#### **RESEARCH ARTICLE**



# Detecting interpersonal relationships in large-scale railway trip data

Kimitaka Asatani<sup>1</sup> · Fujio Toriumi<sup>1</sup> · Junichiro Mori<sup>1</sup> · Masanao Ochi<sup>1</sup> · Ichiro Sakata<sup>1</sup>

Received: 30 December 2017 / Accepted: 24 July 2018 / Published online: 6 August 2018 © The Author(s) 2018

#### Abstract

With increases in the amount of human trajectory data, interest in explaining or predicting human mobility is growing. Owing to the difficulty of associating mobility data with interpersonal relationship data, previous studies on the link between interpersonal relationships and mobility are limited to the specific activities of particular users. In this paper, we propose a method for detecting interpersonal relationships from mobility data, while distinguishing these relationships from those of familiar strangers such as commuters. In the method, persons who take diverse variations within the same activities are recognized as a pair. From IC card data covering the daily mobility of six million people over three years, we detected millions of frequently co-located pairs. Under certain conditions, most of the detected pairs are confirmed as not being familiar strangers, but rather to have an interpersonal relationship. Next, we analyzed the detected pairs and found that the density of the relationships between groups was divided by gender and age and was found to be asymmetric by gender. For example, an elderly male person is not likely to take trips as a pair with a same-gender elderly person, and this result is data-based evidence for the isolation of retired men. In addition, group trips are confirmed to have an extraordinal character and sometimes converge spatiotemporally. These findings indicate that interpersonal relationship is a strong factor to determine their mobility and group observation is potentially useful for event detection.

**Keywords** Human mobility · Interpersonal relationship



K. Asatani acknowledge the support of JSPS KAKENHI-PROJECT (Grant number 16K16167).

The University of Tokyo, 7 Chome-3-1 Hongo, Bunkyo, Tokyo, Japan

### Introduction

Understanding and predicting human mobility is important for many applications such as preventing the spread of disease [1], transportation engineering [13, 28], and event planning [24]. Most previous models focused on human mobility behavior [10, 27] and interactions between humans and the characteristics of a place [9, 12, 15, 22]. Such models could be used to explain the trajectory of a person or population. Some studies suggest that the probability of crime [3] and pattern of land use [11, 14] can be estimated based on human mobility. Despite these studies, human mobility, which results from various motivations, is not fully understood and cannot be well predicted [25]. In daily life, human mobility is often motivated by interpersonal relationships; that is, the need or desire to meet with someone or to travel together.

A previous study [6] shows that social interaction is important for predicting the future location of an individual. In addition, some studies [5, 8, 26, 29] have analyzed the relationship between social ties and human mobility. However, these studies used only limited data for a given time span and/or target group. By putting aside the problem of data availability, researchers have considered human mobility using a theoretical approach [17, 18, 20]. In spite of much attention being paid to human mobility and interpersonal relationships, many questions remain regarding the correlation between interpersonal relationships and human mobility. For instance, how do people interact with each other in their daily life? What types of people interact frequently? What are the characteristics of trips taken with friends?

In a broader context, interpersonal relationships are considered a fundamental component of human societies and are strongly related to the mental and physical health of individuals [19]. Many researchers have discussed the estimation of hidden interpersonal relationships based on online communities [7, 16] and image sharing services [4]. Moreover, some researchers have investigated the connection between interpersonal relationships and activities on the internet [21]. However, these data are limited to the specific activities of each user. By contrast, human trajectories collected over a large area can better represent daily human activities. For people living in large cities, trip data from railway systems reflect their activities of daily life [24]. Connecting interpersonal relationship data with human relationships helps us better understand daily human interactions. However, there are obvious issues with combining such data because of data restrictions and privacy concerns.

To answer the above questions about interpersonal relationships and human mobility in the absence of sufficient data, we have developed a method for extracting interpersonal relationship data from human trajectories. We analyzed data collected for six million people from the Kansai-area (second-largest city area in Japan, including Osaka, Kobe, Nara, and Kyoto) over three years, resulting in 2.2 billion trajectories. These data are sufficient to trace various human activities, including routine and extraordinary movements, because movement by train is essential to the daily activities of citizens, a large proportion of whom travel by



train using an IC card. We extract interactions according to the following simple rule: users who share greater than (or equal to) a certain number of departure/destination station pairs on identical trips (i.e., the same destination and same departure location at the same time) are considered friends.

We extracted millions of frequently co-located pairs from the mobility data. To distinguish these pairs as either familiar strangers who unintentionally take the same trip often (i.e., commuters) or interpersonal related persons, we observed differences in the detected pairs in relation to changes in time window size. In certain condition, the most of detected pairs are assumed to be interpersonal related persons who have strong intention to move together because the increase of time window size has only a small impact on the detected pairs. We observed the pairs and found a significant gender difference in such relationships. Moreover, we revealed certain novel characteristics of group trips, e.g., that they exhibit an extraordinal character and converge spatiotemporally.

#### Methods

We extracted frequently co-located pairs from human trajectories (see "Detection of frequently co-located pairs" section) and identified group trips made as a group/pair (see "Extracting group trips" section). Next, we calculated the density of the relationships according to age and gender (see "Density of relationship" section). In "Labeling relationships and group trips" section, we describe the characteristics of the detected relationships and group trips.

#### Data

Human trajectories were collected from the Kansai-area of Japan (second-largest city area in Japan, including Osaka, Kobe, Nara, and Kyoto) from April 2014 to March 2016. The data comprise trips with six major railway companies and include a total of 500 million trips taken by six million unique anonymized users. Of these, five million (25% of the population of Kansai) used the train over 100 times in three years. The trajectories were recorded by the minute. Of these users, 33% had associated age labels (divided into 10-year bins) and genders. To maintain privacy, we do not describe any details of the users or small groups. The data contain no information indicating which users are friends.

### Detection of frequently co-located pairs

We developed two methods, both of which extract frequently co-located pair. The first is the "identical trip method", in which people who take at least  $t_{id}$  identical trips are detected as frequently co-located pairs. We identify an identical trip using the boarding/dropping station and time, which is recorded by the minute. In the case of time window size  $w = 1(\min)$ , trips in which the individuals of a pair depart and arrive at the same minute are recognized as identical. In the case of  $w \ge 2$ , trips



in which the difference between the two individuals' departure and arrival times is within  $w-1(\min)$  are considered identical trips. People x, y who share at least  $t_{\rm id}$  identical trips are considered frequently co-located pairs  $P_{\rm id}^{t_{\rm id},w}(x,y)=1$  (see Eq. 1). In this equation,  $IT^w(x,y)$  denotes identical trips between users x and y in window size w.

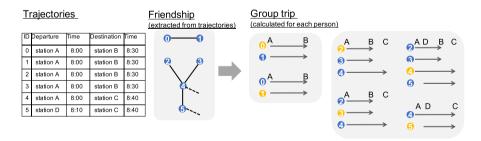
$$P_{\mathrm{id}}^{t_{\mathrm{id}},w}(x,y) = \begin{cases} 1 & n(IT_{(x,y),w}) \ge t_{\mathrm{id}} \\ 0 & \mathrm{else.} \end{cases}$$
 (1)

In the identical trip method, some familiar strangers who share the same commute are likely to be detected as a pair. If we use a sufficiently large  $t_{\rm ur}$  value, we also detect punctual commuters who take exactly the same trip every day. Assuming that people who are connected by interpersonal relationships take many kinds of trips together over a long time period, we proposed the "unique route method", in which people who share at least t destination/departure stations are detected as co-located pairs. More formally, people x, y who share at least  $t_{\rm ur}$  destination/departure station sets are considered frequently co-located pairs  $P_{\rm ur}^{t_{\rm ur},w}(x,y)=1$  (see Eq. 2). In this equation,  $ITS_{(x,y),w}$  denotes a set of destination and departure stations in which users x and y have identical trips in window size w.

$$P_{\text{ur}}^{t_{\text{ur}},w}(x,y) = \begin{cases} 1 & n(\{(s_i, s_j) | (s_i, s_j) \in ITS_{(x,y),w}\}) \ge t_{\text{ur}} \\ 0 & \text{else.} \end{cases}$$
 (2)

# **Extracting group trips**

From the trajectories and the relationship networks constructed by the detected interpersonal relationships, each individual's group trip is extracted as follows. Initially, for each trajectory, we extract trajectories if more than one person boards at the same time from the same station (left in Fig. 1). Then, we form a subgraph of the



**Fig. 1** Example of extracting group trips. From the trajectories and interpersonal relationship graph extracted by the method described in Eq. 2, each individual's group trips are extracted (yellow and blue circles in each figure indicate the target person and his/her friend, respectively). In the simplest case, when person 0 travels with person 1 from station A to B, we recognize that person 0 has traveled with person 1 and person 1 has traveled with person 0 from station A to B. For persons 2 to 5, the group trip is recognized differently for each person. For instance, person 2 departed from station A with persons 3 and 4 and then arrived at station B with person 3. Person 4 departed from station A with persons 2 and 3 and then arrived at station C with person 5



relationship network containing only those people (middle in Fig. 1). For each individual, the other individuals in a connected component are considered companions who start from the departure station of the trajectory (left in Fig. 1). Similarly, we determine which individuals alight at the same destination station.

## **Density of relationship**

Using the detected relationships, we examined which combinations of group types were most likely to exist. Users were divided into groups by age (10–20, 20–30, ..., 70–80 years) and gender. We defined the relative probability of a relationship existing between groups  $g_1$  and  $g_2$  as  $\bar{D}(g_1, g_2)$ , as shown in Eq. 3. The relative probability  $\bar{D}(g_1, g_2)$  is the ratio of the network density of relationships between  $g_1$  and  $g_2$  ( $D(g_1, g_2)$ ) to the network density of relationships between all users D(total,total). If a relationship between two groups has the same probability as relationships across the whole dataset, the relational density of these two groups takes a value of 1. As shown in Eq. 4, the probability of relationships existing between groups  $g_1$  and  $g_2$  is the ratio of the number of relationships  $\sum_{x \in g_1, y \in g_2} P(x, y)$  to the potential number of relationships  $n(g_1) * n(g_2)$ .

$$\bar{D}(g_1, g_2) = D(g_1, g_2) / D(\text{total,total})$$
(3)

$$D(g_1, g_2) = \frac{\sum_{x \in g_1, y \in g_2} P(x, y)}{n(g_1) * n(g_2)}.$$
 (4)

## Labeling relationships and group trips

The relationships between frequently co-located pairs are classified into four clusters: same home-station, same office-station, same home-office-station, and different station. Home station refers to the most frequent station at which a person alights the train for the last time in a day, and office station refers to the most frequently visited station other than the home station.

Next, we label the group trip according to the attribute(s) of the pair(s). Group trips made by at least one same-home-office-station pairs are classified to same-home-office-station group trip. Other group trips made by at least one same-home-station / same-office-station are classified to same-home-station / same-office-station group trips. Other group trips are diff-station group trips.

In addition, group trips were classified into two categories: both-sides group trips and one-side group trips. Both-sides group trips are trips in which at least two people have the same departure and destination station. One-side group trips are trips in which no two people have the same departure and destination station. In Fig. 1, person 5's trip is a one-side group trip. One-side trips are not likely detected because people do not always meet or leave from inside the station. Even if that is the case, we need to observe the differences between both-sides and one-side group trips because the motivation for each group trip is considered different.



### Results

# Interpersonal relationships from frequently co-located pairs

Initially, we confirmed the existence of frequently co-located pairs. Figure 2a, b shows the number of frequently co-located pairs  $N_{\rm id}(t_{id},w) = \sum_{x < y,x,y} P_{\rm id}^{t_{id},w}(x,y)$  (identical trip method) and  $N_{\rm ur}(t_{\rm ur},w) = \sum_{x < y,x,y} P_{\rm ur}^{t_{\rm ur},w}(x,y)$  (unique route method) in relation to changing  $t_{\rm id}$ ,  $t_{\rm ur}$  according to window-size w=1 to 5. In w=1, decreases in  $N_{\rm id}$  seem to follow a power law in  $t_{\rm id} < 100$ . In the cases of  $w \ge 2$  for the identical trip method and each w for the unique route method,  $N_{\rm id}$  decreases more than a power law, but it exhibits a fatter tail than does a normal distribution. If each person takes trips randomly, the decreases are much sharper. Considering this, these pairs are considered to have sufficient similarity to represent trips taken together.

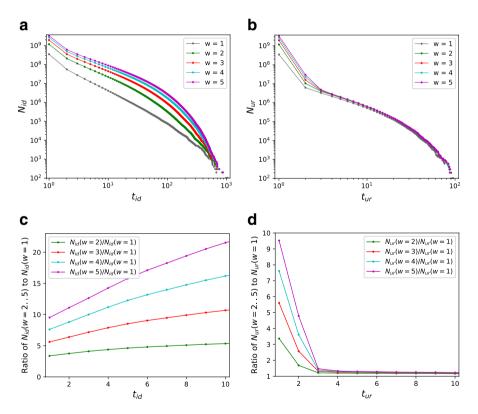


Fig. 2 Number of co-located pairs detected by the identical trip method (a) and unique route method (b) Horizontal axes indicate each threshold for the identical trip method  $t_{\rm id}$  (a) and the unique route method  $t_{\rm ur}$  (b). Vertical axes indicates the number of detected co-located pairs. In each figure, we plot the results according to changing window size  $w = 1 \dots 5$ . Ratio of the number of co-located pairs with  $w = 2 \dots 5$  to that with w = 1 by the identical trip method (c) and unique route method (d) Horizontal axes indicate each threshold of the identical trip method  $t_{\rm id}$  (c) and unique route method  $t_{\rm ur}$  (d). Vertical axes indicate the ratio of the number of detected co-located pairs under the condition of  $(w = 2 \dots 5)$  to that of (w = 1)



To confirm the proportion of detected pairs between persons who have an interpersonal relationship vs. those who are familiar strangers, we focus on the effect of window-size on the number of detected pairs. The increase in detected familiar strangers is easily inferred as w becomes larger because each person in the pair exhibits no intention of precisely adjusting their times of boarding or alighting the train. On the other hand, the increase in detected pairs in which persons have an interpersonal relationship is not large because each person in the pair passes through the ticket gate within just a few seconds. Therefore, if the detection method excludes familiar strangers, the number of detected pairs does not significantly increase as window size increases.

We can see a large increase in the number of detected co-located pairs  $N_{\rm id}$  in the identical trip method in Fig. 2a, c. In each  $t_{\rm id}$ ,  $N_{\rm id}(w=2)$  is more than three times larger than  $N_{\rm id}(w=1)$  (in Fig. 2c). In the case of  $t_{\rm id}=1$ , the number of detected pairs  $N_{\rm id}$  increases 3.37 times as w changes from 1 to 2; this is close to the estimated increase. Assuming an unrelated distribution of the departure time of each person, the probability that a pair of familiar strangers depart the same departure station at the same time in time window size w' is w'/w'' larger compared to time window size w''. This estimation is the same at arrival stations. Thus, the detection probability of identical trips increased  $(w'/w'')^2=4$  times as w changes from w'=2 to w''=1 and is close to the observed increase in detected pairs (3.37). Thus, not a few familiar strangers are included in the detected pairs detected by the identical trip method.

In the case of the unique route method, the increases in the number of detected pairs  $N_{\rm ur}$  with changing w is not significantly large at  $t_{\rm ur} \geq 3$  (in Fig. 2b, d). At  $t_{\rm ur} = 1$ , the increases are the same as for the case with the identical trip method. In the case of  $t_{\rm ur} = 2$ , a large increase in detected pairs with w is also observed. This increase is caused by many pairs of familiar strangers in which persons take an identical trip on their commute route and an identical trip on another route; this is possible because the data recording period is sufficiently large (three years). Under the condition of  $t_{\rm ur} \geq 3$ , the ratio of  $N_{\rm ur}(w=2..5)$  to  $N_{\rm ur}(w=1)$  is very small. In this case, few familiar strangers are assumed to be included in the detected pairs. In other words, the time difference of a trip between these detected pairs converges to a short time and it is assumed that each person has a strong motivation to take exactly the same time on the trip. Accordingly, these pairs are considered to intentionally take identical trips, indicating the existence of an interpersonal relationship.

The 20.0% difference between  $N_{\rm ur}(t_{ur}=3,w=2)$  and  $N_{\rm ur}(t_{ur}=3,w=1)$  comes from a few familiar stranger and interpersonal relationship pairs. Under the condition w=1, a certain proportion of group trips made by interpersonal relationship pairs are not detected as identical trips because they are recorded by minute, and a few seconds difference in recoding time exists in these trip. This problem does not occur for  $w \ge 2$ . Actually, the increases in  $N_{\rm ur}$  between w=2 and 3, 4, 5 are significantly smaller: 5.1% for (w=3), 8.5% for (w=4), and 11.0% for (w=5). Therefore, in w=2, we can avoid (i) missing the detection of interpersonal relationships and (ii) detecting familiar strangers. For these reasons, we use the unique route method with the parameter  $t_{\rm ur}=3$ , w=2 in this paper.

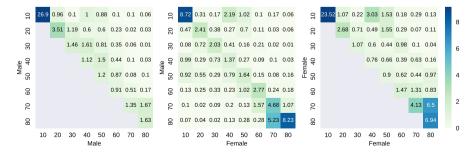


# Density of interpersonal relationships by age and gender

To confirm what type of people are likely to have interpersonal relationships, the density of the relationships between groups  $\bar{D}$  is plotted in Fig. 3. Many interpersonal relationships exist among the combinations of groups of the same age. People in their teens and twenties tend to form groups of the same gender, whereas those older than 30 form male/female groups of the same age. This tendency reflects human life: a person grows up with his/her friends and later has a partner and family. The above findings are, therefore, in line with expectations. However, in terms of gender and gerontology, we can infer some important points from the interpersonal relationship density results:

- 1. The interpersonal relationship density between women in the 40–50 years age bracket and teenagers is almost two times that of men in the 40–50 years age bracket and teenagers. This is illustrative of the fact that fathers are not as often involved in caring for their children as are mothers.
- In older adults (70 years or over), male/male interpersonal relationships are rare compared to female/male and female/female interpersonal relationships. This trend reflects the fact that elderly males have fewer coeval friends with whom to travel.
- 3. Interpersonal relationships between older men and younger women are more prevalent than those between younger men and older women. This reflects the asymmetric nature of relationships between men and women.

We obtained the same result when we changed the value of parameter  $t_{\rm ur}$  from two to ten. These results highlight significant gender differences in the activities of Japanese daily life; indeed, some studies and reports have noted that gender strongly divides the roles of Japanese people [23]. The result that elderly men do not often take trips with a same-gender friend is evidence for the general isolation of elderly men; this was also observed by experts in gender and gerontology based on questionnaire surveys [2]. Accordingly, we have thus confirmed a



**Fig. 3**  $\bar{D}$ : relative density of interpersonal relationships between groups divided by age and gender. The left-hand-side, middle, and right-hand-side figures show  $\bar{D}$  for male/male, male/female, and female/female groups, respectively

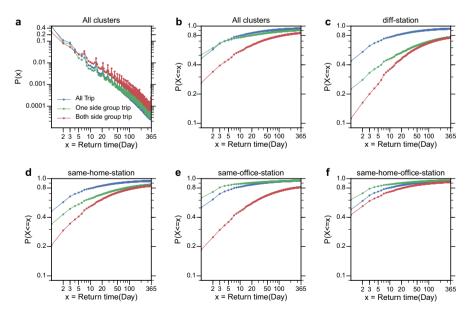


hypothesis proposed in the domain of sociology and gerontology by investigating a large amount of trip data.

# Extraordinal trips with a pair/group

Daily routines regulate the trajectories of people in large cities [10, 25, 27]. Here, we investigate the extraordinary nature of group trips to demonstrate that interpersonal relationships are a determining factor in these human trajectories. The return time (i.e., the difference between the initial arrival time and subsequent arrival times at the same station) of a trip represents the extraordinary nature of the trip. Candidates for the next trip of each group trip are whole trips (not only group trips). In addition, when a user makes several trips to a station on the same day, we consider it a single trip. The return time probability is plotted in the top-left of Fig. 4a. In this figure, the return time probability distribution follows a power law, as found in a previous study [25]. Owing to the weekly activities of individuals, we can observe a slight increase in the ratio of return times in multiples of seven days. The increase on day three is attributed to commuters' return trips between a Friday and the following Monday.

The top-middle in Fig. 4 clearly shows the difference between whole trips and group trips in the cumulative probability distribution of return times. In 43% of trips (blue line in Fig. 4b), people return to the station the next day, whereas in only 26% of both-side group trips (red line in Fig. 4b) does each person return to the station



**Fig. 4** (Cumulative) probability distribution of return time. **a** Probability distribution of the return time of all trips. **b** Cumulative probability distribution of all trips. **c**-**f** That of group trips by each cluster. In each figure, both-sides group trips (red), one-side group trips (green), and all trips (blue) are plotted



the next day. Therefore, the both-side group trips are likely to be extraordinal. This tendency is strongly observed in each cluster (except same home-office-station clusters). In other words, people in these clusters make group trips in daily ordinal activities. However, overall, this result indicates that interpersonal relationship between individuals results in travel to extraordinal places.

The difference in the distribution of the return times of one-side trips compared to those of all trips is different in each cluster. As shown in Fig. 4c, d, in different station and same home-station clusters, individuals do not likely return to the station within a short time. However, we see the opposite effect in other clusters. In the trips of these clusters, people appear to take regular daily trips with friends, e.g., depart from the station near the workplace and travel to each person's home. Thus, one-side trips have different characteristics in terms of extraordinality.

To determine why extraordinal trips are often taken with friends in both-side pair trips, we examined when and where people go with their friends. The unique route method identified 9.7% of all trips as group trips, and these tended to occur at specific times. As shown in Fig. 5b, the proportion of both-sides group trips increased during the daytime on weekdays and at weekends. These trips were not likely to be ordinal trips because the trips were not likely to be taken in commuting hours. In Fig. 5c, the time of both-sides trips is short compared with that of all trips. This implies that during the group/pair decision-making process, people reach an agreement to avoid a high cost for long travel. This tendency is observed in each group except for same-home-station group. In this group, pairs likely go outsides in the weekend and travel a long distance.

# Spatiotemporal convergence of group trajectories and event detection

As discussed above, group trips have extraordinary properties, and the destination converges owing to certain interpersonal relationship clusters. Here, we show that the group trip destination sometimes converges to particular stations at specific times. We observed the temporal sequence of the number of group trips and the number of all trips at a station near a large stadium (Fig. 6a) and at other stations (Fig. 6c, d). In Fig. 6a, we can recognize some distinct peaks, and these seem to almost exactly match

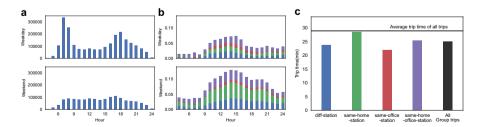
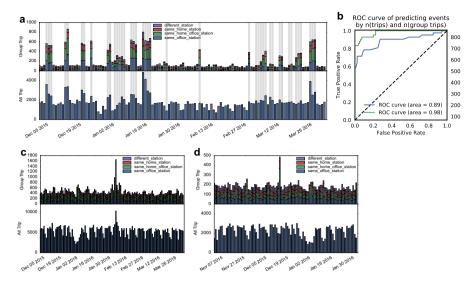


Fig. 5 a Number of trips by hour The number is calculated by weekday and weekend (including holidays). b Hourly ratio of both-sides trip In each figure, the values on weekdays and weekends are calculated separately.  $\bf c$  Average trip time for group trips and all trips. Average trip time of both-sides trips are shown for each cluster of group trips and all group trips





**Fig. 6** Daily number of group trips and total trips. Each sub-figure shows the daily number of group trips (upper) and total number of trips (lower). **a** A station near the stadium. The dates on which events were held (baseball game, concert, etc.) are highlighted in gray. In addition to this, we plot the ROC curve of the accuracy of event detection. This graph shows the predictive power for extracting an event near the station of **a** using the daily number of group trips (red) and total trips (blue) from December 2015 to March 2016. **b**, **c** Stations in a residential area

the days on which events were held at the stadium (although the total number of people per day did not necessarily peak on event day). In this case, the number of group trips has strong detective power of events. We confirmed this by ROC curve of detecting days when events are held by the daily number of group trips and total trips (Fig. 6e). In this figure, the horizontal axis indicates the true positive rate (on the *y*-axis) and the vertical axis indicates the false positive rate (on the *x*-axis) for every possible classification threshold. Figure 6b, c shows another example of the spatiotemporal convergence of group trips; these stations are located in an urban residential area. Based on the large number of group trips, we can identify that events occurred in early February (b) and in mid-December (c). Many other trajectories increase only slightly or not at all.

From these observations, there exists some temporal convergence of human interaction through mobility. Therefore, trips with friends are a reliable indicator of event detection. This means that without any information about passengers (e.g., departure station, gender, or age), the ratio of people who alight a train with friends implies the occurrence of an event near the station.

#### Discussion

In this paper, we propose a method that extracts interpersonal relationships from large-scale mobility data. In the method, people who take diverse variations within the same activity are detected as a pair. A few percents of familiar strangers are



estimated to be included in the detected pairs. This approach allows us to investigate the activities inherent in human interactions. In addition to railway trip data, an increasing amount of recorded information potentially provides information on human interactions, e.g., GPS log data, web server activity, and purchase logs. A few studies have attempted to estimate relationships among users based on social networks [7, 16] or image-sharing services [4]. The method proposed in the present study, which detects friends without accidentally including people in the same activity group, is applicable to other datasets if these data were recorded over a sufficiently long period.

In previous studies, mobility was regulated and restricted by the relationships between places and humans [9, 15, 22, 24, 27]. On a broader scale, human mobility can be understood as an aspect of human–human relationships and place–human relationships, because many trips are conducted for the sole purpose of meeting friends. In this paper, we confirmed that interpersonal relationships are another determining factor in mobility. Trips with friends are likely to be extraordinary trips; it is considered that friendship induces people to show extraordinary behaviors. In addition to this, we found a spatiotemporal convergence in group trips. These results are strong evidence of the links between interpersonal relationships and human mobility. We expect our results to contribute to agent-based approaches [18, 20] for revealing certain fundamental aspects of human mobility.

Moreover, the interaction frequencies are broken down by gender and age and highlight significant gender differences in daily life. For example, compared with older women, older men rarely interact with each other. The isolation of older men is an important issue within our society [2]. In addition to this, the probability of a relationship existing between a 30 and 50 years old women and young people is significantly different from that for 30–50 years old men and young people. The result that middle-aged men are less likely than middle-aged women to travel with young people is strong evidence for the gender division of childcare roles. This result provides strong data-based evidence applicable to policy-making and event planning.

The limitation of the proposed method is that some group trips (e.g., when friends use the train and meet outside the station) are not detected as group trips. To examine the association between interpersonal relationships and mobility, one possible solution is to apply the proposed method to more detailed human trajectory data such as mobile phone logs. We need to investigate topics that this paper does not discuss: group detection based on relationships (e.g., terrorists), transitions of interpersonal relationships over several years, and activities of specific pairs (mother/father and their children, elderly groups, etc.).

**Authors' contributions** KA conceived and designed the experiments, with FT advising on the design. MO and KA preprocessed data and KA carried out the statistical analyses. JM. and MO provided advice on the analysis. KA drafted the manuscript, with assistance from JM, FT and IS conceived the study, designed the study, coordinated the study, and helped draft the manuscript. All authors gave final approval for publication.

### Compliance with ethical standards

Conflict of interest The authors declare no competing interests.



**Data accessibility** The datasets generated and/or analyzed during the current study are not publicly available owing to user privacy issues; people can be easily identified from their train journeys over a period of three years. Therefore, we cannot make the data public. However, they are available from the corresponding author upon reasonable request.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., & Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51), 21484–21489.
- Beach, B., & Bamford, S. (2014). Isolation: The emerging crisis for older men. London: Independent Age.
- 3. Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., & Pentland, A. (2014). Once upon a crime: Towards crime prediction from demographics and mobile data. In *Proceedings of the 16th international conference on multimodal interaction* (pp. 427–434). ACM
- 4. Cheung, M., She, J., & Jie, Z. (2015). Connection discovery using big data of user-shared images in social media. *IEEE Transactions on Multimedia*, 17(9), 1417–1428.
- Cho, E., Myers, S.A., & Leskovec, J. (2011). Friendship and mobility: User movement in locationbased social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1082–1090). ACM
- 6. De Domenico, M., Lima, A., & Musolesi, M. (2013). Interdependence and predictability of human mobility and social interactions. *Pervasive and Mobile Computing*, 9(6), 798–807.
- Dong, W., Dave, V., Qiu, L., & Zhang, Y. (2011). Secure friend discovery in mobile social networks. In *INFOCOM*, 2011 Proceedings IEEE (pp. 1647–1655). IEEE
- 8. Fan, C., Liu, Y., Huang, J., Rong, Z., & Zhou, T. (2017). Correlation between social proximity and mobility similarity. *Scientific Reports*, 7, 11975.
- 9. Giannotti, F., Nanni, M., Pedreschi, D., Pinelli, F., Renso, C., Rinzivillo, S., et al. (2011). Unveiling the complexity of human mobility by querying and mining massive trajectory data. *The VLDB Journal The International Journal on Very Large Data Bases*, 20(5), 695–719.
- Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. arXiv:0806.1256
- 11. Grauwin, S., Szell, M., Sobolevsky, S., Hövel, P., Simini, F., Vanhoof, M., et al. (2017). Identifying and modeling the structural discontinuities of human interactions. *Scientific Reports*, 7, 46677.
- Hawelka, B., Sitko, I., Kazakopoulos, P., & Beinat, E. (2017). Collective prediction of individual mobility traces for users with short data history. *PloS One*, 12(1), e0170907.
- 13. Itoh, M., Yokoyama, D., Toyoda, M., Tomita, Y., Kawamura, S., & Kitsuregawa, M. (2014). Visual fusion of mega-city big data: An application to traffic and tweets data analysis of metro passengers. In *Big data* (*big data*), 2014 Ieee international conference on IEEE (pp. 431–440)
- Lee, M., & Holme, P. (2015). Relating land use and human intra-city mobility. PloS One, 10(10), e0140152.
- Matamalas, J. T., De Domenico, M., & Arenas, A. (2016). Assessing reliable human mobility patterns from higher order memory in mobile communications. *Journal of The Royal Society Interface*, 13(121), 20160203.
- Merritt, S., Jacobs, A.Z., Mason, W., & Clauset, A. (2013). Detecting friendship within dynamic online interaction networks. arXiv:1303.6372
- 17. Miller, H. J. (2005). Necessary spacetime conditions for human interaction. *Environment and Planning B: Planning and Design*, 32(3), 381–401.



- Neutens, T., Witlox, F., Van De Weghe, N., & De Maeyer, P. (2007). Space-time opportunities for multiple agents: a constraint-based approach. *International Journal of Geographical Information* Science, 21(10), 1061–1076.
- Newman, P., & Matan, A. (2012). Human mobility and human health. Current Opinion in Environmental Sustainability, 4(4), 420–426.
- O'Sullivan, D. (2008). Geographical information science: agent-based models. *Progress in Human Geography*, 32(4), 541–550.
- 21. Pflieger, G., Rozenblat, C., Mok, D., Wellman, B., & Carrasco, J. (2010). Does distance matter in the age of the internet? *Urban Studies*, 47(13), 2747–2783.
- 22. Roth, C., Kang, S. M., Batty, M., & Barthélemy, M. (2011). Structure of urban movements: Polycentric activity and entangled hierarchical flows. *PloS One*, 6(1), e15923.
- Leopold, T. A., Ratcheva, V., & Zahidi, S. (2016). The global gender gap report 2016. In World Economic Forum. http://reports.weforum.org/global-gender-gap-report-2016/. Accessed 1 Aug 2018.
- Shimosaka, M., Maeda, K., Tsukiji, T., & Tsubouchi, K. (2015). Forecasting urban dynamics with mobility logs by bilinear poisson regression. In *Proceedings of the 2015 ACM international joint* conference on pervasive and ubiquitous computing (pp. 535–546). ACM
- Song, C., Qu, Z., Blumm, N., & Barabási, A. L. (2010). Limits of predictability in human mobility. Science, 327(5968), 1018–1021.
- Toole, J. L., Herrera-Yaqüe, C., Schneider, C. M., & González, M. C. (2015). Coupling human mobility and social ties. *Journal of The Royal Society Interface*, 12(105), 20141128.
- Wang, Y., Yuan, N. J., Lian, D., Xu, L., Xie, X., Chen, E., & Rui, Y. (2015). Regularity and conformity: Location prediction using heterogeneous mobility data. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1275–1284). ACM
- 28. Yokoyama, D., Itoh, M., Toyoda, M., Tomita, Y., Kawamura, S., & Kitsuregawa, M. (2014). A framework for large-scale train trip record analysis and its application to passengers flow prediction after train accidents. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 533–544). Springer
- Zhang, C., Shou, L., Chen, K., Chen, G., & Bei, Y. (2012). Evaluating geo-social influence in location-based social networks. In *Proceedings of the 21st ACM international conference on Information and knowledge management* (pp. 1442–1451). ACM

