AGILE: GIScience Series, 3, 19, 2022. https://doi.org/10.5194/agile-giss-3-19-2022
Proceedings of the 25th AGILE Conference on Geographic Information Science, 2022.
Editors: E. Parseliunas, A. Mansourian, P. Partsinevelos, and J. Suziedelyte-Visockiene.
This contribution underwent peer review based on a full paper submission.

Unlocking social network analysis methods for studying human mobility

Nina Wiedemann[1], Henry Martin[1, 2], and Martin Raubal[1]

[1]Institute of Cartography and Geoinformation, ETH Zurich, Zurich, Switzerland
[2]Institute of Advanced Research in Artificial Intelligence (IARAI), Vienna, Austria

Correspondence: Nina Wiedemann (nwiedemann@ethz.ch)

Abstract. Planning and operations in urban spaces are strongly affected by human mobility behavior. A better understanding of individual mobility is key to improve transportation systems and to guide the allocation of public space. Previous studies have discovered statistical laws of travel distances, but the topology of movement between places has received little attention. We propose to employ network modelling methods to analyze the effect of spatial and context attributes on individual movement patterns. The perspective of mobility as a network allows to explicitly regard dyadic dependencies of sequential location visits. Here, we consider two methods developed for social networks and provide a formulation of mobility networks to justify their applicability. First, we use the Multiple Regression Quadratic Assignment Procedure to test hypotheses on the influence of location attributes on mobility behavior. Secondly, Stochastic Actor-Oriented Models are applied to model the evolution of mobility networks over time. As a proof-of-concept study, we transform data from one GNSS-based and one check-in based dataset into mobility networks and present results from both methods. We find relations that appear for a majority of samples and thus seem inherent to mobility networks. The differences between individuals and the available datasets are further quantified and discussed. We conclude that the transfer of network modeling methods is an interesting opportunity to study network-related phenomena in geographic information science.

Keywords. Human mobility, network modelling, movement analysis, network dynamics

1 Introduction

1.1 Human mobility research

With the wide-spread availability of affordable tracking technology and large-scale movement datasets, quantitative human mobility analysis emerged as a field (Laube, 2014; Dodge et al., 2016). Human mobility plays an important role in the modelling of disease spreading (Kraemer et al., 2020), in supporting sustainable mobility (Bucher, 2020), studying animal movement patterns (Demšar et al., 2021) and many other subjects (Dodge et al., 2020). Aside from specific applications, there is a rich body of literature that analyzes large-scale movement datasets to gain general insights on human movement. Findings include the identification of power-law properties in our travel patterns (Brockmann et al., 2006), individual characteristic travel distances (González et al., 2008), groups with distinct mobility behavior (Pappalardo et al., 2015) or an individual capacity of regularly visited places (Alessandretti et al., 2018). Others study *changes* of mobility behavior via clustering (Hong et al., 2021; Wang et al., 2018), feature-based anomaly detection (Jonietz and Bucher, 2018) or Bayesian approaches (Zhao et al., 2018).

1.2 Networks for individual human mobility

At the same time, complex network analysis emerged as a field as many phenomena across domains could be explained using graph representations and methods from network science (Strogatz, 2001). There are countless examples of spatial networks (Barthélemy, 2011) and the analysis of these networks is an important application area of geographic information science (Curtin, 2007). However, there is little research available on the graph-based representation and analysis of human mobility. Individual mobility can be represented by a network of visited locations which can be used to analyze human mobility patterns (Schneider et al., 2013), to predict the next visited place (Rinzivillo et al., 2014), to label activities (Martin et al., 2018) or to identify groups of similar movement behavior (Ben-Gal et al., 2019).

Although the network of visited locations is a *lossy* compression of tracking data, the topology of these graphs allows to analyze important aspects of human mobility behavior, such as the distribution of visited location in space and over time. Consider the following hypothesis as an example: In our everyday life, we have the tendency to visit locations consecutively that are close in space, such as the supermarket close to our home or work location. To test this hypothesis, the number of transitions between locations must be related to the geographic distance between them. However, common statistical tests strongly rely on the assumption of sample *independence*. Crucially, an individual's visits to locations are not independent. Such dependencies are explicitly modeled with ties in networks, and we can borrow from the rich field of social network analysis to study their characteristics.

1.3 Social network analysis

Indeed, graph theory formulations have been adapted for use in sociology since decades (Scott, 1988; Kish, 1965; Harary et al., 1965). The importance of social network analysis increased further with the rise of online platforms such as Facebook (Garton et al., 1997; Ellison et al., 2007). Much research in the field has been devoted to quantifying network properties, such as measures for density or cohesion (Smith, 1975) or centrality indices (Landherr et al., 2010), and the latter have actually been used for the analysis of movement (Blanford et al., 2015) and traffic flow (Holme, 2003; Altshuler et al., 2011). Aside from these measures, statistical models were proposed to relate the structure or evolution of networks to explanatory variables (Stadfeld and Amati, 2021). For example, the Multiple Regression Quadratic Assignment Procedure (MR-QAP) (Krackardt, 1987; Freedman and Lane, 1983) allows to test the association between the network and external variables, conditioning on the network structure. Stochastic Actor-Oriented Models (SAOMs) (Snijders, 2001) were developed to explain the evolvement of the network structure over time. Together, network modelling methods from the social sciences provide statistically sound ways to simultaneously test and control for the effect of multiple factors on network structure and dynamics.

1.4 Contribution

In this work, we thus propose to employ network modelling methods to leverage our understanding of the properties and influence factors of individual human mobility. We demonstrate in experiments on two tracking datasets how network modelling may yield insights into the spatial properties and temporal evolution of mobility networks. The results are compared across users and datasets in order to determine which characteristics are inherent to human mobility. In summary, this paper contributes the following:

- We propose a formulation of human mobility networks that enable the application of methods from the social network analysis field.

- Two methods are presented in detail, namely MR-QAP to test for significant relations between network ties and location properties, and SAOMs for the analysis of network evolution. The applicability of these methods on mobility networks is discussed and verified.

- A proof-of-concept study on two diverse datasets is presented. The proposed methods indeed yield interesting findings with respect to the topology of individual mobility, identifying its driving factors and changing characteristics over time.

2 Data and preprocessing

2.1 Datasets

Recent studies on human mobility oftentimes use large datasets of call detail records (CDR) from mobile phones (see for example Schläpfer et al. (2021)). CDR datasets are usually not publicly available and are unlabelled. In contrast, there are small tracking studies that offer high tracking coverage and rich meta information, at the cost of dataset size (Chen et al., 2016). Here, we consider two datasets that were chosen because of their high tracking coverage over a period of more than one year, offering a good balance between dataset size and quality.

The first dataset is an excerpt of the public *Foursquare* dataset. Foursquare is a social network where users can share their location with friends. To do so, they "check in" at each location they visit, and assign a purpose to the visit, e.g. "work" or "bar". The Foursquare global-scale dataset[1] presented in (Yang et al., 2015, 2016) is a large collection of Foursquare data, including more than 33 Million check-ins of 266,909 users over a period of around 18 months (April 2012 to September 2013). Each check-in record comprises the timestamp, geographic coordinates and purpose information. It is not guaranteed that users track every trip they make. However, it can be assumed that the most active users (among all 266,909 users in the dataset) check-in most of their visited locations. Thus, we restrict the dataset to the 100 most active users; specifically the users with most check-ins at the "home" location. This filter both ensures a high tracking coverage and avoids a bias due to missing home locations. After filtering for sufficient coverage (cf. 2.2.2), 42 users are included.

The second dataset is GPS tracking data from a study conducted by the Swiss Federal Railways (Martin et al., 2019). 139 participants were included in the study and were given access to a general public transport pass for all of Switzerland, as well as an electric vehicle for their personal use. The subjects participated in surveys about their socio-demographic information, and were asked to track

[1] https://sites.google.com/site/yangdingqi/home/foursquare-dataset

their movements with an app. The MyWay[2] app automatically collects GPS data and divides the track into *staypoints*, i.e. a cluster of trackpoints where the user stays at one place, and *triplegs*, i.e. sections of trips done with the same mode of transport. *Staypoints* are manually assigned a "purpose" by the users, e.g. "work". After preprocessing, we found 114 users with sufficient tracking coverage to be included in our study.

2.2 Data preprocessing

We define the individual mobility network based on the locations visited by a user as nodes and the transitions between locations as weighted edges. The locations are directly provided in the Foursquare dataset as the check-in locations. However, for Foursquare there is no information on the transitions between locations. We thus assume that every subsequent appearance of locations in the datasets corresponds to a transition between them, except if two subsequent check-ins are more than 12 hours apart. For the Green Class study we have high resolution tracking data available. We use the Python package Trackintel[3] (Martin et al., 2022) to aggregate the provided *staypoints* and *triplegs* into *locations* and *trips*, according to the data model proposed by Axhausen (2007). Specifically, *staypoints* are clustered into *locations* with the DBSCAN algorithm (parameters: $minPts = 1$, $\epsilon = 30$, i.e. a search radius of 30m). A sequence of *triplegs* is transformed into a *trip* if it is not interrupted by an activity-labelled *staypoint* or by a temporal gap of more than 25min.

2.2.1 Graph representation

The tracking data is transformed into networks. Let $\mathcal{D}(t)_{\text{sequential}} = [l_1, l_2, \dots, l_n]$ be the ordered sequence of locations visited during the time period t. A location l_i is defined by the geographic coordinates of its center, and comes along with a purpose. The sequence is first converted into pairs of subsequent locations,

$$\mathcal{D}(t)_{\text{pairs}} = [(l_i, l_{i+1}) \mid i \in [1, n] \wedge \text{ no gap between } l_i, l_{i+1}]$$

Next, a weighted directed graph is created. Let G_u^t denote the graph of user u at time period t. We first define $O(\mathcal{D}, e)$ as the count of element e in the list $\mathcal{D}$ (the number of occurrences). Then $G_u^t(V_u^t, E_u^t)$ is defined as

$$V_u^t = \{l_i \mid \forall l_i \in \mathcal{D}(t)_{\text{sequential}}\} \quad (1)$$

$$E_u^t = \{e \mid \forall e \in \mathcal{D}(t)_{\text{pairs}}\} \quad (2)$$

$$w(e) = O\big(\mathcal{D}(t)_{\text{pairs}}, e\big) \; \forall e \in E_u^t \quad (3)$$

Intuitively, the visited locations become nodes in the graph, and the transitions between locations are directed edges in the graph, weighted by the transition count. An exemplary location graph is shown in Figure 1.

[2] https://play.google.com/store/apps/details?id=ch.sbb.myway
[3] https://github.com/mie-lab/trackintel

2.2.2 Splitting into time periods and limiting networks to core activity

The experiments in section 4 analyze how the network structure evolves over time. This analysis requires access to several instances of a person's mobility network over time. Thus, we divide the tracking period into non-overlapping time periods of 120 days. A period of nearly four months guarantees a robust coverage of the reoccurring activities of a user. It is also the maximal time period that still results in three distinct time periods for both datasets. We further limit our analysis to the stable part of an individual's mobility, including only locations that are visited on a regular basis. A related concept was termed *activity set* by Alessandretti et al. (2018). Here, we define the stable part of an individual mobility network by the distinct time periods: A location is part of the core C_u if and only if it appears in all three graphs, formally $l \in C_u \iff l \in V_u^t \; \forall t$. We restrict all graphs G_u^t to its subgraph defined by the core, $G_u^t[C_u]$. All users with less than 10 nodes in the core set ($|C_u| < 10$) are excluded.

2.2.3 Node and edge properties

Last, we prepare the following edge and node attributes: 1) the distance between two locations, 2) the distance of a location from the node with purpose "home" (*distance from home*), and 3) the *purpose* of a location. Note that the first is an edge attribute whereas the latter two are node attributes. The distances are computed with the Haversine distance between the geographic coordinates. The purpose was simplified to distinguish only four categories, namely "home", "work", "shopping" and "leisure".

In the following, we will analyze the network topology with respect to these properties. Thereby it is important to distinguish between *alter* and *ego* effects for node attributes. The *ego* effect is the effect of the attribute of location l_i on the creation of an (l_i, l_j) tie, while the *alter* effect refers to the attribute of location l_j. In other words, we can relate an edge either on the properties of its origin node (the *ego*) from which the edge is *outgoing*, or the properties of its destination node (the *alter* with respect to its *incoming* edge).

3 Analyze topology-context relations in mobility networks

Given the graph and the location attributes for an individual, we demonstrate how to test hypotheses on the relation between network edges and location properties. Based on the available information and our prior knowledge about human mobility, we consider the following hypotheses:

- H1: Ties occur more often between locations that are geographically close to each other.

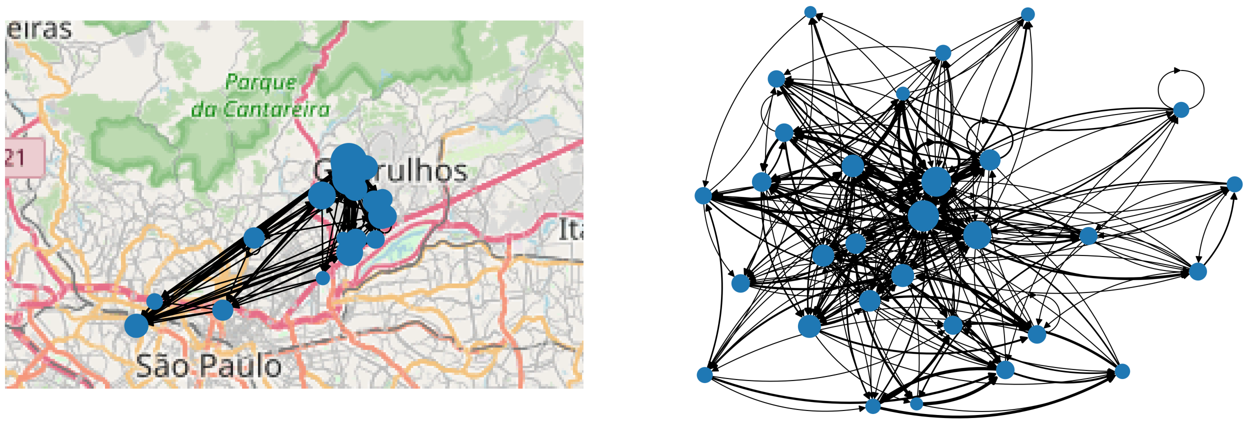

Figure 1. An exemplary location graph in two views: The left image shows the actual geographic layout, while the right plot shows the graph in a spring layout for visibility. In the location graph, nodes represent geographic locations and edges are direct trips of the user between two locations. In both visualizations, the size of nodes (blue) is proportional to the number of visits to a location, and the edge thickness shows the number of transitions between two locations, i.e. the edge weight.

- H2: Locations that are close to the "home" location are more popular.
- H3: There are more transitions between locations with the same purpose.

These hypotheses have been studied in similar forms. For example, Papandrea et al. (2016) find that the movement from place to place depends rather on the temporal than on the spatial distance between them (related to H1). They also explain visitation patterns with point-of-interest information, similar to H3. All research on the scales of human mobility (Brockmann et al., 2006; González et al., 2008) are very related to H2; however, their statistical laws regard the distribution of travel distances and do not necessarily transfer to the core set of a user's locations as considered here. Noulas et al. (2012) also use the Foursquare dataset and compare properties of individual mobility across cities, similar to our comparison of users and datasets.

The advantage of network modeling methods is that it provides a statistically sound way to test for multiple influence factors at once (such as the spatial distribution of locations or their spatial relation to the user's home), while *controlling* for the sequential ordering of location visits (e.g., the network structure). Specifically, we propose to use the Multiple Regression Quadratic Assignment Procedure (MR-QAP) (Krackhardt, 1988). MR-QAP is applied to the weighted graphs of all users separately to test H1-H3, and the results are compared across users.

3.1 Multiple-Regression Quadratic Assignment Procedure

3.1.1 Background

The Quadratic Assignment Procedure was developed by Mantel (1967); Hubert (1986) and Krackardt (1987) and has revealed insights in a variety of network analysis problems, including patent citations (Park et al., 2013), teacher collaborations (Noben et al., 2022), inter-physician networks (Mascia et al., 2015) and refugee flow (Johnson and Schon, 2019). Here, a QAP approach is preferred over other models, e.g. exponential random graph models (Lusher et al., 2013), for two reasons: a) Monte-Carlo-based estimation methods are more complex and might not converge, while QAP offers a simple approach but still rich interpretations, and b) QAP can be applied to weighted graphs without changes.

The Quadratic Assignment Procedure makes use of permutation tests in order to deal with the challenge of dyadic dependencies. Permutation tests is a type of resampling that is commonly used to find the empirical distribution of test statistics under the null hypothesis (Good, 2013). In the case of networks, permutation tests offer a way to estimate the distribution of network statistics conditioned on the observed network structure. In QAP regression, only the rows and columns of the matrix of the independent variable Z are permuted, leaving the network structure intact. For each permutation, the regression model is fitted and the parameters are collected. The resulting parameter distribution is compared to the parameters of the observed network.

3.1.2 Method

To extend QAP to multiple independent variables $Z_1,\ldots,Z_n$, MR-QAP was proposed (Freedman and Lane, 1983). Since permuting the rows and columns of a single Z_k is problematic with multiple variables, the empirical distribution is instead computed with *residual permutation*. As in normal linear or logistic regression models, the aim is to estimate the coefficients of the model

$$g(X) = \theta_0 + \theta_1 Z_1 + \cdots + \theta_n Z_n + \epsilon\ , \tag{4}$$

where the tie variable in X is the dependent variable, θ_0 is the intercept, ϵ is the vector of residuals, and θ_k are the coefficients that describe the association between Z_k and the ties in the network X.

There are two established methods for residual permutation, the Freedman and Lane (1983) method and Double Semi-Partialing (Dekker et al., 2007). Here we apply the latter since it was particularly recommended for testing linear models of continuous or skewed network count data (Dekker et al., 2007). In short, Double Semi-Partialing proceeds as follows: One variable Z_k is regressed on Z_{-k}, i.e. all independent variables except for Z_k. The residuals are computed and permuted, and Z_k^* is again computed using the permuted residuals. Finally, Z_k^* is plugged in the original model (Equation 4) and the model is fitted again, yielding the parameters of one permutation $(\theta_1^1,\ldots,\theta_n^1)$. The process is repeated and after m permutation runs, the estimates $\theta^1,\ldots,\theta^m$ determine the empirical distribution of θ in the given network structure, i.e. the expected estimates θ under the null hypothesis that no external variable has any effect.

Our implementation uses the `netlm` function of the `sna` (Butts, 2008) package in **R**. The location graph X is the dependent variable. To test H1 - H3, we formalize each hypothesis in an independent variable (Z_1, Z_2 and Z_3) and test its effect on the network structure X. Each of the Z_k is a description of pair-wise relationship of the available nodes and it is therefore a matrix of dimension $|V_u| \times |V_u|$. We define Z_1 as the matrix of pairwise distances of locations. H2 describes the effect of the location's distance from home on its "popularity", which is its weighted indegree. From the perspective of an edge, the hypothesis can be answered by relating its weight to the properties of the node attached to the incoming edge (the tie's *alter*). Therefore, Z_2 is the distance-from-home property of the tie's *alter*. Finally, Z_3 is defined as the binary matrix indicating whether or not the locations have the same purpose. The `netlm` function applies MR-QAP with $m = 5000$ residual permutations to get a good estimate of the empirical distribution of $\theta_0 - \theta_3$.

3.1.3 Exemplifying MR-QAP for one user

The model is fitted for all users separately. An example output is shown in Table 1, displaying the results for the same user of the Foursquare dataset that is shown in Figure 1. The first column in Table 1 lists the intercept θ_0 and the estimated coefficients θ_1 to θ_3. The intercept is 0.36, indicating that the number of transitions between two locations is positive with an edge weight of 0.36 if the other influence factors (distances, distance from home and purpose equality) were zero. With respect to the three hypothesis, only H1 is supported for this user ($p = 0.01$). Note that the absolute values of the estimates θ are not comparable between distance- and purpose attributes since they are provided in different units.

	θ	$P(\leq b)$	$P(\geq b)$	**p-value**
Intercept	0.36	1.00	0.00	0.00
Distance	-0.02	0.00	1.00	0.01
Distance from home (alter)	-0.01	0.12	0.88	0.26
Same purpose	-0.09	0.15	0.85	0.30

Table 1. Output of the Multiple Regression Quadratic Assignment Procedure for a single user. The regression model supports the hypothesis that there are more ties between locations that are close to each other.

3.2 Comparing MR-QAP results across users and datasets

Based on the results of all individual users, we seek to understand which relations are inherent to mobility networks (consistent across all users) and which ones are user-specific. The regression coefficients were thus estimated for all of the 156 users (114 Green Class and 42 Foursquare) and for all of their three time periods G_u^1, G_u^2, G_u^3 separately (see subsection 2.2). Five users were excluded because their core locations (the locations appearing in all three time periods) all had the same purpose ("leisure"), leading to a ill conditioned statistical model. We take the average of the coefficients over time slots, yielding one coefficient per user per variable.

Figure 2 shows the distribution of the coefficients over all users. Clearly, the direction of the effects (positive or negative coefficient) is the same for most users, indicating that they are characteristic for mobility networks. A preference for ties between locations of low distance is observed in 21% of the users (H1), whereas in only 6% cases there is evidence for the relation of the distance from home to the location's popularity (H2). A significantly negative correlation of the same-purpose attribute is shown for 37% of the users, contrary to the statement in H3. We hypothesize that the low evidence for H2 and the evidence contradicting H3 are due to the strong importance of the *home*- and *work*- locations in weighted networks. Since by far most of the transitions between locations start or end at *home*, and there is usually only one location labelled as *home*, most

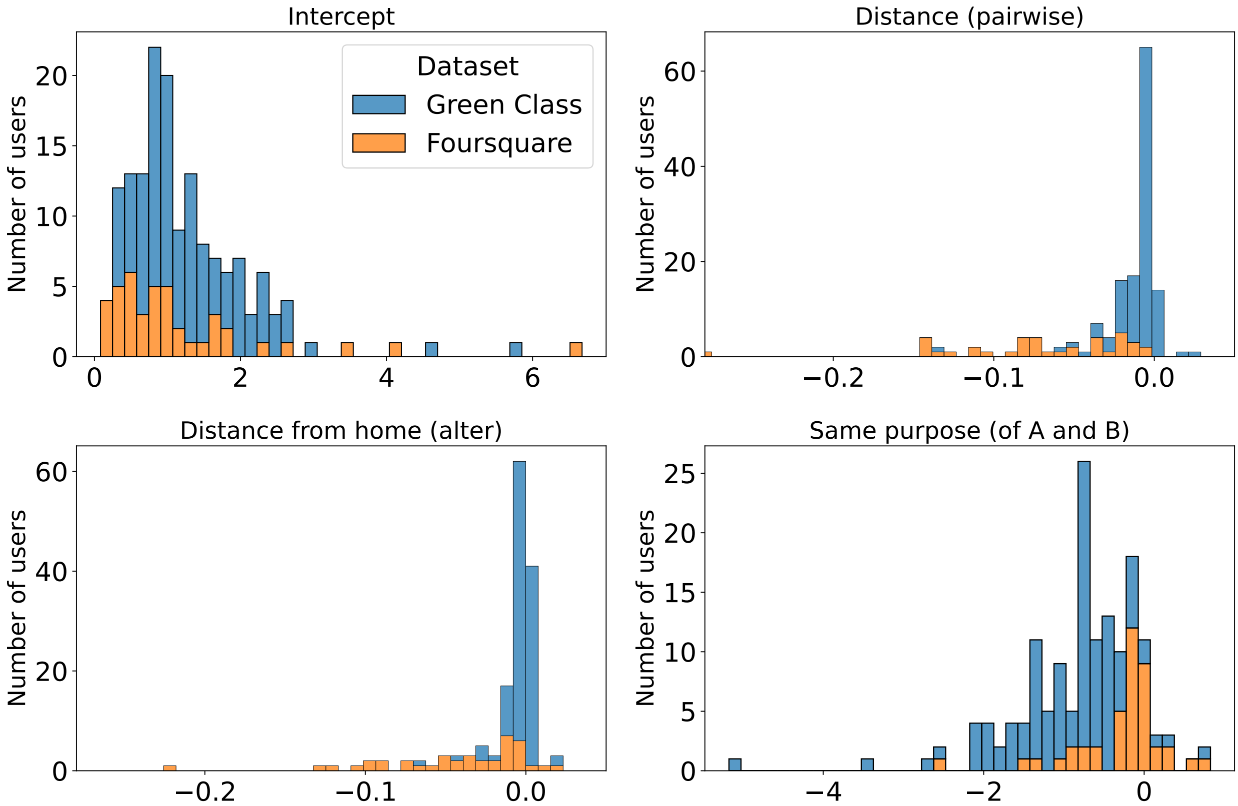

Figure 2. Distribution of the MR-QAP coefficients over all users of both datasets (42 Foursquare and 114 Green Class). For a vast majority of users, the estimated intercept is positive and the distances between locations and their purpose accordance have a negative effect on the edge weight. The effect of the alter's distance from home is close to zero for Green Class users but has a negative effect for many Foursquare users.

transitions occur between locations of different purpose, contradicting H3. Furthermore, the distance from home might play only a minor role because of the high importance of the *work* location which is not necessarily very close to *home*. Experiments on unweighted networks and further attributes could provide further insights but are out of the scope of this paper.

3.3 Comparing inter- and intra-user variances

Last, we validate the stability of user-wise estimates over time to distinguish differences between individuals from differences due to network instability. The variance of the coefficients over users (as shown in Figure 2) is now compared to the intra-user variances (the variance of a single user's results over time). Let $\theta_{u,t,k}$ now denote the coefficient θ_k estimated for the graph G_u^t. For inter-user variance we first take the average over time $\overline{\theta}_{u,k} = \frac{1}{n}\sum_t \theta_{u,t,k}$ and then the variance over users $\mathrm{Var}_u(\overline{\theta}_{u,k})$. For intra-user variance we instead compute the variance over time steps and then aggregate the results: $\frac{1}{n}\sum_u \mathrm{Var}_t(\theta_{u,t,k})$. Note that Var_t is the average over *three* time steps only and must be taken with a grain of salt. The aggregated result over users should however give a good estimate of intra-user variance. Table 2 shows the results by variable. The fact that intra-user variance is much lower than inter-user variance indicates that there is a certain consistency in the mobility behavior over time, and most of the coefficient-variances are due to actual differences between individuals.

	Intra-user std	Inter-user std
Intercept	0.269	1.019
Distance	0.017	0.109
Distance from home (alter)	0.014	0.100
Same purpose	0.164	0.806
Average	0.116	0.509

Table 2. Comparing intra- and inter user variances of the estimated coefficients in MR-QAP. The variance over time for one user ("intra-user") is significantly lower than the variance over users, confirming the stability of user-wise results.

4 Modelling network dynamics

MR-QAP can yield insights about how attributes are associated with the existence or weight of ties. However, in

MR-QAP the network structure is *fixed*, such that it can not test any hypotheses on the network topology. For example, we are interested to understand whether there is a tendency for reciprocity or transitivity in the structure of mobility networks. Intuitively, we would expect that a user may use a link between two locations in the opposite direction as well (reciprocity), and that users may take a shortcut between two locations that were previously visited only via a third location (transitivity). To examine such structural effects we propose to consider how mobility networks evolve and to determine the factors that drive changes in their topology.

Network dynamics, i.e. changes to the connections in a network over time, can be modelled with Stochastic Actor-Oriented Models (SAOMs) (Snijders, 2001, 2017; Snijders et al., 2010). The underlying model of SAOMs is a continuous-time Markov chain where network states are observed at discrete time steps. Here, we regard the unweighted location graphs since we are interested to study the *existence* of connections between locations (instead of the *number* of transitions). Over the course of several months, transitions between pairs of locations appear that have not been used before, and other transitions are not used anymore. We regard the former as adding a tie to the network in one step and the later as a tie deletion. In the following, we first explain the SAOM method and describe our formulation of mobility networks that enables its application. We then answer the following questions using SAOMs applied to the available datasets.

- Q1: Is there a tendency for ties to be reciprocated, or for new ties to close a transitive triad?
- Q2: How does the purpose of a location affect the formation of new ties to other locations?
- Q3: Can we observe a preference for new ties to be added between locations that are close to another (or are ties between distant locations dropped more often)?

4.1 Stochastic Actor Oriented Models (SAOM)

SAOMs have been used to model the evolution of friendship (Boda et al., 2020), collaborative networks (Cao et al., 2017), or financial networks (Chu et al., 2021). In social sciences, the nodes in the graph are actors while the edges describe a relation between actors, e.g. friendship. SAOMs take an actor-centered perspective and model changes of network ties as a Poisson process, where an actor i decides to change one of its ties based on behavioral or network states $s_{ik}(G)$, with k as the index of the attribute or network property Φ_k. For example, Φ_1 could be reciprocity and $s_{11}(G)$ is then the number of reciprocal dyads attached to actor 1. As explained in Snijders et al. (2010), the decision whether to add or drop a tie is based on the following objective function:

$$f(i,G,\beta) = \sum_k \beta_k \cdot s_{ik}(G) \ ,$$

where G is the current network and β is the vector of behavioral / structural preferences. Intuitively, the greater $|\beta_k|$, the more relevant is the k-th attribute to the network evolution. The probability to change a specific tie is then given by the normalized value of the objection function:

$$P(i \longrightarrow j|x,\beta) = \frac{\exp\left(f(i,x^{\pm ij},\beta)\right)}{\sum_k \exp\left(f(i,x^{\pm ik},\beta)\right)}$$

where $x^{\pm ik}$ is the network state where the edge from i to k is changed (dropped or added dependent on the current state). Together, the network statistics $s_{ik}(x^{\pm ij})$, weighted by β_k, are compared to the network statistics in all other scenarios; i.e. when a different tie $i \longrightarrow k$ is changed or nothing changes ($x^{\pm ii}$).

4.1.1 A SAOM-compatible formulation of human mobility networks

One may object that an actor-oriented perspective is unsuitable in the context of mobility, because the nodes in our network are *locations* (and not actors). In response, we draw attention to the user's movement in the network. At any point in time, the user is located at one node and transitions to a new node along known or new connections. In other words, the *actor* in the location graph is the user at a location, *deciding* about a location's transition. The actor can decide to take a new transition from location l_i to l_j that was not done before (equivalent to adding a tie between l_i and l_j). On the other hand, we define edges to be dropped if the corresponding transitions have not been used for more than 120 days. We further justify the application of SAOMs on mobility data by considering four key assumptions of SAOMs.

1. **The network panel data are the outcome of a continuous-time Markov chain:** Arguably, this assumption is fulfilled to similar extent as for social networks. The formation of new friendships is not only dependent on the current state of the network, but also on previous time steps. In the location graph, the actor can choose freely to visit a new location based on the current state, but edge deletion is dependent on previous decisions.
2. **Actors control their outgoing ties:** With the proposed formulation this assumption is fully satisfied. The actor (the user at a location) controls its next goal, i.e. the target node of a new tie.
3. **Only one tie can change at a time:** While this assumption is not necessarily fulfilled for social networks (e.g. two friendships may be formed at the same time), it is inherently fulfilled by physics here: A user can not be at two locations simultaneously and thus ties can change only sequentially.

4. **Actors have full knowledge of the network:** Again, this property is satisfied even better for location graphs than for social networks. In the location graph, the actor can remember most of its visited places and the trips inbetween, whereas social agents are not aware of all relations between the other agents.

We conclude that the assumptions of SAOMs are fulfilled and SAOMs can be applied. We will further comment on its validity in subsubsection 4.3.1 and 4.3.2.

4.1.2 Implementation and model details

We model selection processes with a SAOM implemented in the RSiena package (Ripley et al., 2011). The input for user u are the three networks $\hat{G}^0_u, \hat{G}^1_u$ and $\hat{G}^2_u$ where $\hat{G}$ is the unweighted version of G. We further provide the stable attributes for each location which do not change from $\hat{G}^0$ to $\hat{G}^2$, i.e. geographic distances between locations, their distance from home and the location purposes. Here we consider both *ego* and *alter* effects for the monadic covariates (distance from home and purpose), as shown in Table 3. In terms of a SAOM, the distinction between *ego* and *alter* means that an actor decides about new ties based on features of its current location (*ego*) or the features of the selected destination node (*alter*).

In addition to these attributes, we also include network-structure statistics, namely reciprocity, transitivity, outdegree density and outdegree activity. The outdegree density effect must be included (Ripley et al., 2011) because it balances the creation and termination of ties over time and thus acts as an intercept. The outdegree activity effect on the other hand "reflects tendencies for actors with high out-degrees to send out extra outgoing ties 'because' of their high current out-degrees."(Ripley et al. (2022), p.46). This effect was included to take into account the scale-free nature of human mobility networks.

The SAOM model is fitted with the `siena07` function of the RSiena package, with four subphases in phase 2 and 3000 iterations in phase 3 as recommended. `siena07` implements the Robbins-Monro (Robbins and Monro, 1951) algorithm to estimate parameters with the Methods of Moments. The model is fitted for all subjects of both datasets individually. Model convergence is measured with the t-ratios of convergence which is supposed to be below 0.1 (Snijders et al., 2010). Here, we observe that the model converges well for a large majority of users; the criterion is fulfilled for 142 out of 156 models. We exclude the ones that did not sufficiently converge. Furthermore, as in (Snijders et al., 2010) we compute the p-values of all effects, dividing the estimates by their standard error. As before, the output for user 327 is given as an example in Table 3.

effect	θ	std	p-value	t-ratios
rate (period 1)	9.11	2.15		
rate (period 2)	15.91	6.58		
reciprocity	0.65	0.23	0.00	0.04
transitive triplets	0.26	0.07	0.00	0.00
outdegree (density)	-2.11	0.17	0.00	0.06
outdegree (activity)	0.03	0.02	0.09	0.03
distance	-0.12	0.02	0.00	-0.02
purpose alter	0.31	0.09	0.00	-0.05
purpose ego	0.29	0.11	0.01	0.02
same purpose	-0.31	0.17	0.06	0.02
dist_home alter	0.01	0.02	0.49	-0.05
dist_home ego	0.02	0.02	0.33	0.01

Table 3. Output of SAOM fitted to a single user (User 327, see Figure 1). The model converged well, i.e. all t.conv values are below 0.1. For this user, there is evidence for a significant effect of reciprocity, transitivity, pairwise distance and purpose (ego and alter) on the formation of new ties.

4.2 Comparing SAOM coefficients across users and datasets

We aggregate the results in Table 4. In a majority of subjects, there is evidence for reciprocity as a selection bias for building new ties (77%), whereas the transitivity is only significant for 27% (Q1). As expected, the mobility networks as scale-free networks oftentimes show a tendency to form new ties from locations that have high outdegree already (41% are significant for a positive effect of outdegree activity). The outdegree (density) statistic is strongly negative, indicating that over the observed time steps more ties were removed than deleted. With respect to location properties, the negative effect of the distance between two locations (Q3) is strongly supported (62%), and there is also a negative effect of the same-purpose relation (Q2), as observed in the MR-QAP analysis (44%).

Finally, we compare the results of the estimated coefficients between both datasets. Clearly, there is a strong correspondence between the effects observed in the Green Class and Foursquare data. There are only very few significant estimates that point in the opposite direction (Table 4 left). This indicates the existence of general properties of human mobility behavior that appear in very diverse datasets. Meanwhile, the magnitude of coefficients and rates differ. The higher rate in the Foursquare data implies stronger changes between the networks of subsequent time periods (compare section 4.3.1). We also compare the coefficients of Green Class users to the ones of Foursquare users in a two-sided independent t-test, and find significant differences for most effects (rate 1 and 2, reciprocity, outdegree density, distance, same purpose, and distance home ego). Further work could investigate which properties of the data cause the observed differences.

	Significant (%) > 0	< 0	Average coefficient Green Class	Foursquare
rate (period 1)			3.65 ± 1.97	6.45 ± 4.65
rate (period 2)			4.17 ± 2.20	7.53 ± 5.50
reciprocity	77	1	1.75 ± 0.87	0.72 ± 0.56
transitive triplets	27	1	0.12 ± 0.27	0.21 ± 0.19
outdegree (density)	0	91	-2.19 ± 0.80	-1.90 ± 0.55
outdegree (activity)	41	0	0.05 ± 0.08	0.03 ± 0.07
distance	0	62	-0.03 ± 0.04	-0.10 ± 0.14
purpose alter	32	2	0.33 ± 0.75	0.28 ± 0.36
purpose ego	9	3	0.07 ± 0.87	0.20 ± 0.26
same purpose	1	44	-1.06 ± 1.05	-0.24 ± 0.32
distance home alter	11	1	-0.00 ± 0.04	-0.02 ± 0.10
distance home ego	11	1	0.00 ± 0.04	-0.05 ± 0.15

Table 4. Aggregated coefficients of a SAOM fitted to all users. The left part shows the percentage of users with significantly positive or significantly negative coefficients. The right part lists the average coefficient values by dataset. When mobility networks evolve, there is a tendency for reciprocal and transitive ties, and ties between locations that are close to each other. While most effects appear to similar extent in both datasets, the change rate is larger in the Foursquare dataset.

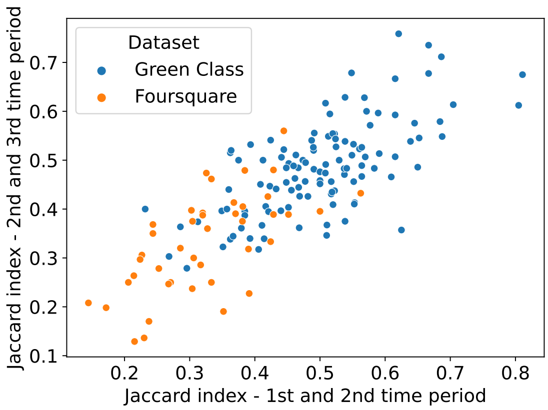

Figure 3. Correlation of the two Jaccard indices over three time periods. The Jaccard index is in general higher for the Green Class users, which thus seem to have more stable location networks. The indices of the two time steps are strongly correlated, providing evidence for a user-specific network stability. Most indices lie between 0.3 and 0.6 such that a SAOM is applicable.

4.3 Validation

4.3.1 Measuring network-stability with Jaccard indices

For a successful application of SAOMs, the network should change over time, although not too much. The network change can be quantified with the Jaccard index (Jaccard, 1900), as explained in Snijders et al. (2010). The Jaccard index between $\hat{G}_u^t$ and $\hat{G}_u^{t+1}$ is the fraction of the number of ties present in both $\hat{G}_u^t$ and $\hat{G}_u^{t+1}$ divided by the union of ties in $\hat{G}_u^t$ and $\hat{G}_u^{t+1}$. Formally, with N_{11} as the number of ties present in both graphs, and N_{10} and N_{01} the number of ties present only in the first or second graph respectively, the Jaccard index is defined as

$$\mathcal{J}(\hat{G}_u^t, \hat{G}_u^{t+1}) = \frac{N_{11}}{N_{11} + N_{10} + N_{01}}$$

If there are many changes, the Jaccard index is low. According to Snijders et al. (2010), a Jaccard index above 0.6 is preferred for a SAOM, but an index between 0.3 and 0.6 is acceptable. Here, we find an average Jaccard index of 0.45 (± 0.13). Figure 3 depicts the distribution of indices and the relation between $\mathcal{J}(\hat{G}_u^1, \hat{G}_u^2)$ and $\mathcal{J}(G_u^2, G_u^3)$. We first observe that in a large majority of cases, the Jaccard index is sufficient for the application of a SAOM. Secondly, Figure 3 shows that the networks are more stable in the Green Class dataset, corresponding to the difference in rates in the model (Table 4). Third, there is a strong correlation (Pearson $r = 0.78$) between both Jaccard indices. It seems that the change rate in mobility networks are highly user-specific and stable over time.

4.3.2 Goodness-of-Fit evaluation

While the SAOM aims to fit reciprocity and transitivity statistics of the observed networks, it does not take into account other network characteristics, such as the indegree distribution. Thus, the model Goodness-of-Fit (GoF) can be estimated from such unmodelled statistics. As common in applications of SAOMs, we compare the indegree-, outdegree- and geodesic distribution (Kalish, 2020). The latter refers to the distribution of all-pair shortest path distances in a graph. We apply the RSiena function `sienaGOF` which runs multiple simulations and computes p-values to quantify the differences between observed and simulated networks. The p-values are computed with a Monte Carlo test based on the Mahalanobis distance (Ripley et al., 2011). The null hypothesis is that the degrees (or geodesic distances) of simulated and observed networks stem from the same distribution. Figure 4 showcases the GoF results for our running example, user 327. In the optimal case, the degree- or geodesic distribution of the simulated networks should match the one of the observed network. In Figure 4 the distributions correspond very well, and there is no evidence for significant differences ($p \geq 0.744$).

We evaluate the GoF for all included users, yielding $142 * 3$ p-values for the three GoF statistics. 75% of those values are above 0.05, demonstrating that for a majority of cases the observed network statistics lie within the 95% confidence interval of the simulated networks' distribution. 88% are within the 99% confidence interval. Since these statistics were not considered in the model fit, the results show a general ability of the model to yield similar network structures. Follow-up work could investigate whether including further covariates or parameter tuning may improve the GoF according to these metrics.

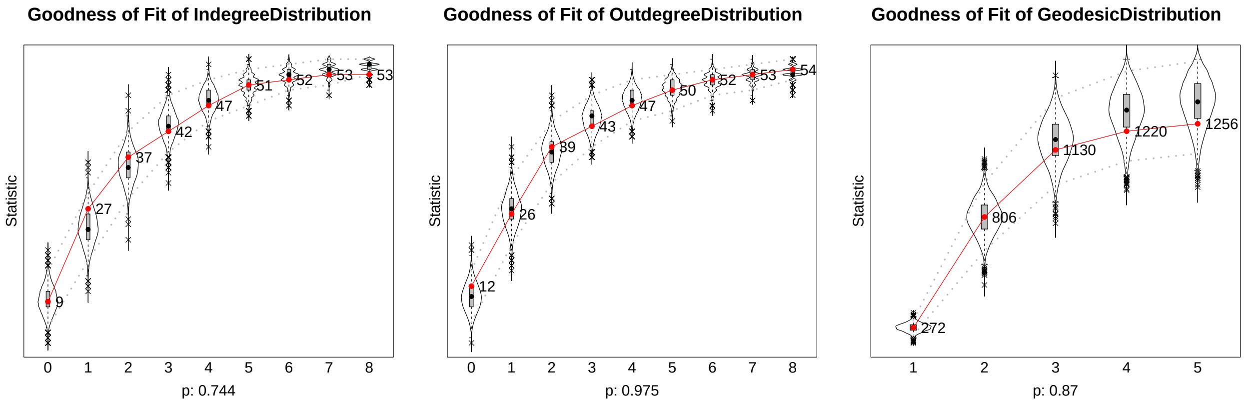

Figure 4. The Goodness-of-Fit for a SAOM can be evaluated by comparing the simulated and observed distributions of indegree, outdegree and geodesic distances. For this exemplary user (user 327) the model obtains a good fit as the distributions are not significantly different ($p > 0.05$). The distributions of the observed networks (red line) correspond well to the distribution of simulated networks (violin plots).

5 Discussion and conclusion

While human mobility behavior has been studied in many aspects, its network characteristics have received little attention. In contrast to previous work on human mobility (Alessandretti et al., 2020; González et al., 2008; Brockmann et al., 2006; Pappalardo et al., 2015; Schläpfer et al., 2021), the methods presented here not only regard the scales of mobility, but its *topology*, i.e. the relations between locations. We provided results on two datasets to demonstrate the new questions that can be answered about human mobility with network modelling methods.

First, the influence of location properties on the network structure could be quantified with MR-QAP. Our analysis provided evidence for a negative effect of the distance between locations; however, the results are strongly influenced by the prominence of the *home* and *work* locations in mobility networks. Our comparison across users and across distinct time periods corroborates the hypothesis that these characteristics are inherent to human mobility for the most part, for a smaller part due to individual choices, and only marginally related to changes over time. Furthermore, the evolvement of the unweighted networks over time was modelled with a SAOM. The SAOM converged well and offers insights in the processes how connections between locations are created and disregarded over time. There is evidence for a tendency towards reciprocal and transitive ties, as well as for a tendency to form new ties starting from locations that already show high activity. Together, both methods revealed interesting properties of individual human mobility and our analysis thereby demonstrates the value of network modelling methods in movement analysis.

We see several opportunities for future work. Using SAOMs for weighted networks could be used to answer questions related to changes in the *quantity* of transitions between locations. Other methods such as Exponential Random Graph Models (ERGMs, see for example Lusher et al. (2013)) could yield insight into the processes that form location graphs. Equipping the location graph with additional attributes may yield further insights about human mobility. For example, preferences for transitions between locations could be related to a location's surroundings, i.e. POI data, or its general visit frequency of other people. Thus, we conclude that the application of network modelling methods is an interesting endeavour for future research on mobility, and could reveal novel findings on the factors that drive human movement.

Data and software availability

The source code to reproduce our results on the public Foursquare dataset is available at https://github.com/NinaWie/network_analysis.

Acknowledgements

This work was supported by the ETH Zurich Foundation [MI-01-19].

Many thanks to Dr. Viviana Amati and Lea Tamberg for their valuable feedback.

References

Alessandretti, L., Sapiezynski, P., Sekara, V., Lehmann, S., and Baronchelli, A.: Evidence for a conserved quantity in human mobility, Nature human behaviour, 2, 485–491, 2018.

Alessandretti, L., Aslak, U., and Lehmann, S.: The scales of human mobility, Nature, 587, https://doi.org/10.1038/s41586-020-2909-1, 2020.

Altshuler, Y., Puzis, R., Elovici, Y., Bekhor, S., and Pentland, A.: Augmented betweenness centrality for mobility prediction in transportation networks, Finding Patterns of Human Behaviors in Network and Mobility Data (NEMO), pp. 1–12, 2011.

Axhausen, K. W.: Definition of movement and activity for transport modelling, in: Handbook of transport modelling, Emerald Group Publishing Limited, 2007.

Barthélemy, M.: Spatial networks, Physics reports, 499, 1–101, 2011.

Ben-Gal, I., Weinstock, S., Singer, G., and Bambos, N.: Clustering users by their mobility behavioral patterns, ACM Transactions on Knowledge Discovery from Data (TKDD), 13, 1–28, 2019.

Blanford, J. I., Huang, Z., Savelyev, A., and MacEachren, A. M.: Geo-located tweets. Enhancing mobility maps and capturing cross-border movement, PloS one, 10, e0129 202, 2015.

Boda, Z., Elmer, T., Vörös, A., and Stadtfeld, C.: Short-term and long-term effects of a social network intervention on friendships among university students, Scientific reports, 10, 1–12, 2020.

Brockmann, D., Hufnagel, L., and Geisel, T.: The scaling laws of human travel, Nature, 439, 462–465, https://doi.org/10.1038/nature04292, 2006.

Bucher, D.: Spatio-Temporal Information and Communication Technologies Supporting Sustainable Personal Mobility, Ph.D. thesis, ETH Zurich, 2020.

Butts, C. T.: Social network analysis with sna, Journal of statistical software, 24, 1–51, 2008.

Cao, D., Li, H., Wang, G., Luo, X., Yang, X., and Tan, D.: Dynamics of project-based collaborative networks for BIM implementation: Analysis based on stochastic actor-oriented models, Journal of Management in Engineering, 33, 04016 055, 2017.

Chen, C., Ma, J., Susilo, Y., Liu, Y., and Wang, M.: The promises of big data and small data for travel behavior (aka human mobility) analysis, Transportation research part C: emerging technologies, 68, 285–299, 2016.

Chu, A. M., Chan, L. S., and So, M. K.: Stochastic actor-oriented modelling of the impact of COVID-19 on financial network evolution, Stat, 10, e408, 2021.

Curtin, K. M.: Network analysis in geographic information science: Review, assessment, and projections, Cartography and geographic information science, 34, 103–111, 2007.

Dekker, D., Krackhardt, D., and Snijders, T. A.: Sensitivity of MRQAP tests to collinearity and autocorrelation conditions, Psychometrika, 72, 563–581, 2007.

Demšar, U., Long, J. A., Benitez-Paez, F., Brum Bastos, V., Marion, S., Martin, G., Sekulić, S., Smolak, K., Zein, B., and Siła-Nowicka, K.: Establishing the integrated science of movement: Bringing together concepts and methods from animal and human movement analysis, International Journal of Geographical Information Science, 35, 1273–1308, 2021.

Dodge, S., Weibel, R., Ahearn, S. C., Buchin, M., and Miller, J. A.: Analysis of movement data, 2016.

Dodge, S., Gao, S., Tomko, M., and Weibel, R.: Progress in computational movement analysis–towards movement data science, 2020.

Ellison, N. B., Steinfield, C., and Lampe, C.: The benefits of Facebook "friends:" Social capital and college students' use of online social network sites, Journal of computer-mediated communication, 12, 1143–1168, 2007.

Freedman, D. and Lane, D.: A nonstochastic interpretation of reported significance levels, Journal of Business & Economic Statistics, 1, 292–298, 1983.

Garton, L., Haythornthwaite, C., and Wellman, B.: Studying online social networks, Journal of computer-mediated communication, 3, JCMC313, 1997.

González, M. C., Hidalgo, C. A., and Barabási, A.-L.: Understanding individual human mobility patterns - Supplementary, Nature, 453, 779–782, https://doi.org/10.1038/nature06958, 2008.

Good, P.: Permutation tests: a practical guide to resampling methods for testing hypotheses, Springer Science & Business Media, 2013.

Harary, F., Norman, R. Z., and Cartwright, D.: Structural models: An introduction to the theory of directed graphs, Wiley, 1965.

Holme, P.: Congestion and centrality in traffic flow on complex networks, Advances in Complex Systems, 6, 163–176, 2003.

Hong, Y., Xin, Y., Martin, H., Bucher, D., and Raubal, M.: A Clustering-Based Framework for Individual Travel Behaviour Change Detection, in: 11th International Conference on Geographic Information Science (GIScience 2021)-Part II, 2021.

Hubert, L.: Assignment methods in combinational data analysis, vol. 73, CRC Press, 1986.

Jaccard, P.: Contribution au problème de l'immigration post-glaciare de la flore alpine, Bull Soc Vaudoise Sci Nat, 36, 87–130, 1900.

Johnson, J. and Schon, J.: Classifying refugee flows: A hierarchical clustering analysis of multiple regression-quadratic assignment procedure (MR-QAP) results that identify the drivers of refugee flow ego networks 2011-2016, in: AGU Fall Meeting Abstracts, vol. 2019, pp. GC11A–02, 2019.

Jonietz, D. and Bucher, D.: Continuous trajectory pattern mining for mobility behaviour change detection, in: LBS 2018: 14th International Conference on Location Based Services, pp. 211–230, Springer, 2018.

Kalish, Y.: Stochastic actor-oriented models for the co-evolution of networks and behavior: An introduction and tutorial, Organizational Research Methods, 23, 511–534, 2020.

Kish, L.: Coleman: Introduction to Mathematical Sociology (Book Review), Social Forces, 44, 252, 1965.

Krackardt, D.: QAP partialling as a test of spuriousness, Social networks, 9, 171–186, 1987.

Krackhardt, D.: Predicting with networks: Nonparametric multiple regression analysis of dyadic data, Social networks, 10, 359–381, 1988.

Kraemer, M. U., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D. M., Group†, O. C.-. D. W., du Plessis, L., Faria, N. R., Li, R., et al.: The effect of human mobility and control measures on the COVID-19 epidemic in China, Science, 368, 493–497, 2020.

Landherr, A., Friedl, B., and Heidemann, J.: A critical review of centrality measures in social networks, Business & Information Systems Engineering, 2, 371–385, 2010.

Laube, P.: Computational movement analysis, Springer, 2014.

Lusher, D., Koskinen, J., and Robins, G.: Exponential random graph models for social networks: Theory, methods, and applications, vol. 35, Cambridge University Press, 2013.

Mantel, N.: The detection of disease clustering and a generalized regression approach, Cancer research, 27, 209–220, 1967.

Martin, H., Bucher, D., Suel, E., Zhao, P., Perez-Cruz, F., and Raubal, M.: Graph convolutional neural networks for human activity purpose imputation, in: NIPS spatiotemporal workshop at the 32nd Annual conference on neural information processing systems (NIPS 2018), 2018.

Martin, H., Becker, H., Bucher, D., Jonietz, D., Raubal, M., and Axhausen, K. W.: Begleitstudie SBB Green Class-Abschlussbericht, Arbeitsberichte Verkehrs-und Raumplanung, 1439, 2019.

Martin, H., Hong, Y., Wiedemann, N., Bucher, D., and Raubal, M.: Trackintel: An open-source Python library for human mobility analysis, arXiv preprint arXiv:2206.03593, 2022.

Mascia, D., Di Vincenzo, F., Iacopino, V., Fantini, M. P., and Cicchetti, A.: Unfolding similarity in interphysician networks: the impact of institutional and professional homophily, BMC health services research, 15, 1–8, 2015.

Noben, I., Brouwer, J., Deinum, J. F., and Hofman, W. A.: The development of university teachers' collaboration networks during a departmental professional development project, Teaching and Teacher Education, 110, 103 579, 2022.

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., and Mascolo, C.: A tale of many cities: universal patterns in human urban mobility, PloS one, 7, e37 027, 2012.

Papandrea, M., Jahromi, K. K., Zignani, M., Gaito, S., Giordano, S., and Rossi, G. P.: On the properties of human mobility, Computer Communications, 87, 19–36, 2016.

Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., and Barabási, A.-L.: Returners and explorers dichotomy in human mobility, Nature communications, 6, 1–8, 2015.

Park, J. H., Kwahk, K.-Y., Han, H., and Kim, Y.: A Study on the Determinants of Patent Citation Relationships among Companies: MR-QAP Analysis, Journal of Intelligence and Information Systems, 19, 21–37, 2013.

Rinzivillo, S., Gabrielli, L., Nanni, M., Pappalardo, L., Pedreschi, D., and Giannotti, F.: The purpose of motion: Learning activities from individual mobility networks, in: Data Science and Advanced Analytics (DSAA), 2014 International Conference on, pp. 312–318, IEEE, 2014.

Ripley, R. M., Snijders, T. A., Boda, Z., Vörös, A., and Preciado, P.: Manual for RSiena, University of Oxford, Department of Statistics, Nuffield College, 1, 2011, 2011.

Ripley, R. M., Snijders, T. A., Boda, Z., Vörös, A., and Preciado, P.: Manual for RSiena, University of Oxford, Department of Statistics, Nuffield College, 1, 46, http://www.stats.ox.ac.uk/~snijders/siena/RSiena_Manual.pdf, 2022.

Robbins, H. and Monro, S.: A stochastic approximation method, The annals of mathematical statistics, pp. 400–407, 1951.

Schläpfer, M., Dong, L., O'Keeffe, K., Santi, P., Szell, M., Salat, H., Anklesaria, S., Vazifeh, M., Ratti, C., and West, G. B.: The universal visitation law of human mobility, Nature, 593, 522–527, https://doi.org/10.1038/s41586-021-03480-9, 2021.

Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., and González, M. C.: Unravelling daily human mobility motifs, Journal of The Royal Society Interface, 10, 20130 246, 2013.

Scott, J.: Social network analysis, Sociology, 22, 109–127, 1988.

Smith, R. A.: Measuring neighborhood cohesion: A review and some suggestions, Human Ecology, 3, 143–160, 1975.

Snijders, T. A.: The statistical evaluation of social network dynamics, Sociological methodology, 31, 361–395, 2001.

Snijders, T. A.: Stochastic actor-oriented models for network dynamics, Annual Review of Statistics and Its Application, 4, 343–363, 2017.

Snijders, T. A., Van de Bunt, G. G., and Steglich, C. E.: Introduction to stochastic actor-based models for network dynamics, Social networks, 32, 44–60, 2010.

Stadfeld, C. and Amati, V.: Network mechanisms and network models, in: Research Handbook on Analytical Sociology, Edward Elgar Publishing, 2021.

Strogatz, S. H.: Exploring complex networks, nature, 410, 268–276, 2001.

Wang, Y., Qin, K., Chen, Y., and Zhao, P.: Detecting anomalous trajectories and behavior patterns using hierarchical clustering from taxi GPS data, ISPRS International Journal of Geo-Information, 7, 25, 2018.

Yang, D., Zhang, D., Chen, L., and Qu, B.: Nationtelescope: Monitoring and visualizing large-scale collective behavior in lbsns, Journal of Network and Computer Applications, 55, 170–180, 2015.

Yang, D., Zhang, D., and Qu, B.: Participatory cultural mapping based on collective behavior data in location-based social networks, ACM Transactions on Intelligent Systems and Technology (TIST), 7, 1–23, 2016.

Zhao, Z., Koutsopoulos, H. N., and Zhao, J.: Detecting pattern changes in individual travel behavior: A Bayesian approach, Transportation research part B: methodological, 112, 73–88, 2018.