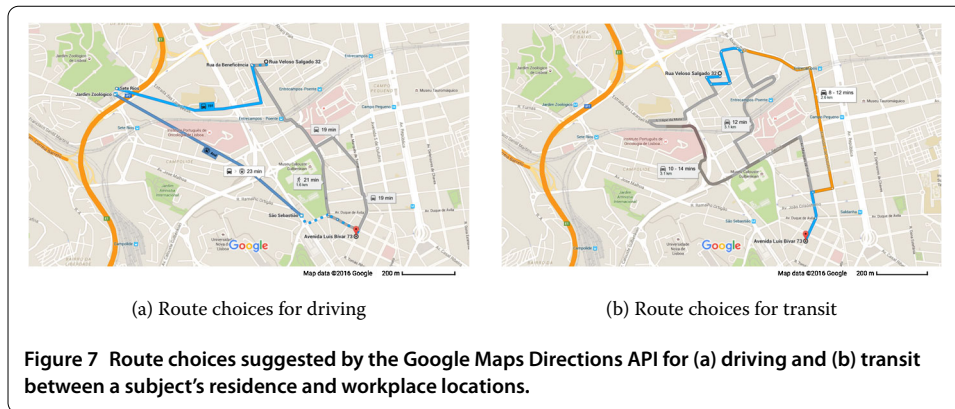between them, normalized by the total call duration between the ego-subject and all ties as given by Eq. (1).

$$s(i) = \frac{c(i)}{\sum_{i=1}^{N} c(i)}, \tag{1}$$

where $s(i)$ is the tie strength between the ego-subject and the ith tie, $c(i)$ is the total call duration between the ego-subject and ith tie, and the denominator is the total call duration between and the ego-subject and all associated ties where N is the total number of associated ties.

Thus, the value of tie strength ranges from 0 (lowest strength) to 1 (highest strength). Figure 6(b) shows a histogram of social tie strength across all ego-subjects and their ties. Similar to Figure 6(a), a logarithmic scale was used because of the nature of statistical distribution of the social tie strength values. The average tie strength is 0.018 while the minimum and maximum values are 2.13×10^{-16} and 0.99, respectively.
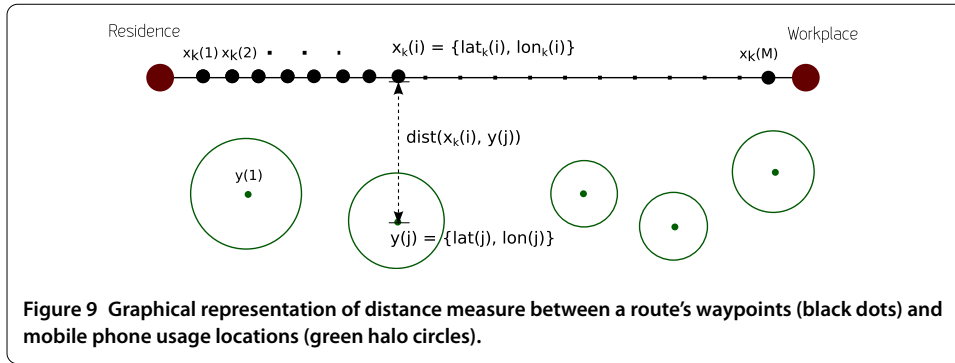
3.4 Transport mode inference

With the estimated residence and workplace locations, we further inferred about the subject's mode choice for commuting, based on the mobile phone usage history. We used the Google Maps Directions API [80] as a tool to help us make a sensible inference about the commute mode choice. With the Google Maps Direction API, for each subject, we made an HTTP request for *waypoints* by using residence and workplace locations as the *origin* and *destination* parameters. Public transportation and private car are the two most common modes of today's transportation. The influential factors and attitudes towards the choice of these modes have been the subject of both transport and behavioural science studies [81–83]. In Portugal, there have been some recent efforts by public authorities at central and local levels to promote the provision of soft modes (e.g., walking and cycling), but the implementation of a soft mobility network is in its early stage. This results in a rare use of soft modes for daily commuting, particularly in urban areas [84]. Therefore, we considered *driving* and *public transit* as the transport mode choices in this study, so for each subject, two HTTP requests were made regarding the mode choice. Along with the

(a) Route choices for driving (b) Route choices for transit

Figure 7 Route choices suggested by the Google Maps Directions API for (a) driving and (b) transit between a subject's residence and workplace locations.

Figure 8 Route choices suggested by the Google Maps Directions API for *driving* (blue lines marked with dots) and *transit* (magenta line marked with asterisks) between a subject's residence and workplace locations. Each location of mobile phone usage history is marked with a circle whose size represents the amount of usage.

origin and *destination* parameters, "driving" and "transit" were used for *mode* parameter. The response from each HTTP request includes *routes* i.e., route choices between origin and destination by each given transport mode (i.e., driving and transit). The number of route choices returned from the request varies with the given origin and destination locations. These route choices were possible route choices suggested by the Google Maps Directions API. Figure 7 shows an example of route choices between residence and workplace locations of one of the ego-subjects. There are three route choices for each given transport mode in this example.

With these suggested routes, our task here was to identify the most probable route taken by the subject, based on the traces of the subject's mobile phone usage. Each response *routes* (suggested route) that we obtained from our HTTP request includes *legs*, which are waypoints (or geographic points) along the route. Figure 8 shows the suggested routes (according to the obtained *legs* information) for both driving and transit (three choices for each mode, in this example) between a subject's residence and workplace locations, along with the subject's mobile phone usage history locations, each marked with a green halo circle whose size corresponds to the amount of usage (the number of connections to the cellular network) - i.e., a larger size means a higher number of connections, which can imply a more frequent visited location. Suggested driving routes are in blue while transit routes are in magenta. Residence and workplace locations are marked with solid red circles.

Intuitively, the route that is geographically closer to the locations of mobile phone connectivity i.e., visited locations, is more likely to be the actual route taken by the subject. As an example (in Figure 8), the driving route choice on the upper right corner position of

Figure 9 Graphical representation of distance measure between a route's waypoints (black dots) and mobile phone usage locations (green halo circles).

the figure appears to be the most probable route taken for commuting, as the route lies in the area of frequent mobile phone usage locations. As the subject has collectively (over a year) connected to the cell towers along the route, which sensibly suggests that this route is likely to be the route taken by the subject - and hence driving is the mode choice.

Systematically, we interpolated the obtained routes so that each route has 100 waypoints to guarantee the fairness in comparison among different routes, regarding the distance to the mobile phone usage locations. The task here was to find the nearest route to the mobile usage locations. The number of data points (waypoints) of 100 for our data interpolation was arbitrarily chosen.

Computationally, for each route k (denoted by r_k); $r_k = \{x_k(1), x_k(2), x_k(3), \ldots, x_k(M)\}$, where $x_k(i)$ is the ith waypoint of route k and M is the number of waypoints (100), we computed the distance (m_k) between the M waypoints and a set of all N mobile phone usage locations, as following:

$$m_k = \frac{1}{M} \sum_{i=1}^{M} c(i), \tag{2}$$

where $c(i)$ is the average haversine distance ($\mathrm{dist}(x, y)$) [85] from the waypoint i to all N mobile usage locations, i.e.,

$$c(i) = \frac{1}{N} \sum_{j=1}^{N} \mathrm{dist}\big(x_k(i), y(j)\big) \cdot w(j). \tag{3}$$

Each waypoint $x_k(i)$ consists of geographic coordinates (latitude, longitude); $x_k(i) = \{\mathrm{lat}_k(i), \mathrm{lon}_k(i)\}$ while each mobile phone usage location ($y(j)$) also consists of geographic coordinates, i.e., $y(j) = \{\mathrm{lat}(j), \mathrm{lon}(j)\}$ (as shown in Figure 9), therefore $c(i)$ can be calculated as:

$$c(i) = \frac{1}{N} \sum_{j=1}^{N} 2r$$

$$\cdot \arcsin\left(\sqrt{\sin^2\left(\frac{\mathrm{lat}(j) - \mathrm{lat}_k(i)}{2}\right) + \cos\big(\mathrm{lat}_k(i)\big) \cos\big(\mathrm{lat}(j)\big) \sin^2\left(\frac{\mathrm{lon}(j) - \mathrm{lon}_k(i)}{2}\right)}\right)$$

$$\cdot w(j), \tag{4}$$

where $w(j)$ is a weight function that varies with the level of connectivity at the usage location j as follows:

$$w(j) = \frac{1}{f(j)} \sum_{n=1}^{N} f(n), \qquad (5)$$

where $f(j)$ is the number of connections (i.e., calls made or received) at the mobile usage location j, so a mobile usage location with a higher connectivity has a lower weight, and vice versa.
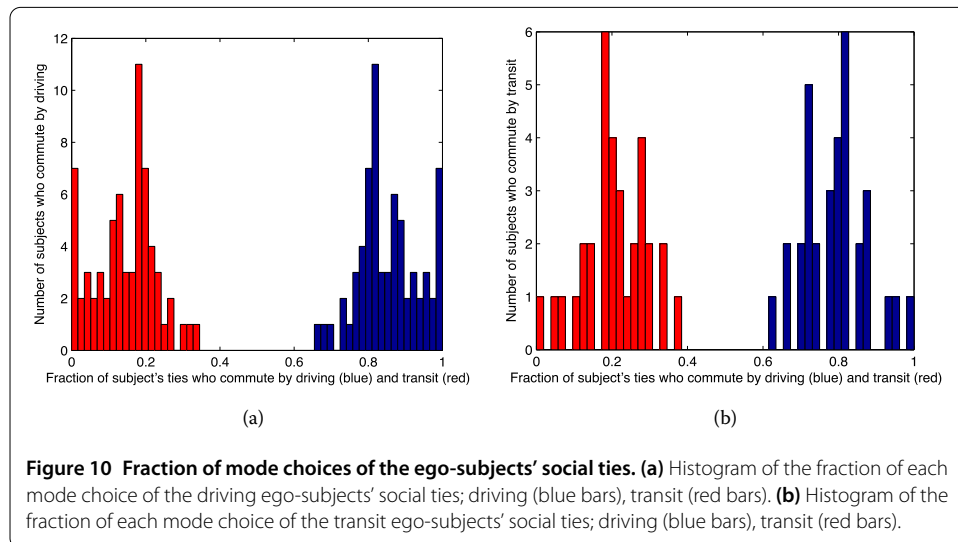
The distance m_k is computed for all suggested routes, and the route with the minimum m_k is then chosen as the most probable route taken by the subject, and the mode choice is determined accordingly. With this transport mode inference algorithm, we have found that for all 5,405 subjects, driving is the mode choice of 4,500 subjects while transit is the choice of 905 subjects, which account for 83.26% (driving) and 16.74% (transit). These percentages are in line with the actual survey data, which are 85% for driving and 15% for transit, reported by Eurostat [86] and ECORYS Transport of Portugal [87]. For our ego-subjects, driving is the choice of 67 subjects while transit is the choice of 33 subjects.

4 Results

4.1 Commute mode choices of social ties

The inferred commute transport modes and social ties allowed us to further investigate on the mode choices of the ego-subjects' social ties. We separated the ego-subjects into two groups according to their commute mode choices; driving and transit. As our initial interest was to observe the characteristic distributions of social ties' mode choices that were believed to influence the choices of the ego-subjects, so we inspected the mode choices the ego-subjects' ties for each ego-subject's group.

For each ego-subject, we examined the number of driving and transit ties from which the fraction of each tie's mode choice was calculated. Figure 10(a) shows a histogram of the fraction of each mode choice of the driving ego-subjects' social ties. The average and standard deviation of the fractions were calculated to be 0.855 and 0.828 for the driving

(a)　　　　　　　　　　　　　　　　　　　　　(b)

Figure 10 Fraction of mode choices of the ego-subjects' social ties. (a) Histogram of the fraction of each mode choice of the driving ego-subjects' social ties; driving (blue bars), transit (red bars). **(b)** Histogram of the fraction of each mode choice of the transit ego-subjects' social ties; driving (blue bars), transit (red bars).

ties, and 0.145 and 0.828 for the transit ties, respectively. It can be observed that there is a much higher fraction of social ties whose commute mode choice are driving, which may have influenced the mode choice decision for the driving ego-subjects.

On the other hand, by examining the mode choices of the transit ego-subjects' social ties, we found that the fraction of social ties whose choice is driving is still higher than that of the social ties with transit being their mode choice, as shown in Figure 10(b). The average and standard deviation are (0.791, 0.852) for driving ties, and (0.209, 0.852) for transit ties.

Since driving is the mode choice of the majority, its fractions appear to overwhelm the transit ties' fractions in both mode choice groups of the ego-subjects. Although, this is the case, it is still observed that there is a slight shift of these fractions where the average of the transit ties' fraction (0.209) is slightly higher than the driving ego-subjects' transit ties' (0.145), as well as a downshift in the driving ties' fractions for transit ego-subjects. This minor evidence may have softly suggested that the mode choice made by the members of the ego-subject's social network can influence the ego-subject's mode choice as well.

To ensure that this result is not entirely due to the unbalanced distribution of car and transit users, we re-generated the result with a random reference system that randomly shuffles the mode choices among people in the dataset. Firstly, we reshuffled the driving mode choices among the people and re-measured the average fractions of driving and transit social ties for each set of driving and transit ego-subjects. We repeated this experiment for 10 times. Secondly, we reshuffled the transit mode choices and re-measured the results. We've observed that the results obtained from this random reference system are different from the result in Figure 10 as there are overlaps (histograms shown in Tables 3 and 4 in the Appendix) and hence the difference between the average fractions of driving and transit ties (Tables 1 and 2 in the Appendix) are less than the results observed in Figure 10, which are (0.855, 0.145) for driving subjects and (0.791, 0.209) transit subjects. The average difference between the fractions of driving and transit ties from the random reference system is 0.587 for driving subjects and 0.527 for transit subjects, while the result in Figure 10 has the differences of 0.710 and 0.582 for driving and transit subjects, respectively - which are overall 12.3% and 5.5% greater than the random references. This may suggest that the result in Figure 10 is not entirely due to the unbalanced distribution of car and transit users but perhaps also the social influence.

There may be multiple realizations for why the mode choices of the ego-subjects resemble the mode choices of ego-subject´s ties for each ego-subject´s group. Pike [59] explained that other members in the ego-subject´s social network could influence the ego-subject. Alternatively, it may be that all members in the ego-subject´s social network have a similar choice context such as similar costs associated with each mode, and that may result in the same mode choice among the ego-subject's social network. It is also possible that members of the subject´s social network are all prone to make the same choice because of shared attitudes on transportation, and this preliminary finding has led us to our further investigation (presented in the later sections).

4.2 Social distance

The strength of tie defines the level of social closeness or relationship. Strong ties are people who are socially close to us and whose social circles tightly overlap with our own. Typically they are people we trust and with whom we share several common interests. On the contrary, weak ties represent mere acquaintances. Different strengths of social ties can

Figure 11 Fraction of social ties sharing a common commute mode choice with the individual ego-subjects, across a range of tie strength values.
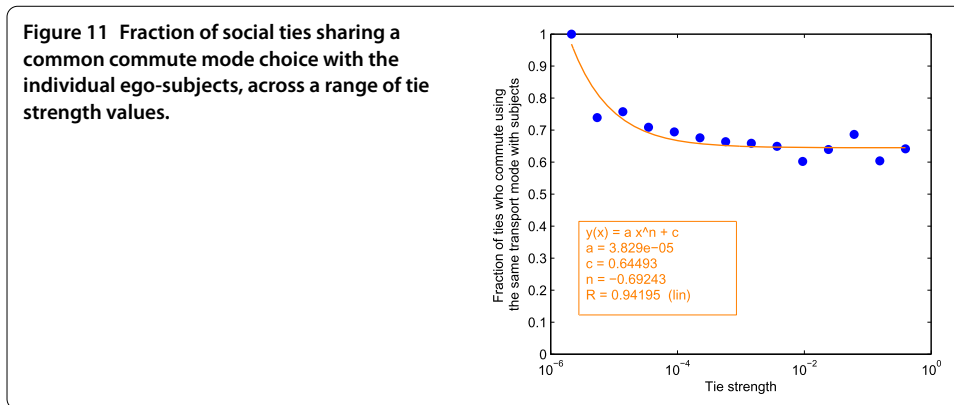
influence various behaviors differently, for example, receipt of information [88], mobility [60], and word-of-mouth referral [89].

Along this line of behavior understanding, here we were interested in how tie strength relates the transport mode choice decision. So, we extended our preliminary investigation on social ties' commute mode choices to examine the tie strength. In particular, we inspected the fraction of social ties who share a common mode choice with the individual ego-subjects, across a range of tie strength values (calculated using Eq. (1)).

Interestingly, as the strength of tie increases, the fraction of ties sharing a common mode choice decreases. This relationship can be fitted with a power law equation $y(x) = 3.829 \times 10^{-5}x^{-0.69243} + 0.64493$ with $r = 0.941$, as shown in Figure 11. The result may suggest that weaker ties tend to have a higher influence on the ego-subject's mode choice, as a higher fraction of them share a common mode choice with the ego-subjects.

To ensure that the observed result is not entirely due to the overwhelming fraction of weak ties, we re-produced this result by considering a network with the same number of connections, degree, and overall fraction of weak ties as originally, but with the transport modes reshuffled among people. We repeated the experiment for 10 trials. We've observed that the results obtained from this setup (shown in Table 6 in the Appendix) are rather random when compared with the result in Figure 11, which is more structured and can be nicely fitted by an equation.

We further examined two separate groups of the ego-subjects; driving and transit. A quite similar result is also observed for the driving ego-subjects (shown in Figure 12(a)), but not the same for the transit ego-subjects (Figure 12(b)). Interestingly, for driving ego-subjects, two distinct trends are observed. When the tie strength is less than the average tie strength, the portion of driving ties decreases as the tie strength increases ($y(x) = 0.0015311x^{-0.37327} + 0.81391$ with $r = 0.943$). However, when the tie strength is higher than the average, the portion of driving ties increases with the tie strength value ($y(x) = 0.63963x^{-0.031178} + 0.26833$ with $r = 0.975$).

The average tie strength has been used to determine strong ties and weak ties [60] where those with a tie strength value of greater than the average are classified as strong ties, and otherwise classified as weak ties. So, the result in Figure 12(a) suggests that as the tie strength increases, weak ties become less influential while strong ties become more influential for the driving ego-subjects. Note that our notion of being influential here means being likely that ego-subject is to share a common choice as reflected by the fraction of social ties' mode choices observed.
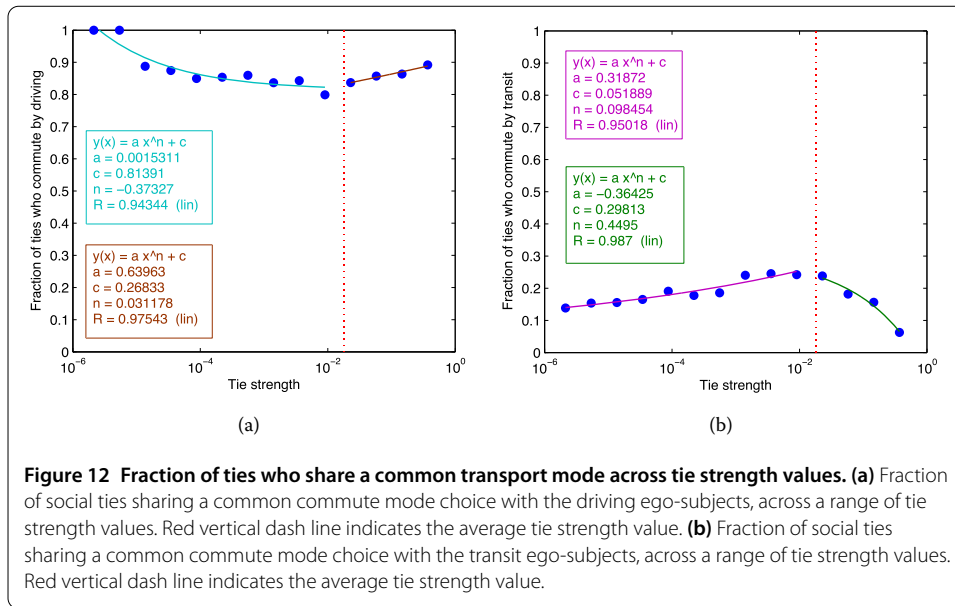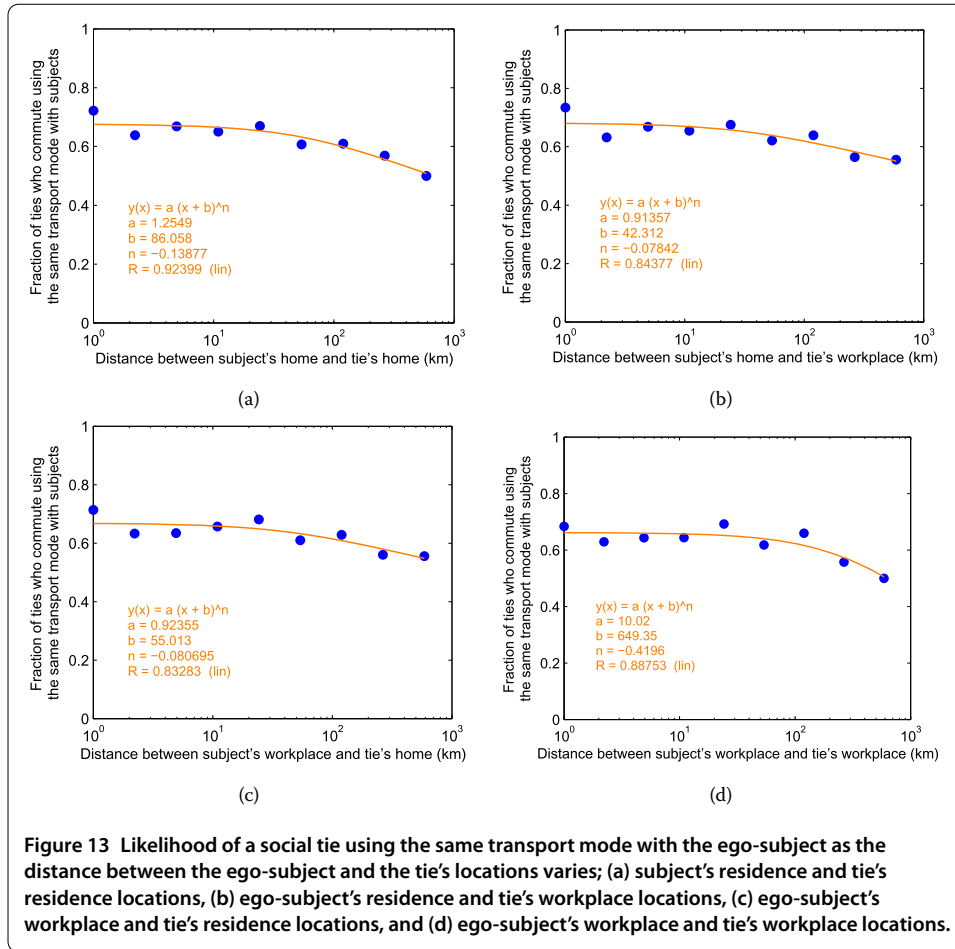
Figure 12 Fraction of ties who share a common transport mode across tie strength values. (a) Fraction of social ties sharing a common commute mode choice with the driving ego-subjects, across a range of tie strength values. Red vertical dash line indicates the average tie strength value. **(b)** Fraction of social ties sharing a common commute mode choice with the transit ego-subjects, across a range of tie strength values. Red vertical dash line indicates the average tie strength value.

The observed higher influence of weaker ties here follows the phenomenon described by Granovetter [79] that individuals are often influenced by others with whom they have tenuous or even random relationships - i.e., weak ties. Our observation of the similar level of influence across tie strengths is in line with a study of the impact of strong and weak ties on the acceptance of a new product by Goldenberg et al. [90] that concluded that the influence of weak ties is as strong as the influence of strong ties, and their effect approximates or exceeds that of strong ties, in all stages of the product life cycle. In addition, they also observed that when personal networks are small, weak ties were found to have a stronger impact on information dissemination than strong ties, which is likely the case for our result in Figure 12(a). Our result thus adds to the literature from the perspective of the influence of social ties on the choice of commuting by a car.

On the other hand, for the transit ego-subjects, opposite trends are observed (Figure 12(b)). As the tie strength increases, the influence of weak ties become more profound ($y(x) = 0.31872x^{-0.098454} + 0.051889$ with $r = 0.950$), whereas the influence of strong ties become less reflective ($y(x) = -0.36425x^{-0.4495} + 0.29813$ with $r = 0.987$). Thus, Figure 12 suggests that strong ties are more important to determine if driving is the person's transport mode choice, whereas weak ties are more important to determine if public transit is the person's choice. People commuting by transit may have a larger chance to create homophilic weak links, as they share space and spend some time together. Driver-commuters, on the other hand, are traveling either by themselves or with people who are presumably strong links. This result complements the study of de Kleijn [91], which found that stronger ties have a greater influence than weaker ties on choice of travelling by public transport, and it extends the observation with a remark of the influence that drops as the tie strength increases for the strong ties.

4.3 Physical distance

After having investigated on the influence of social distance on mode choices, here we were interested in exploring how geographical distance plays a role in mode choice de-

Figure 13 Likelihood of a social tie using the same transport mode with the ego-subject as the distance between the ego-subject and the tie's locations varies; (a) subject's residence and tie's residence locations, (b) ego-subject's residence and tie's workplace locations, (c) ego-subject's workplace and tie's residence locations, and (d) ego-subject's workplace and tie's workplace locations.

cision. In particular, as we've already observed that distance in social relationship is an influential factor, so we wanted to further explore if the physical distance to the social ties is also another influential factor for the mode choice decision. For instance, are friends who live (or work) nearby more influential than those who live (or work) farther away? It is the question concerning landscape and transport infrastructures that may influence the transport mode choice decision, as different geographical areas may be structured with different physical arrangements.

So, we examined the portion of ties who share a common mode choice with the individual ego-subjects across a range of geographical distance between them. Residence and workplace locations were considered. We observed the trend that as the distance increases, the portion of ties who share a common mode choice drops, conceivably (as shown in Figure 13). The drop is approximately 10% from 0 km to 600 km in all cases. The fitted curve equations and their corresponding correlation coefficient r are: $y(x) = 1.2549(x + 86.058)^{-0.13877}$ with $r = 0.924$ (between the ego-subject's and tie's residence locations), $y(x) = 0.9135(x + 42.312)^{-0.07842}$ with $r = 0.843$ (between the ego-subject's residence and tie's workplace locations), $y(x) = 0.92355(x + 55.013)^{-0.080695}$ with $r = 0.832$ (between the ego-subject's workplace and tie's residence locations), and $y(x) = 10.02(x + 649.35)^{-0.4196}$ with $r = 0.888$ (between ego-subject's and tie's workplace locations).
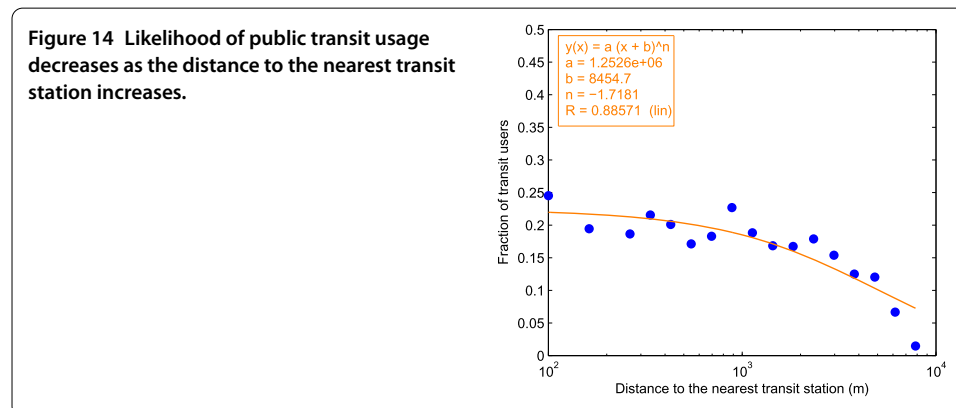
The observed results suggest that social ties who are geographically closer are more influential for the commute mode choice than the ones who are farther away. Plausibly, landscape and transport infrastructures may also play a role in transport mode choice as it circumscribes spatial arrangement and thus constraints for the area's transportation. So, for those who are within nearby areas, their transport mode choices tend to be more similar.
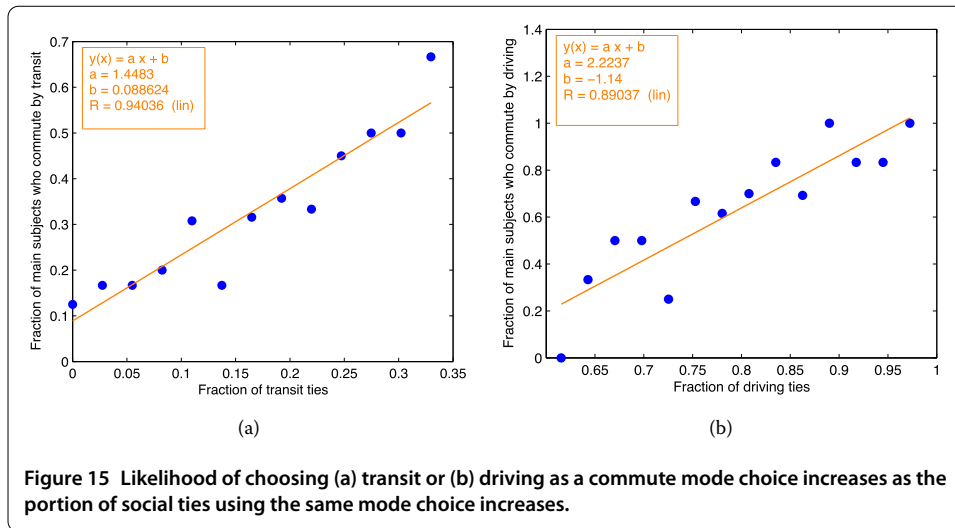
Previous studies approached the interplay of mobility patterns and social relationship from a different angle. Their primary objective was measuring the correlations between tie strength and mobility similarity and showed that mobility similarity can be used to classify social relationships [62, 63]. Our study offers another perspective that physical distance to the social ties, from working/living nearby to sharing/attending sets of similar locations, can be influential factor for the mode choice decision.

We continued to ask the question concerning the physical space or spatial arrangement, particularly in this case it refers to the public transit accessibility - i.e., how public transit infrastructure influences the decision of public transit usage.

Accessibility of public transit is important in evaluating existing services and predicting travel demands. Access distance [64] is one of the accessibility measures. So for each of our 4,405 subjects, we measured the distance to the nearest public transport station. We used the Google Places API via the Nearby Place Search [65] with "Type" parameter set to 'transit_station'. We then examined the fraction of the subjects who are transit users as the distance to the nearest public transit station varies from 100 m to 10 km. Expectedly, the result (in Figure 14) shows that as the access distance increases the portion of transit users decreases. This relationship can also be described by a fitted power law equation: $y(x) = 1.2526 \times 10^6 (x + 8{,}454.7)^{-1.7181}$ with $r = 0.888$. Hence, public transit access distance is also one of the influential factors for commute mode choice decision.

Our current analysis considers the accessibility as the ease of access to the public transport stations. Our future studies will also explore other aspects of public transport use, such as out-of-vehicle and in-vehicle-times, speed, directness of travel, and number of transfers for specific origin-destination connections, and so on. In a study conducted in Lisbon, Papaioannou and Martinez [92] emphasized on the primary role of accessibility in people´s mode choice decision. This however is not always the case, in major urban areas, where residents have access to variety of transportation options, distance can be a less influential factor and travellers tend to choose transport modes that provide better service rather than being nearby.

Figure 14 Likelihood of public transit usage decreases as the distance to the nearest transit station increases.

Figure 15 Likelihood of choosing (a) transit or (b) driving as a commute mode choice increases as the portion of social ties using the same mode choice increases.

4.4 Ego-network effect

We were back to the question of how our social network influences our mode choice decision. Do we have a tendency of having the same choice with our social network? Is a person likely to be a public transit user if most of people in the person's social network are public transit users? To find out about this ego-network effect, we scrutinized the fraction of number of social ties in each ego-subject's social network who share a common commute mode choice.

We inspected the fraction of social ties who share a common commute mode choice with the main subject. The result (in Figure 15) shows that the there is a higher likelihood of the subject using the same commute mode choice with his/her ties when the portion of the ties using the particular mode choice increases within his/her social network. It is the case for both mode choices: (a) transit and (b) driving. For transit, the trend can be described by a fitted linear equation $y(x) = 1.4483x + 0.088624$ with $r = 0.940$, while for the driving, the fitted linear equation is $y(x) = 2.2437x - 1.14$ with $r = 0.89$.

The result is a supporting evidence of homophily in social networks [8]. Our result suggests that people who are already connected within personal network can also influence one another's behavioral characteristics such as commute mode choice decision, as shown in this study. Our data does not allow us to go further in an explanation of psychological factors implied in the individual decision-making, however the influence of social expectations on intention and habits to use a car or public transport has been observed in psychological studies [93]. Our research indicates that the social influence can be particularly effective when originating in one's interpersonal relations.

5 Conclusions

The advance in information and telecommunication technologies such as personal mobile phones has opened up new exciting opportunities for behavior studies among which is the opportunistic behavioral sensing that incorporates human being as part of sensing infrastructure via cellular network. By analyzing a longitudinal mobile phone data that include both location and communication logs, we were able to infer transport mode choices of 4,405 mobile phone users that were considered as the subjects in our study, as well as investigate on the social influence on commute mode choice decision. Our results show

that strong ties are more important to determine if driving is the person's transport mode choice. On the other hand, weak ties are more important to determine if public transit is the person's choice. These results seem to be in line with the threshold model of social influence [7]. However, it indicates also that the influence of others can be differentiated according to their relation with ego. Not only all individuals are not equally influenced by others [94], but different links can be more or less influentials in the mode choice. Furthermore, we've observed that social ties that are geographically closer are more influential for the commute mode choice than the ones who are farther away. Our analysis also shows that the public transit access distance is one of the influential factors for the commute mode choice decision, as the results have pointed out that the portion of transit users decreases as the access distance increases. In addition, we've observed that the likelihood of choosing either transit or driving as a commute mode choice increases with the portion of social ties choosing that particular mode choice. Hence, social network evidently does influence transport mode choice.

There are nonetheless some limitations to the observations we present in this study. The first of these is the discontinuous nature of the location traces in our dataset. Since individuals were only located when connections with the cellular network were established, we were able to only identify a subset of all the locations visited in the course of a day. However, we believe that our aggregation of mobility patterns and the longitudinal nature of our data compensates for this. The second limitation is the coarse spatial resolution of the location traces, which is determined by the granularity of cellular tower coverage. Although traditional surveys can achieve much higher spatial resolution, in practice such surveys are only conducted for small samples of the whole population and for very limited periods of time. Another limitation is the unavailability of the actual transport mode ground-truth information at the individual level, which restricts our validation to an aggregate level (national level). Thus, the individual-level validation will be part our future investigation that explores possible means for ground-truth data collection. Our future investigation will also include the development of map matching algorithm that can be applied to map waypoints (*legs* information) to actual road network for higher precision. Furthermore, there are other significant places besides residence and workplace such as grocery store, restaurant, and so on that can be explored for more complete understanding of travel behavior, in particular, mode choice behavior, which may vary significantly with the type of these contextual places, as seen in service usage patterns [27, 28]. Moreover, there are several relevant factors that can influence transport mode choice decision such as age, socio-economic status, geographical constraints, social influence, awareness to environmental damage, time of commuting, and so on. Social influence, the focus of this study, is one of these many influential factors. Although social influence may not be the strongest determinant of travel behavior, it contributes interestingly to some extent, and its effect has often been left out when modeling the mode choice decision process. So, this study provides an observational view of travel behavior from the social network context perspective.

We believe that this study sheds new light on the transport behavior research from the social-context perspective, and we hope that our findings suggest new ways to use mobile phone data to investigate the social influence on transport mode choice behavior. This present analysis is based on a static ego-centric social network. In the future, with time-evolving information such as transport mode adoption, and network evolution, it

will allow for investigation of more insights into social influence on mode adoption over time. Additional data sources such as socioeconomic status in different areas can also be considered in future investigation. There is still a number of remaining interesting open questions, such as how much does each of the social and spatial elements influence the mode choice decision, what role do the social norms play in the decision, and how does transport mode decision change over time?

Appendix

Tables 1 and 2 show the numeric results obtained from a random reference system where mode choices are randomly reshuffled among people in the dataset. The experiment was carried out for 10 trials each with a different random set of mode choice assignments. The purpose of the experiment is to compare the obtained results with the result in Figure 10 to ensure that that the result in Figure 10 is not due entirely to the unbalanced distribution of car and transit users. Tables 3 and 4 show the corresponding histograms of the results in Tables 1 and 2, respectively.

Table 1 Results obtained from a random reference system where driving mode choices are randomly reshuffled among people in the dataset. Experiment is composed of 10 trials each with a different random set of driving mode choice assignments

Trial	Average fraction of social ties who commute by driving and transit			
	Driving ego-subjects		Transit ego-subjects	
	Driving ties	Transit ties	Driving ties	Transit ties
1	0.796	0.204	0.771	0.229
2	0.79	0.21	0.758	0.242
3	0.79	0.21	0.754	0.246
4	0.806	0.194	0.737	0.263
5	0.795	0.205	0.769	0.231
6	0.781	0.219	0.774	0.226
7	0.773	0.227	0.811	0.189
8	0.799	0.201	0.761	0.239
9	0.773	0.227	0.763	0.237
10	0.782	0.218	0.777	0.223

Table 2 Results obtained from a random reference system where transit mode choices are randomly reshuffled among people in the dataset. Experiment is composed of 10 trials each with a different random set of transit mode choice assignments

Trial	Average fraction of social ties who commute by driving and transit			
	Driving ego-subjects		Transit ego-subjects	
	Driving ties	Transit ties	Driving ties	Transit ties
1	0.798	0.202	0.76	0.24
2	0.808	0.192	0.758	0.242
3	0.793	0.207	0.753	0.247
4	0.798	0.202	0.765	0.235
5	0.804	0.196	0.77	0.23
6	0.809	0.191	0.747	0.253
7	0.798	0.202	0.765	0.235
8	0.794	0.206	0.741	0.259
9	0.796	0.204	0.772	0.228
10	0.791	0.209	0.763	0.237

Table 3 Histograms of the results shown in Table 1

Trial	Histograms

Table 3 (*Continued*)

Trial	Histograms
6	
7	
8	
9	
10	
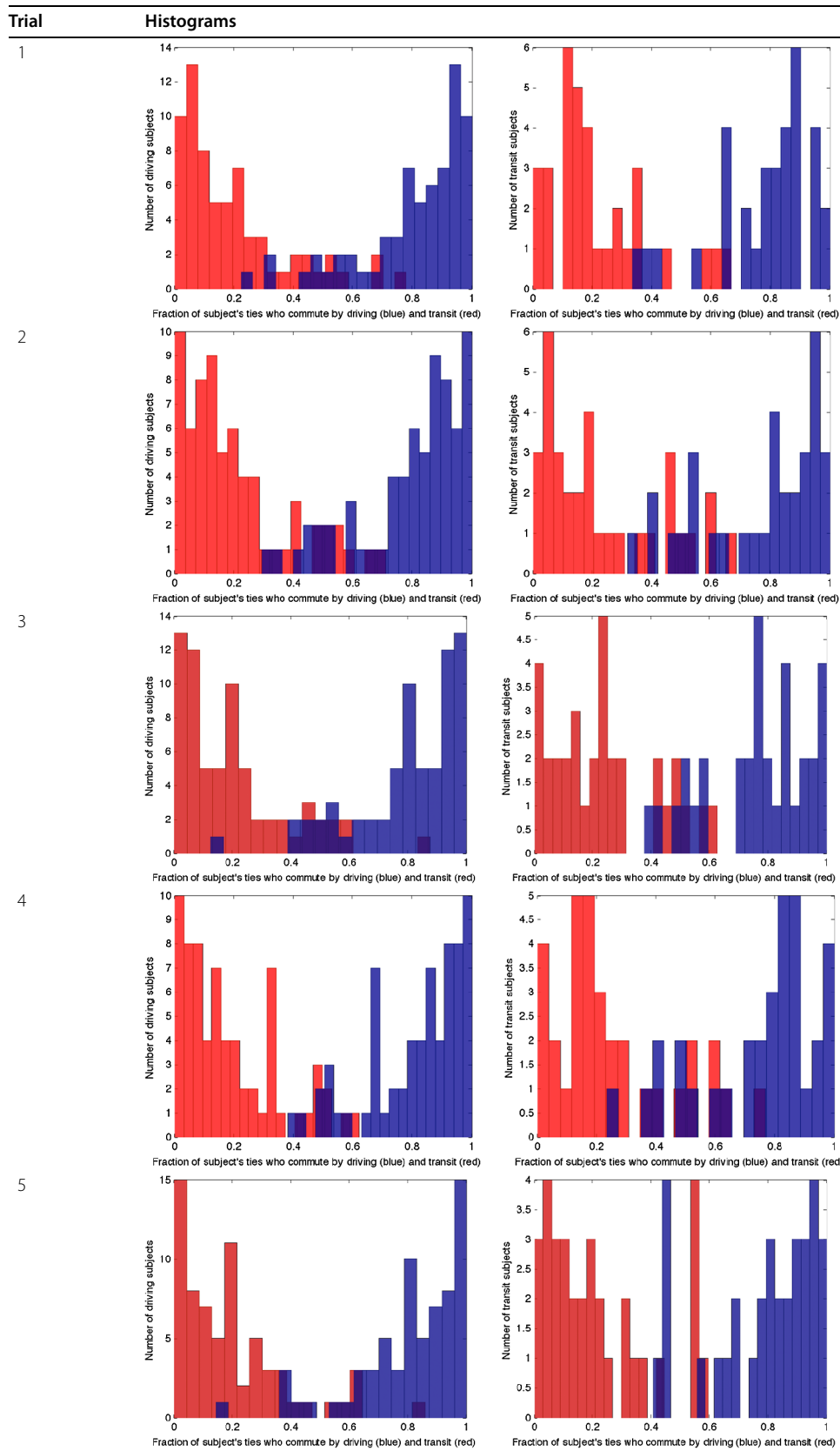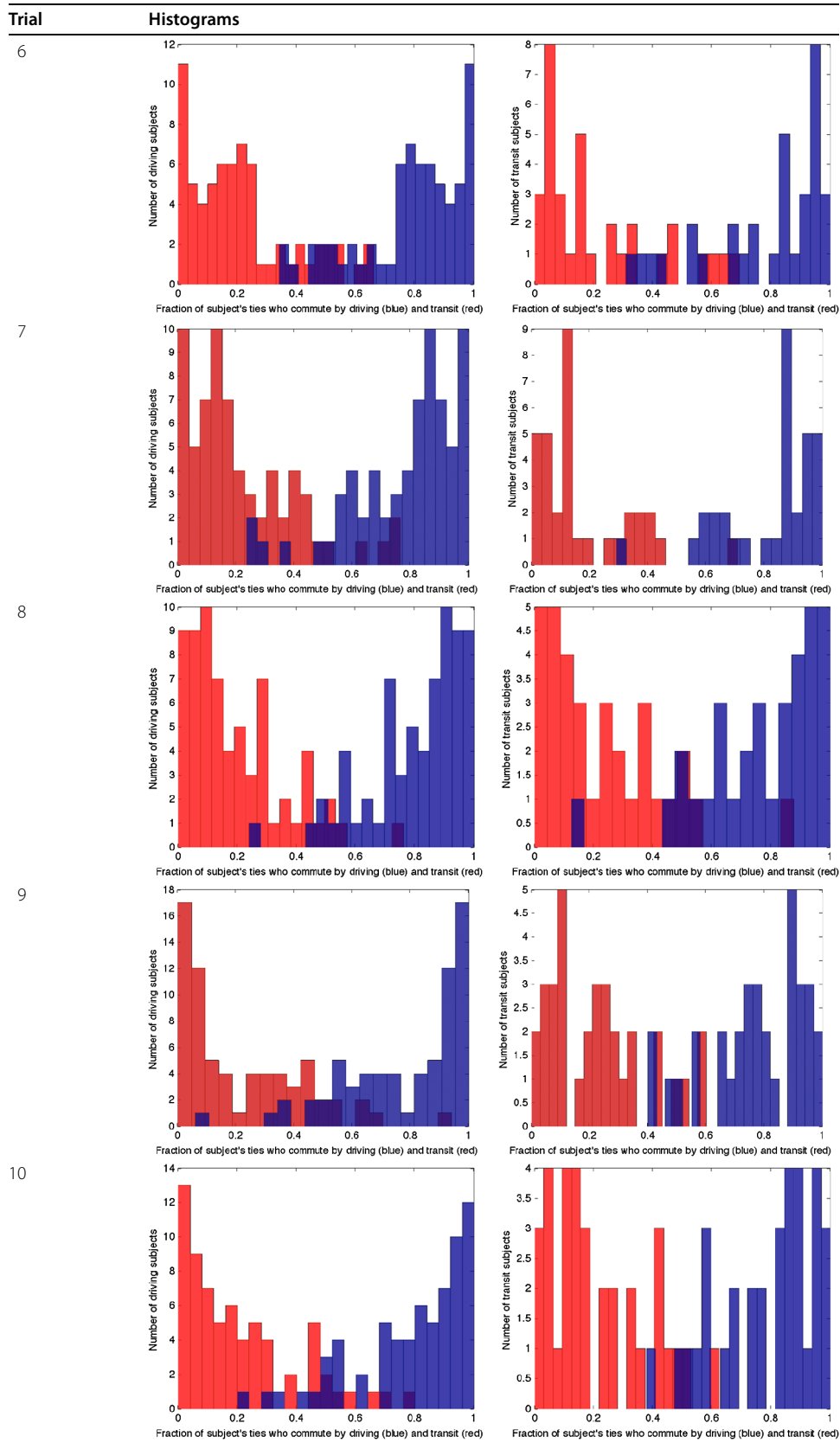
Table 4 Histograms of the results shown in Table 2

Trial	Histograms
1	
2	
3	
4	
5	

Table 4 (*Continued*)

Trial	Histograms
6	
7	
8	
9	
10	
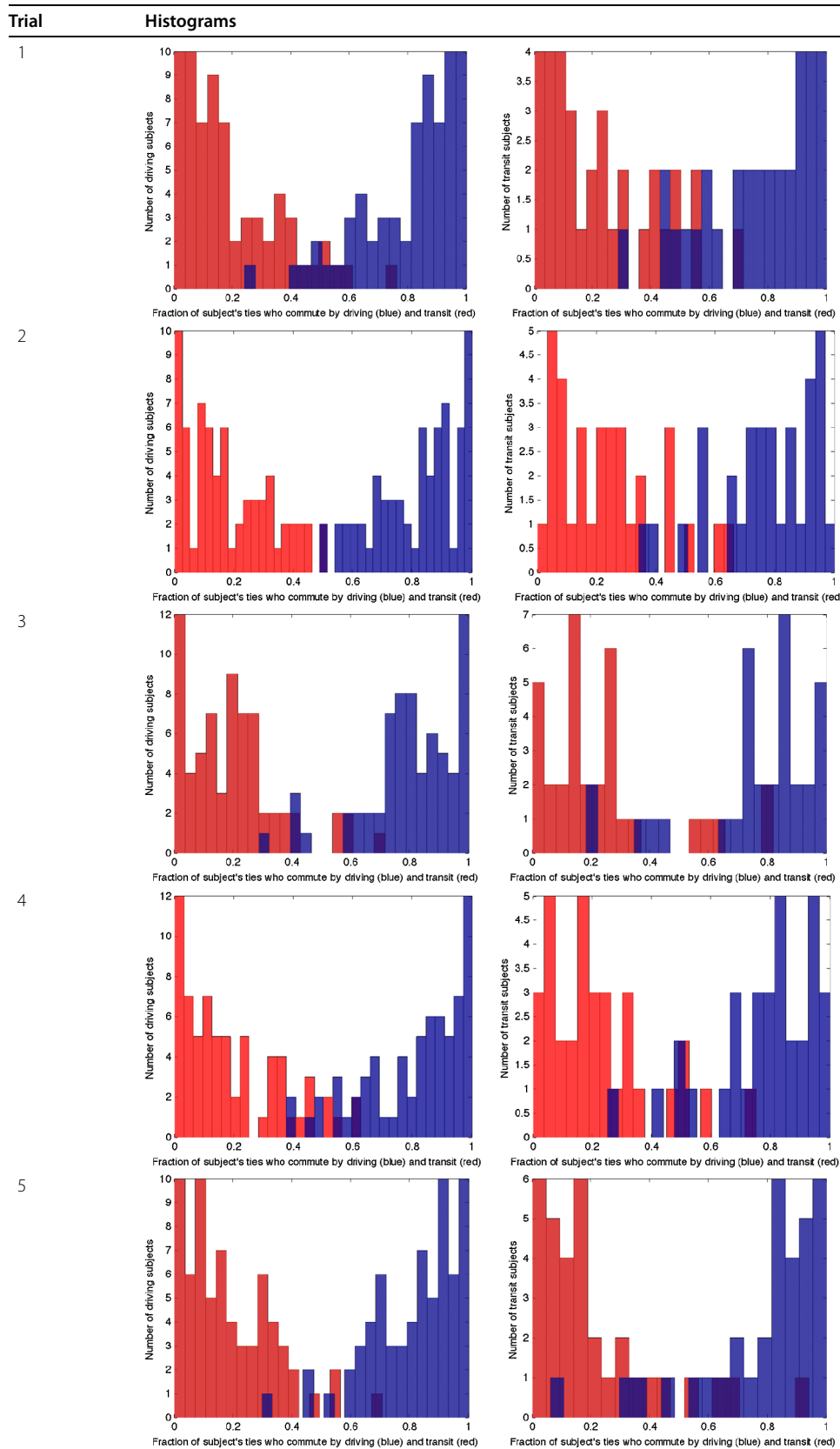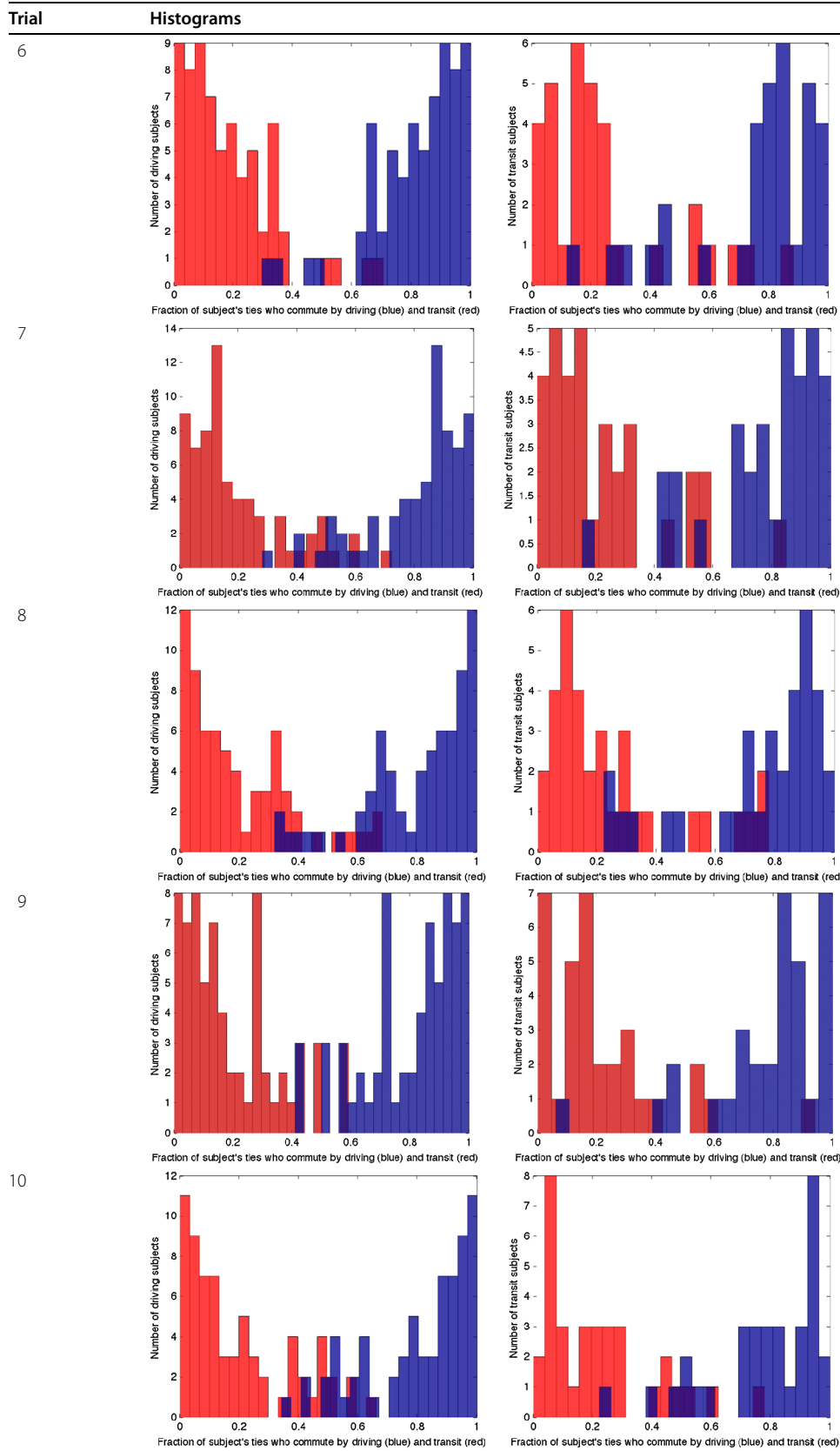
Table 5 shows the calculated values of the correlation coefficient (*r*) of candidate functions that fit the data points shown in Figures 11-15, from which the best fitted function is chosen and shown in bold type.

Table 6 shows the re-produced results of Figure 11 by a random reference system - i.e., considering a network with the same number of connections, degree, and overall fraction of weak ties as originally, but with the transport modes reshuffled among people. Experiment was carried out for 10 trials to produce some evidence that ensures that the result observed in Figure 11 is not entirely due to the overwhelming fraction of weak ties.

Table 5 Calculated values of the correlation coefficient (*r*) of the fitted candidate functions

Figure	Correlation coefficient (*r*) of fitted function			
	Linear: $ax + b$	Exponential: ae^{bx}	Power-law 1: $a(x + b)^n$	Power-law 2: $ax^n + c$
11	0.073	0.078	0.139	**0.941**
12(a)	0.263, 0.896	0.263, 0.969	0.228, 0.109	**0.943, 0.975**
12(b)	0.634, 0.968	0.551, 0.107	0.103, 0.863	**0.950, 0.987**
13(a)	0.914	0.802	**0.924**	0.637
13(b)	0.815	0.803	**0.844**	0.486
13(c)	0.825	0.803	**0.834**	0.486
13(d)	0.801	0.803	**0.888**	0.486
14	0.867	0.761	**0.886**	0.654
15(a)	**0.940**	0.937	0.119	0.923
15(b)	**0.890**	0.651	0.101	0.610

Table 6 Re-generated result of Figure 11 by considering a network with the same number of connections, degree, and overall fraction of weak ties as originally, but with the transport modes reshuffled among people. Experiment was carried out for 10 trials

Trail	Result	Trail	Result
1		3	
2		4	

Table 6 (*Continued*)

Trail	Result	Trail	Result
5		8	
6		9	
7		10	
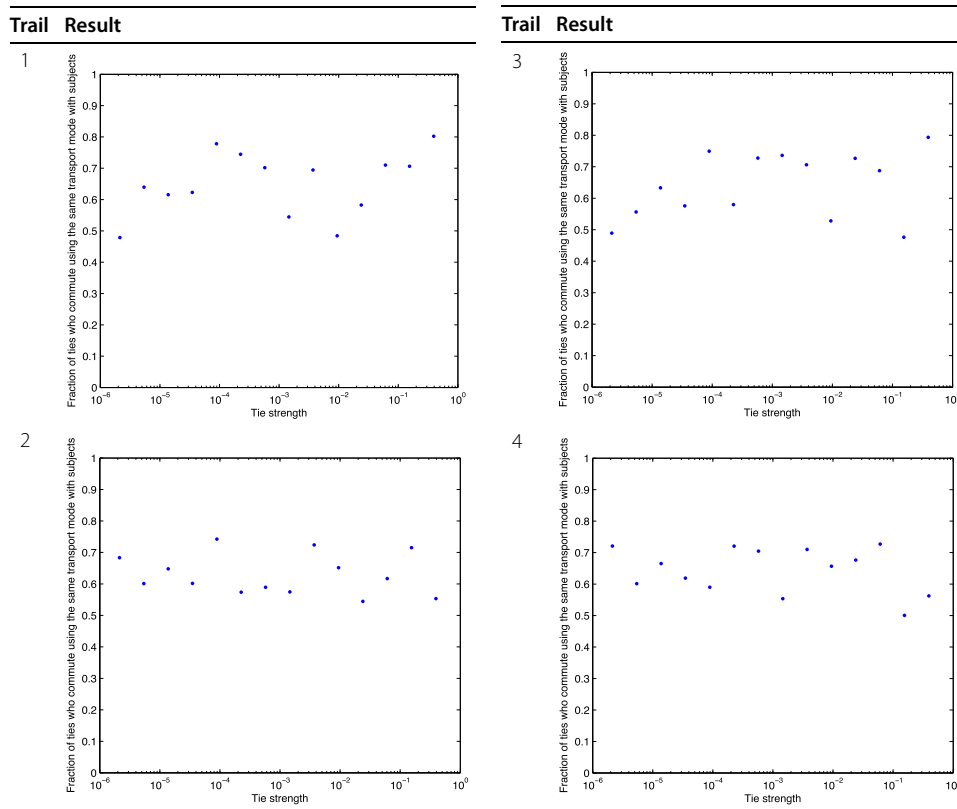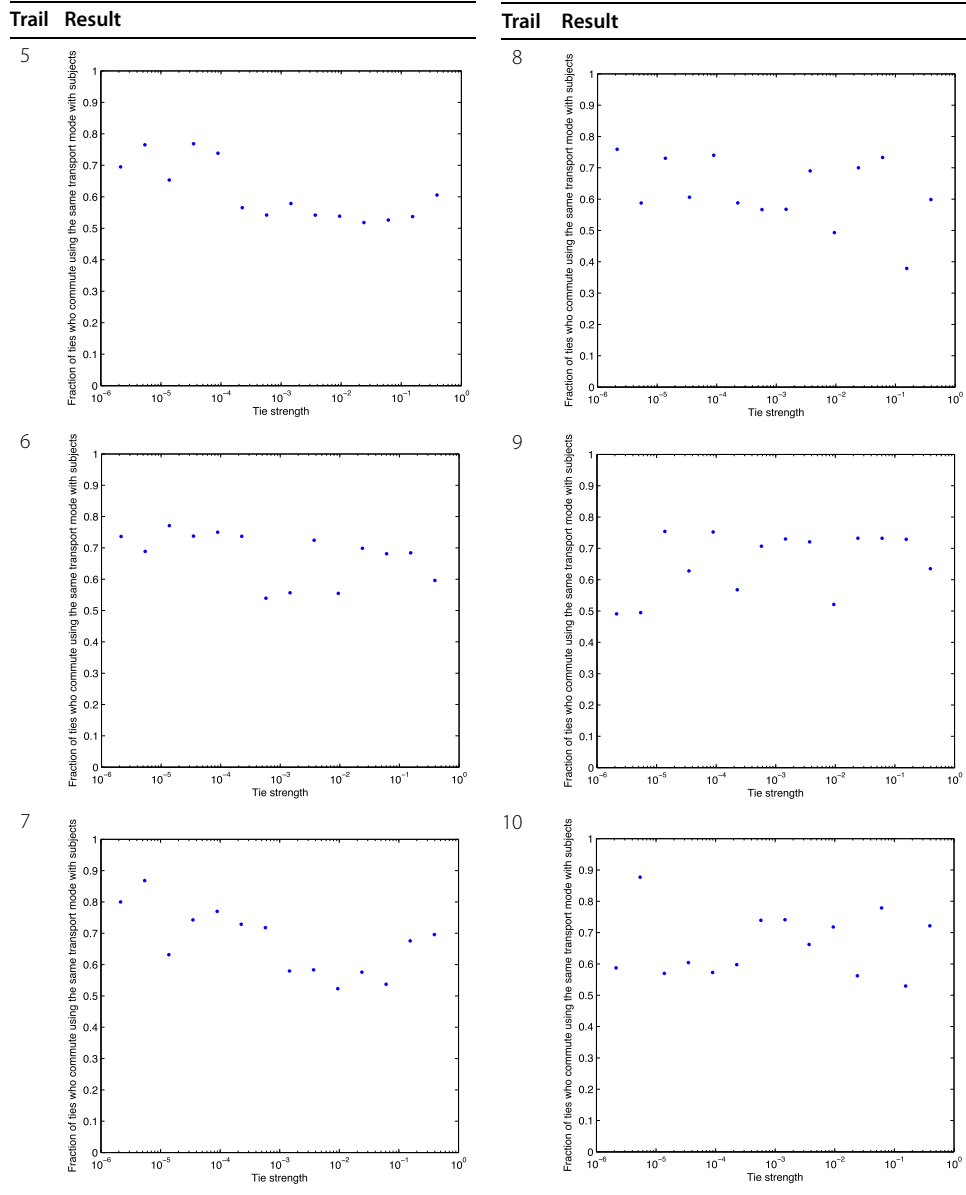
Availability of data and materials
The dataset used in this study was provided by the France Telecom. A sample of data can be made available on request to other researchers for academic, non-commercial purposes by the authors.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
SP conceived of and designed the study. SP, TS, MD, ZS, JN, and CB analyzed the data. TS preprocessed the data. SP and TS developed the algorithm for transport inference. SP, TS, MD, ZS, JN, and CB wrote the manuscript. All authors read and approved the final manuscript.

Author details
[1]Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand.
[2]Department of Civil Engineering, Schulich School of Engineering, University of Calgary, Calgary, Canada. [3]Sociology and

Economics of Networks and Services Department, Orange Labs, Issy-les-Moulineaux, France. [4]Center for Informatics and Systems, University of Coimbra, Coimbra, Portugal.

Publisher's Note

References

1. Litman T (2013) The new transportation planning paradigm. ITE J 83(6):20-28
2. Goodwin P (2011) Three views on 'peak car'. World Transp. Policy Pract. 17(4):8-17
3. Schwanen T, Banister D, Anable J (2012) Rethinking habits and their role in behaviour change: the case of low-carbon mobility. J Transp Geogr 24:522-532
4. Domencich T, McFadden D (1975) Statistical estimation of choice probability function. In: Urban travel demand: a behavioral analysis, pp 101-125
5. Ortúzar J, Willumsen L (2011) Modeling transport, 4th edn. Wiely, New York
6. Axhausen K (2003) Social networks and travel: some hypotheses. Zürich
7. Granovetter M (1978) Threshold models of collective behavior. Am J Sociol 83:1420-1443
8. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Annu Rev Sociol 27:415-444
9. Newman MEJ (2003) Mixing patterns in networks. Phys Rev E 67:026126
10. Kowald M, Arentze T, Axhausen K (2015) Individuals' spatial social network choice: model-based analysis of leisure-contact selection. Environ Plan B, Plan Des 42(5):857-869
11. Bar-Gera H (2007) Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: a case study from Israel. Transp Res, Part C, Emerg Technol 15(6):380-391
12. Demissie MG, Correia GH, Bento C (2013) Intelligent road traffic status detection system through cellular networks handover information: an exploratory study. Transp Res, Part C, Emerg Technol 32(1):76-78
13. Herrera J, Work D, Herring R, Ban X, Jacobson Q, Bayen A (2010) Evaluation of traffic data obtained via GPS-enabled mobile phones: the mobile century field experiment. Transp Res, Part C, Emerg Technol 18(4):568-583
14. Liu H, Danczyk A, Brewer R, Starr R (2008) Evaluation of cellphone traffic data in Minnesota. Transp Res Rec 2086(1):1-7
15. Caceres N, Wideberg J, Benitez F (2007) Deriving origin-destination data from a mobile phone network. IET Intell Transp Syst 1(1):15-26
16. Demissie M, Phithakkitnukoon S, Sukhvibul T, Antunes F, Bento C (2016) Inferring origin-destination flows using mobile phone data: a case study of Senegal. In: 13th international conference on electrical engineering/electronics, computer, telecommunications and information technology, Chiang Mai
17. Iqbal MS, Choudhury CF, Wang P, González MC (2014) Development of origin-destination matrices using mobile phone call data. Transp Res, Part C, Emerg Technol 40(1):63-74
18. Pan C, Lu J, Di S, Ran B (2006) Cellular-based data-extracting method for trip distribution. Transp Res Rec 1945(1):33-39
19. White J, Wells I (2002) Extracting origin destination information from mobile phone data. In: 11th international conference on road transportation and control, London
20. Alexander L, Jiang S, Murga M, Gonzalez M (2015) Origin destination trips by purpose and time of day inferred from mobile phone data. Transp Res, Part C, Emerg Technol 58(1):240-250
21. Colak S, Alexander L, Alvim B, Mehndiretta S, Gonzalez M (2015) Analyzing cell phone location data for urban travel: current methods, limitations and opportunities. In: Transport research board, transit cooperation research program, Washington
22. Demissie M, Phithakkitnukoon S, Sukhvibul T, Antunes F, Gomes R, Bento C (2016) Inferring passenger travel demand to improve urban mobility in developing countries using cell phone data: a case study of Senegal. IEEE Trans Intell Transp Syst 17(9):2466-2478
23. Demissie M, Correia G, Bento C (2015) Analysis of the pattern and intensity of urban activities through aggregate cellphone usage. Transportmetrica A: Transp Sci 11(6):502-524
24. Soto V, Frías-Martínez E (2011) Robust land use characterization of urban landscapes using cellphone data. In: Adjunct proceedings of 9th international conference on pervasive computing, San Francisco
25. Toole J, Ulm M, González M, Bauer D (2012) Inferring land use from mobile phone activity. In: ACM SIGKDD international workshop on urban computing, Beijing
26. Trestian I, Ranjan S, Kuzmanovic A, Nucci A (2009) Measuring serendipity: connecting people, locations and interests in a mobile 3G network. In: ACM IMC, Chicago
27. Jo H-H, Karsai M, Karikoski J, Kaski K (2012) Spatiotemporal correlations of handset-based service usages. EPJ Data Sci 1:10
28. Karikoski J, Soikkeli T (2013) Contextual usage patterns in smartphone communication services. Pers Ubiquitous Comput 17(3):491-502
29. Steenbruggen J, Borzacchiello MT, Nijkamp P, Scholten H (2013) Mobile phone data from GSM networks for traffic parameter and urban spatial pattern assessment: a review of applications and opportunities. GeoJournal 78(2):223-243
30. Calabrese F, Ferrari L, Blondel V (2015) Urban sensing using mobile phone network data: a survey of research. ACM Comput Surv 47(2):25
31. Blondel V, Decuyper A, Krings G (2015) A survey of results on mobile phone datasets analysis. EPJ Data Sci 4:10
32. Mutz D (1992) Impersonal influence: effects of representations of public opinion on political attitudes. Polit Behav 14:89-122
33. Helbing D, Farkas I, Vicsek T (2000) Simulating dynamical features of escape panic. Nature 407:487-490

34. Hirshleifer D, Teoh SH (2003) Herd behaviour and cascading in capital markets: a review and synthesis. Eur Financ Manag 9:25-66
35. Krumme C, Cebrian M, Pickard G, Pentland S (2012) Quantifying social influence in an online cultural market. PLoS ONE 7(5):e33785
36. Schweitzer F, Mach R (2008) The epidemics of donations: logistic growth and power-laws. PLoS ONE 3:e1458
37. Sridha S, Srinivasan R (2012) Social influence effects in online product ratings. J Mark 76(5):70-88
38. Mavrodiev P, Tessone CJ, Schweitzer F (2013) Quantifying the effects of social influence. Sci Rep 3:1360
39. Prechter R (2001) Unconscious herding behavior as the psychological basis of financial market trends and patterns. J Psychol Financ Mark 2:120-125
40. Wenzel M (2005) Misperceptions of social norms about tax compliance: from theory to intervention. J Econ Psychol 26:862-883
41. Leahey TM, Kumar R, Weinberg BM, Wing RR (2012) Teammates and social influence affect weight loss outcomes in a team-based weight loss competition. Obesity 20(7):1413-1418
42. Salomon I (1985) Telecommunications and travel - substitution or modified mobility. J Transp Econ Policy 19:219-235
43. Hagerstrand T (1970) What about people in regional science? Pap Reg Sci 24:7-21
44. Janelle DG, Goodchild MF, Klinkenberg B (1988) Space-time diaries and travel characteristics for different levels of respondent aggregation. Environ Plan A 20:891-906
45. Lenntorp B (1976) Paths in space-time environments: a time geographic study of movement possibilities of individuals. Environ Plan 9(8):961-972
46. Pred A (1981) Of paths and projects: individual behavior and its societal context. In: Behavioral problems in geography revisited. Methuen, New York, pp 231-255
47. Harvey AS, Taylor ME (2000) Activity settings and travel behavior: a social contact perspective. Transportation 27(1):53-73
48. Arentze T, Timmermans H (2008) Social networks, social interactions, and activity-travel behavior: a framework for microsimulation. Environ Plan B, Plan Des 35:1012-1027
49. Gordon P, Kumar A, Richardson HW (1989) Gender differences in metropolitan travel behavior. Reg Stud 23:488-510
50. Hanson S, Hanson P (1981) The impact of married women's employment on household travel patterns - a Swedish example. Transportation 10(2):165-183
51. Hanson S, Hanson P (1981) The travel-activity patterns of urban residents - dimensions and relationships to sociodemographic characteristics. Econ Geogr 57:332-347
52. Pas EI (1984) The effect of selected sociodemographic characteristics on daily travel-activity behavior. Environ Plan A 16:571-581
53. Lu XD, Pas EI (1999) Socio-demographics, activity participation and travel behavior. Transp Res, Part A, Policy Pract 33(1):1-18
54. Carrasco JA, Hogan B, Wellman B, Miller EJ (2008) Collecting social network data to study social activity-travel behavior: an egocentric approach. Environ Plan B, Plan Des 351(6):961-980
55. Ben-Akiva M, Lerman SR (1985) Discrete choice analysis: theory and applications to travel demand. MIT Press, Cambridge
56. Gliebe JP, Koppelman FS (2002) A model of joint activity participation between household members. Transportation 29(1):49-72
57. Scott DM, Kanaroglou PS (2002) An activity-episode generation model that captures interactions between household heads: development and empirical analysis. Transp Res, Part B, Methodol 36(10):875-896
58. Páez A, Scott DM (2007) Social influence on travel behavior: a simulation example of the decision to telecommute. Environ Plan A 39:647-665
59. Scott DM, Dam I, Páez A, Wilton RD (2012) Investigating the effects of social influence on the choice to telework. Environ Plan A 44(5):1016-1031
60. Phithakkitnukoon S, Smoreda Z, Olivier P (2012) Socio-geography of human mobility: a study using longitudinal mobile phone data. PLoS ONE 7(6):e39253
61. Song C, Qu Z, Blumm N, Al B (2010) Limits of predictability in human mobility. Science 327(5968):1018-1021
62. González MC, Hidalgo CA, Barabási A (2008) Understanding individual human mobility patterns. Nature 453:779-782
63. Song C, Koren T, Wang P, Barabási A (2010) Modelling the scaling properties of human mobility. Nat Phys 6:818
64. Calabrese F, Di Lorenzo G, Ratti C (2010) Human mobility prediction based on individual and collective geographical preferences. In: International conference on intelligent transportation systems, Madeira Island, Portugal
65. Becker R, Cáceres R, Hanson K, Isaacman S, Ji M, Martonosi M, Rowland J, Urbanek S, Varshavsky A, Volinsky C (2013) Human mobility characterization from cellular network data. Commun ACM 56(1):74-82
66. Wang P, González MC, Hidalgo CA, Barabási A (2009) Understanding the spreading patterns of mobile phone viruses. Science 324(5930):1071-1076
67. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, Buckee CO (2012) Quantifying the impact of human mobility on malaria. Science 338(6104):267-270
68. Hidalgo CA, Rodriguez-Sickert C (2008) The dynamics of a mobile phone network. Physica A 387(12):3017-3024
69. Onnela JP, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási A-L (2007) Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci USA 104(18):7332-7336
70. Phithakkitnukoon S, Dantu S (2011) Mobile social group sizes and scaling ratio. AI Soc 26(1):71-85
71. Phithakkitnukoon S, Calabrese F, Smoreda Z, Rattti C (2011) Out of sight out of mind - how our mobile social network changes during migration. In: International conference on social computing, Boston, MA
72. Eagle N, de Montjoye Y, Bettencourt L (2009) Community computing: comparisons between rural and urban societies using mobile phone data. In: International conference on computational science and engineering, Vancouver, BC
73. Eagle N, Macy M, Claxton R (2010) Network diversity and economic development. Science 328(5981):1029-1031
74. Phithakkitnukoon S, Leong T, Smoreda Z, Olivier P (2012) Weather effects on mobile social interactions: a case study of mobile phone users in Lisbon, Portugal. PLoS ONE 7(10):e45745

75. Onnela J-P, Arbesman S, González M, Barabási A-L, Christakis N (2011) Geographic constraints on social network groups. PLoS ONE 6(4):e16939
76. Krings G, Calabrese F, Ratti C, Blondel V (2009) Scaling behaviors in the communication network between cities. In: International conference on computational science and engineering, Vancouver, BC
77. Calabrese F, Smoreda Z, Blondel V, Ratti C (2011) Interplay between telecommunications and face-to-face interactions: a study using mobile phone data. PLoS ONE 6(7):e20814
78. Domenico MD, Lima A, Musolesi M (2013) Interdependence and predictability of human mobility and social interactions. Pervasive Mob Comput 9(6):798-807
79. Granovetter M (1973) The strength of weak ties. Am J Sociol 78(1):1360-1380
80. Google (2016) Google maps directions API. [Online]. Available: https://developers.google.com/maps/documentation/directions/. [Accessed 7 January 2016]
81. Chee WL, Fernandez JL (2013) Factors that influence the choice of mode of transport in penang: a preliminary analysis. Proc, Soc Behav Sci 91(10):120-127
82. Beirao G, Cabral JS (2007) Understanding attitudes towards public transport and private car: a qualitative study. Transp Policy 14:478-489
83. Anwar AHMM (2009) Paradox between public transport and private car as a modal choice in policy formulation. J Bangladesh Inst Plann 2:71-77
84. Viegas FAR (2008) Critérios para a Implementação de Redes de Mobilidade Suave em Portugal. Universidade Técnica de Lisboa Instituto Superior Técnico, Lisbon
85. Sinnott RW (1984) Virtues of the haversine. Sky Telesc 68(2):159
86. Eurostat (2011) Instituto Nacional De Estatistica (Statistics Portugal). Modal split of passenger transport. [Online]. Available: https://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=124229996&att_display=n&att_download=y. [Accessed 10 3 2016]
87. ECORYS Transport (2006) ECORYS transport. Study on strategic evaluation on transport investment priorities under structural and cohesion funds for the programming period 2007-2013. [Online]. Available: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/pdf/evasltrat_tran/portugal.pdf. [Accessed 10 3 2016]
88. Lauwerijssen P (2011) Tie strength and the influence of perception: obtaining diverse or relevant information. Tilburg University, Tilburg
89. Brown JJ, Reingen PH (1987) Social ties and word-of-mouth referral behavior. J Consum Res 14(3):350-362
90. Goldenberg J, Libai B, Muller E (2001) Talk of the network: a complex systems look at the underlying process of word-of-mouth. Mark Lett 12(3):211-223
91. Kleijn MSD (2015) The influences of an individual's social network on the choice of travelling by public transport. Padualaan
92. Papaioannou D, Martinez L (2015) The role of accessibility and connectivity in mode choice. A structural equation modeling approach. In: 18th Euro Working Group on Transportation, EWGT 2015, Delft, The Netherlands
93. Donald I, Cooper S, Conchie AS (2014) An extended theory of planned behaviour of the psychological factors affecting commuters' transport mode choice. J Environ Psychol 40:39-48
94. Watts D, Dodds P (2009) Threshold models of social influence. In: The Oxford handbook of analytical sociology. Oxford University Press, Oxford, pp 475-497