Capturing relationship strength: a choice model for leisure time, frequency of interaction and ranking in name generators

Abstract

In the past few decades, the travel behaviour literature has devoted increasing attention to understanding the demand for leisure and social travel and the engagement in leisure activities. Some of the studies in this field have adopted a social network perspective, acknowledging that it is mainly the people involved motivating such activities and travel. It is in this literature that the present study places itself. We develop a joint choice model to analyse the share of time spent in leisure activities with each social contact, the frequency of interaction by different modes and the ranking in a name generator. We show that these different decisions are linked by an underlying latent factor that we refer to as relationship strength. As this relationship strength cannot be directly observed, we use a number of different indicators for it in measurement models, including what we believe to be a novel use of the position in which a social contact is ranked in the name generator. The study sheds light on the concept of relationship strength, which is believed to be crucial for understanding social interactions and leisure activity engagement. The results of our joint model are in line with expectations and improve the understanding of relationship strength on the basis of the nature of the relationship and homophily measures.

Keywords: social interactions, hybrid choice model, time use, name generator

1 Introduction

In the last few decades, many studies have discussed the importance of understanding engagement in leisure and social activities, both because they make up a substantial part of daily life and because of its implications for travel (Mokhtarian et al., 2006; Tilahun and Levinson, 2017). In 2017, 25% of all trips and 40% of the miles travelled per year in England were for social and leisure purposes (Department for Transport, 2018). These shares have not greatly changed from the mid-nineties, and similar statistics were reported for other European countries such as Switzerland (Ohnmacht et al., 2009) and Germany (Stauffacher et al., 2005). The time spent conducting leisure activities from the mid-seventies until today has increased for men, while there is no clear trend for women (Gimenez-Nadal and Sevilla, 2012). The latest statistics for the UK report that men spend on average six hours and nine minutes a day on leisure activities, as opposed to the five hour and 29 minutes for women (Office for national statistics, 2015).

Differently from commuting and other "compulsory" forms of travel, understanding the determinants of trips with leisure purposes presents some additional difficulties as this is believed to depend not only on destination but also on the people involved in the trip/activity. There is a limited literature on the factors influencing the engagement in leisure activities (as opposed to the work on commuting and shopping trips), within which a small number of studies have adopted a social network approach by trying to link the characteristics of the relationships within personal networks to engagement in social and leisure activities.

An example is represented by Carrasco and Miller (2009), who use a disaggregate perspective of personal social networks by incorporating the specific characteristics of social contacts as well as the overall network statistics in a multilevel model to explain engagement in social activities. Their paper serves as a proof of concept as it demonstrates the crucial role of the characteristics of social contacts (or *alters*) on the frequency of social activities by the respondent (or *ego*). In the dataset used by Carrasco and Miller (2009), *egos* are asked to complete a standard "name generator" questionnaire, i.e. to list their *alters*, differentiating those who are perceived as *very close*. This piece of information is later used in the model as an independent variable. A very similar name generator was used by van den Berg et al. (2013) to study social activity-travel patterns and ICT use. A slightly different wording is instead used by Maness (2017), who employs a survey in which respondents are asked to list the people "who are especially significant in their lives", where this information is then used to model leisure activity frequency and variety. Another variant, adopted by Kim et al. (2017) in a paper to assess the role of social influence in car-sharing decisions, consists in asking respondents to assess their closeness to each *alter* on a 5-point Likert scale ranging from "Not close" to "Very close".

Data distinguishing close and weak ties have also been used in models of social interactions. Kowald (2013), like Carrasco and Miller (2009), uses closeness (as stated by the *egos*) to explain interaction frequency, finding that close social contacts were contacted more frequently than others. The results obtained by Kowald (2013) via multilevel models to separately account for *ego* and *ego-alter characteristics* were replicated by Calastri et al. (2017), who made use of advanced choice models, confirming the finding related to closeness in interactions. Frei (2012) also incorporated *ego-alter* information in a model of social interaction, highlighting the importance of distance between the two, when this is available. While this list is not exhaustive, it gives an idea of the nature of the studies that have made a link between engagement in leisure activities or interactions and the concept of relationship strength. Note that these studies have generally focused on modelling either leisure travel or social interactions. Like in the rest of the travel behaviour literature, there is a shortage of work attempting to jointly examine different dimensions of behaviour and choices that may be linked with each other.

A key component of this area of research is a quantification of the notion of closeness between individual people. Several surveys directly ask respondents about closeness to other people and then use this as an explanatory variable in their models. A potential risk with such an approach is that stated closeness is not an error free measure and is subject to a number of respondent driven biases, not unlike answers to attitudinal questions (Kroesen and Chorus, 2018). It is thus arguably more appropriate to treat the answers to such questions as indicators of strength rather than direct measures. This then opens up the important possibility that a set of different such data points can be used for each ego-alter pair to gain a deeper and more robust understanding of relationship strength. Strength would then be recognised to be a complex latent construct, of which "stated" closeness and other variables are indicators rather than direct measures (c.f. Calastri et al., 2018b; Kim et al., 2017).

The concept of strength has since been defined in different ways and applied to understand different processes, such as job search (Granovetter, 2018), emotional support (Schaefer et al., 1981) and buyer-seller relationships (De Cannière et al., 2010; Stanko et al., 2007). More recently, different studies have tried to infer strength of online relationships, arguing for example that the frequency of online interactions is a measure of strength (Jones et al., 2013). The idea of strength of relationship has been adopted by travel researchers in the past few decades, while a parallel but separate literature in sociology has investigated this concept since its introduction by Granovetter (Granovetter, 1977). Already in the 1980s, Marsden and Campbell (1984) were discussing the need of further study of the concept of relationship strength, while describing a number of indicators to measure this construct, namely a measure of closeness, time spent and intensity. Since then, the list of measures of strength has been extended to also include communication reciprocity (Friedkin, 1980) and frequency of interaction (Gilbert et al., 2008), among others. This opens up the possibility of attempting to combine several such measures, as we do in our paper.

The objective of this paper is to gain insights about the social determinants of the choices of leisure time use, frequency and mode of interaction and ranking in the name generator. In addition, we investigate the correlation between these different choices and test whether the same factors affect the individual decisions that we are analysing.

We make use of a rich dataset to first unveil a common underlying construct to explain three objective measures, two of which have been used in the sociological literature as indicators of tie strength, namely frequency of interaction and time spent together. The third measure used is the order in which the contacts are named in a name generator when asking respondents to list the people with whom they were interacting out of work. We choose to use this measure for two main reasons. First of all, name generator and name interpreter surveys tend to be quite burdensome due to the large amount of information to be recalled. The ranking of social contacts in the name generator is collected in any case, with no additional effort for the respondent, and we seek to investigate whether this measure is an indicator of relationship strength. Our hypothesis is that respondents will instinctively first think of those people they are closer to. Secondly, we believe that even though the ranking is reported by the *ego* only, it is an indirect measure of closeness, and as such potentially preferable as it avoids the subjectivity in the assessment by each ego and makes use of a piece of information which is collected in any name generator survey without adding further respondent burden.

After defining our latent measure as relationship strength, we use it to explain the three separate indicators (time spent, frequency of interaction and ranking in the name generator) by the means of choice models. In particular, we apply a Tobit model to model the share of leisure time spent with each *alter*, five ordered logit models for the frequency of interaction (one for each of the five modes of interaction considered in this study) and an exploded logit for the ranking.

Our key research questions can therefore be summarised as follows:

- Are social interaction and ranking of alters in a name generators both affected by an unobserved construct which can be interpreted as relationship strength? Can this effect be shown quantitatively?
- What are the key characteristics of a social relationship that affect its strength?
- Can ranking of alters in name generators be seen as a proxy of relationship strength, offering potential for the reduction of respondent burden in social network surveys?

In the following section, we introduce the dataset used for the analysis and highlight its particular suitability to investigate the concept of tie strength. In Section 3, we describe the modelling methodology used for each of the three indicators and to create the link among them. In Section 4, we analyse the results of our preferred model, discuss its implications and draw some general conclusions.

2 Data

The dataset used in the present study was collected in Leeds (UK) in 2017. This online and smartphone-based survey aimed to collect a wide range of information, with a particular focus on social networks, life-course events and travel behaviour. For the purpose of the present paper, we exploit in particular the social network and travel behaviour data. With respect to the former, participants were asked to complete a name generator, in which they list all their social contacts with whom they "choose to regularly interact outside of work, either in person or via phone or digital media". After listing their social connections, they were asked to complete a *name interpreter*, in which they provide a number of details for each person they had listed. This included gender, age, nature of the relationship (friend, partner and so on), home location and frequency of interaction face to face, by phone (call and text), by email and via online social networks (OSN). After completing the survey, participants downloaded a customised version of the smartphone app *RMove* (Resource Systems Group, 2017) and used it for two weeks to track all their trips and activities, specifying who among the people listed in the name generator was with them at the time of each activity. We also make use of the information collected in the background questionnaire, such as socio-demographics of the respondents. More details about the survey can be found in Calastri et al. (2018a).

	N.egos	**% egos**	**N. ego-alter pairs**	**% ego-alter pairs**
Age				
18-29	107	24%	628	13%
30-49	224	51%	1,081	22%
over 50	112	25%	558	12%
Different age	-	-	2,538	53%
Gender				
Male	185	42%	1,044	22%
Female	258	58%	2,034	42%
Different sex	-	-	1727	36%
Location				
Live in Leeds	293	66%	1,711	36%
Live in W Yorkshire	42	9%	164	3%
Live elsewhere in the UK	108	24%	720	15%
Different areas	-	-	2,210	46%
Type of alter named in the name generator				
Friend	-	-	2,393	50%
Sibling	-	-	412	9%
Partner	-	-	328	7%
Colleague	-	-	313	7%
Parent	-	-	482	10%
Other relative	-	-	729	15%
Other acquaintance	-	-	145	3%

Table 1: Descriptive statistics

The present study uses a subsample of the original dataset. It includes only respondents who have provided full information about their trips in *RMove* and who did not present any inconsistency in the timing of activities. The final sample contains 443 *egos* who named $4,805$ *alters*. Table 1 summarises the sample, providing socio-demographic and network information at the *ego* and *alter* level.

The share of the different socio-demographic characteristics of the respondents included in the sample are in line with those in the original dataset (c.f. Calastri et al., 2018a). The sample was collected with the idea of achieving a variety of socio-demographic and mobility characteristics, and not with representativeness in mind. For further details please see (c.f. Calastri et al., 2018a). As shown by Table 1, the sample mainly includes people who live in Leeds as a result of the sampling strategy, but this does not affect the analysis carried out in this paper, as we aim to understand a specific process rather than being able to infer behavioural results for the entire population. We can also see that half of the *alters* listed in the name generator are "friends" and that the number of acquaintances not falling in any of the other categories is rather small.

Respondents could list a minimum of one social contact and a maximum of 40 (in line with similar studies), where we observe an average of 10.84 *alters* each. As Figure 1 shows, the distribution of the number of *alters* named is skewed to the right.

As mentioned above, *egos* were asked to specify the level of interaction with each *alter* by five different modes. For each mode, they stated the frequency of interaction from a dropdown menu listing 8 options, ranging from "multiple times per day" to "never". Figure 2 represents the average frequency of interaction between *egos* and *alters* for each type of *alter* and mode of interaction, where the thickness of the links represent the frequency of interaction and the colour distinguishes between *alter*-types. The data represented in the diagram reflect our expectations: for example, there is a frequent communication by any mode with partners, while interactions with colleagues and other relatives are mainly face-to-face.

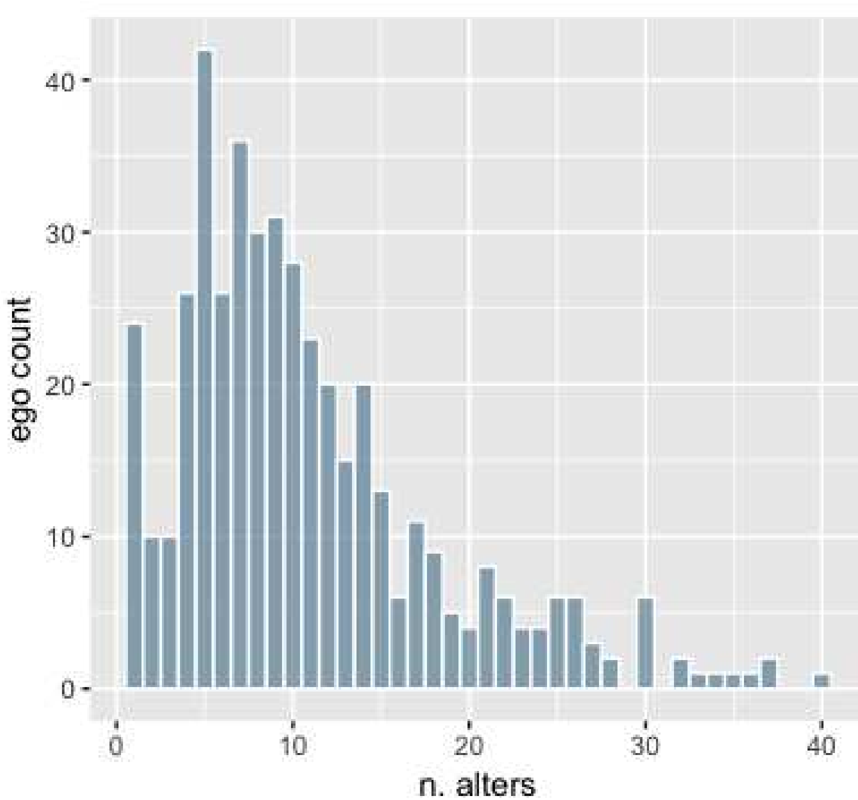

Figure 1: Number of $alters$ named in the name generator

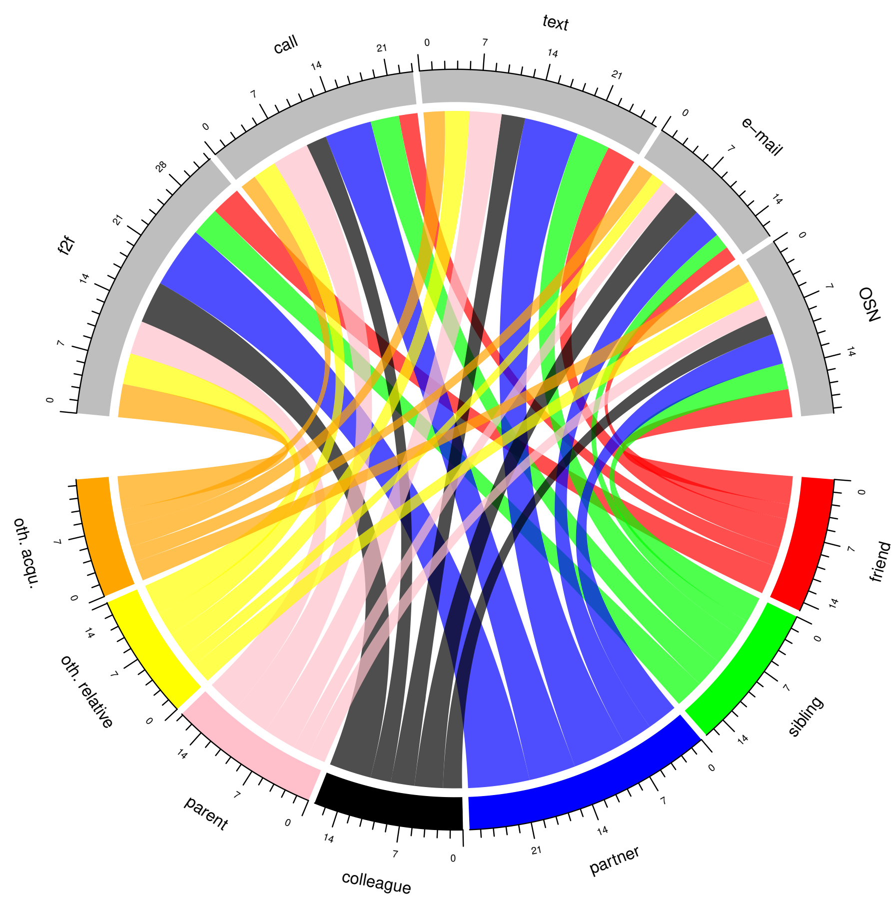

Figure 2: Interactions with *alters* by different modes

The *RMove* travel app collected detailed information about the different activities conducted by respondents over two weeks. Every time they would stop moving (i.e. travelling), participants were prompted to enter information such as travel mode, activity at the destination and who they were travelling with/meeting to perform the activity. For the given sample, we collected on average 5.2 trips per participant per day.

For our analysis of leisure time spent by each *ego* with a given *alter*, we use the share of the overall time spent conducting leisure and social activities with the given *alter*. If multiple people were present, we attributed the time to each of them, implying that the sum of the "shares" obtained can be larger than one. While this may sound like an unusual approach, it was adopted as the alternative would have been to devise a rule for assigning shares to each *alter*, and that would have been excessively subjective.

Figure 3 illustrates these average shares for each *alter*-type, where the vertical axis represents the share and the horizontal axis the position of the named contacts in the name generator. The graphs show how the share of time spent with each *alter*-type tends to decrease with the position in the name generator. An exception is seen for "partners", with high time shares independently of the ranking in the list, while we also observe one outlier among "siblings".

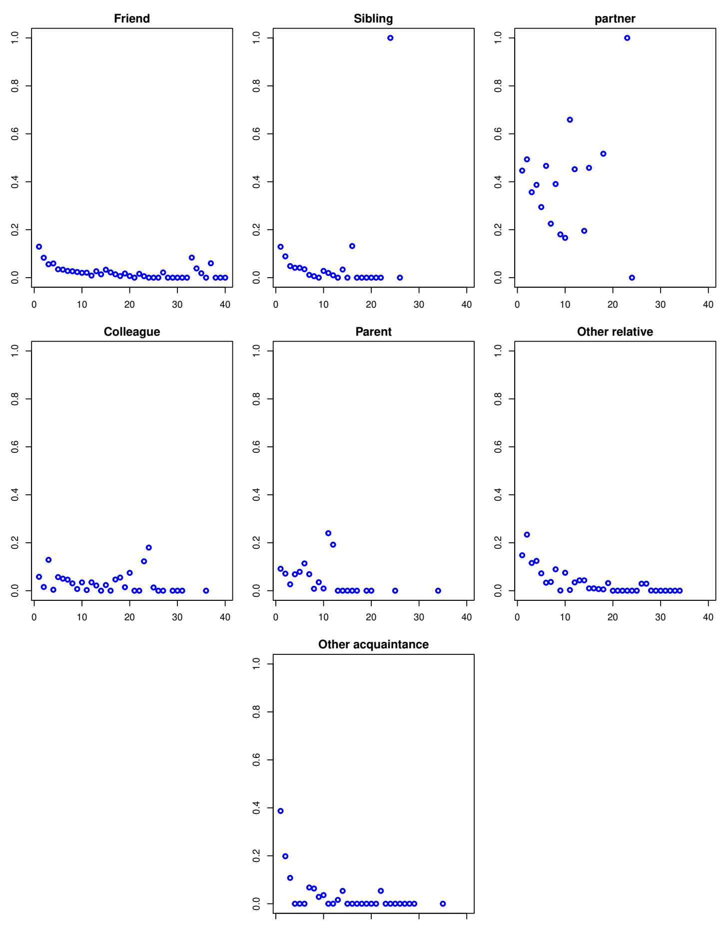

Figure 3: Average leisure time share spent with each *alter* type

3 Methods

Our modelling approach consists in developing separate models for each of the indicators and then test their correlation by means of a latent variable (LV). As mentioned in Section 1, we estimate:

- a Tobit model for the leisure time use component;
- five ordered logit models for the frequency of interaction by each of the five modes;
- an exploded logit model for the ranking in the name generator.

We specify the individual models as well as the equation of the LV as functions of *ego-alter* characteristics. Therefore, all models are formulated at the *ego-alter*-level. The model framework is illustrated in Figure 4. To make the illustration parsimonious and understandable, in the figure we slightly simplified the proposed model: it is important to remember that *ego* and alter-level characteristics were also used to explain the frequency of interaction and ranking choices, and that we do not explicitly show boxes for the utility of each choice.

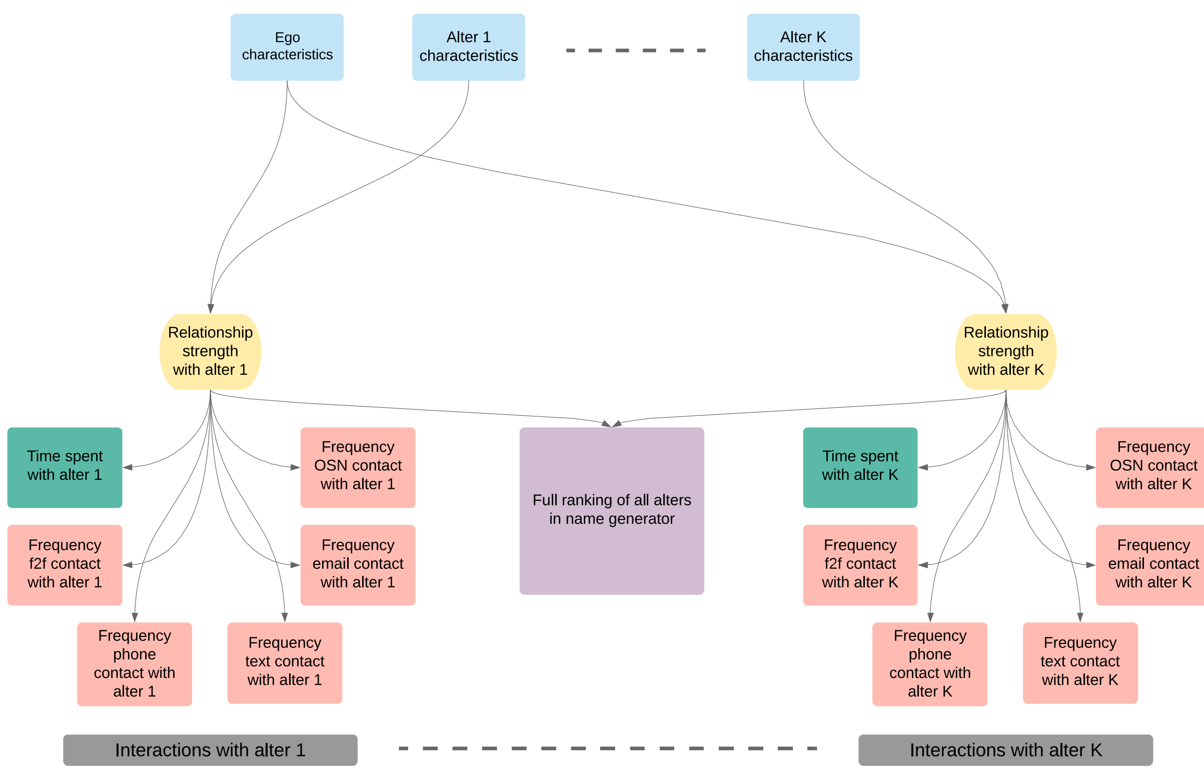

Figure 4: Model framework

In the remainder of this section, we will describe in detail the structure of each of the three components.

3.1 Leisure time use

A Type I left and right-censored Tobit model (Amemiya, 1984) is used to model the share of leisure time spent with each *alter*. The Tobit model assumes that there is an unobservable variable ${t_n}^*$, representing the real value of the dependent variable, i.e. in our case the percentage of leisure time an *ego* would like to spend with an *alter*, where an *ego-alter* pair is identified as n. The observed percentage of leisure time spent together by an *ego-alter* pair n is linearly related (through a set of coefficients $\boldsymbol{\beta}$, also including a constant β_0) to a set of independent variables $\boldsymbol{x_n}$. The observable variable t_n, is defined as a step-wise function. In particular, the dependent variable of our model is censored below and above 1, and we get that the likelihood of the n^{th} observation, i.e. the share of leisure time that *ego-alter* pair n spend together, is given by:

$$T_{t_n}(\sigma, \boldsymbol{\beta}) = \begin{cases} \frac{1}{\sigma}\phi\left(\frac{t_n - \boldsymbol{\beta x_n}}{\sigma}\right) & \text{if } 0 < {t_n}^* < 1 \\ 1 - \Phi\left(\frac{\boldsymbol{\beta x_n}}{\sigma}\right) & \text{if } {t_n}^* \leq 0 \\ 1 - \Phi\left(1 - \frac{\boldsymbol{\beta x_n}}{\sigma}\right) & \text{if } {t_n}^* \geq 1, \end{cases} \tag{1}$$

where σ is the standard deviation of the Normal distribution used for the error term of the model and ϕ and Φ are the density and the cumulative distribution function of the standard Normal distribution, respectively.

3.2 Frequency of interaction

The frequency of interaction with each *alter* by the five different modes is described by five variables ranging from 1 to 8, where a higher number represents more frequent interaction. Given the ordered nature of the variables, we make use of ordered logit models. As mentioned above, the lowest level of these variables, 1, corresponds to no interaction between the *ego-alter* pair by a given mode. In order to distinguish this case of "no contact" from the others, we estimate binary logit models (BL) to determine if there is any interaction at all, and we only apply the ordered logit models (OL) for values of the variables between 2 and 8. The model capturing the frequency of face-to-face interaction between the *ego-alter* pair n ($f_{n,f2f}$) has the following probability F:

$$F_{f_{n,f2f}}(\mu_{f2f}, \boldsymbol{\kappa}, \boldsymbol{\delta}) = BL_{f_{n,f2f}}(\mu_{f2f}) \cdot OL_{f_{n,f2f}}(\boldsymbol{\kappa}, \boldsymbol{\delta}). \tag{2}$$

The first term, BL, is the likelihood of the binary logit component, given by:

$$BL_{f_{n,f2f}}(\mu_{f2f}) = \begin{cases} \frac{1}{1+e^{\mu_{f2f}}} & \text{if } f_{n,f2f} = 1 \\ \frac{e^{\mu_{f2f}}}{1+e^{\mu_{f2f}}} & \text{if } f_{n,f2f} > 1 \end{cases} \tag{3}$$

The second term, OL, is the likelihood of the ordered component, given by:

$$OL_{(f_{n,f2f})}(\boldsymbol{\kappa}, \boldsymbol{\delta_{f2f}}) = \begin{cases} 1 & \text{if } f_{n,f2f} = 1 \\ \left(\frac{1}{1+e^{(-\kappa_q + \boldsymbol{\delta_{f2f} x_n})}} - \frac{1}{1+e^{(-\kappa_{q-1} + \boldsymbol{\delta_{f2f} x_n})}} \right) & \text{if } f_{n,f2f} = q,\ q > 1, \end{cases} \tag{4}$$

where the ordered logit model is only used for levels strictly greater than 1 and ignored for the lowest level, setting the likelihood to 1.

While the equations above only describe face-to-face interaction $f_{n,f2f}$ between *ego-alter* pair n, equivalent models for the other modes can be simply obtained by replacing $f2f$ by *call*, *text*, *email* and OSN. The mode-specific constant μ is the only estimated parameter entering the binary logit model in Equation 3 and $\kappa_1, \kappa_2 ... \kappa_q ... \kappa_Q$ are thresholds for each of the $Q + 1$ levels of the dependent variable. Note that the vector of parameters $\boldsymbol{\delta}$, measuring the impact of a set of variables x_n, is mode-specific.

3.3 Ranking in the name generator

In this component of the model we analyse the order in which social contacts are listed in the name generator by each survey respondent. Differently from the other components, this model is at the *ego* level, as it is the *ego* who chooses how to rank his/her *alters*. In order to model the ranking of a specific *ego*, we use an exploded logit model.

As all our models are estimated at the *ego-alter* level, we represent the decision of whether to position a specific *alter* in the available slot that needs to be filled in the name generator. In this case the exploded logit takes the form of a simple logit model, which will depend on the specific *alter* characteristics as well as on those of the remaining *alters* to be listed.

For example, let's say an *ego* has named n *alters* in the first n positions in the name generator. Assuming that he/she will name N *alters* in total, the probability of an *alter* being named next (in position $n+1$) and therefore before the remaining $n+2...N$ *alters* is equal to:

$$R_{n+1}(\boldsymbol{\theta}) = \frac{e^{\boldsymbol{\theta x}_{\text{alter n+1}}}}{\sum_{j=\text{alter n+1...alter N}} e^{\boldsymbol{\theta x}_{\text{alter j}}}} \tag{5}$$

where $\boldsymbol{\theta}$ represents a vector of estimated parameters measuring the impact of a set of independent variables $\boldsymbol{x_n}$ on the probability of the given *alter* being listed in a given position. It is clear that the last *alter* mentioned will have a probability of being named in the last position equal to 1, given that all the other *alters* have already been ranked.

3.4 The linked model

As mentioned above, the three model components just described are linked by means of a latent variable α_n. This recognises the fact that we can capture some of the variations across individual ego-alter pairs in this relationship strength through socio-demographic characteristics but there remains additional random variation. The latent variable α_n, which is used in all the different model components, is defined at the *ego-alter* level, so that there are up to 40 latent variables for each *ego*. It is specified as follows:

$$\alpha_n = \boldsymbol{\gamma x_{(n,\alpha_n)}} + \eta_n \tag{6}$$

where $\boldsymbol{\gamma}$ is a vector of parameters measuring the impact of *ego-alter* characteristics $\boldsymbol{x_{(n,\alpha_n)}}$ on the latent variable and η_n is a disturbance distributed at the *ego-alter* level following a Normal distribution with mean equal to 0 and standard deviation equal to 1.

A set of τ parameters (one for each model component) measure the additive effect of a latent variable α_n on the different utilities, so that, for example, the utility of the Tobit model $\boldsymbol{\beta x_n}$ becomes $\boldsymbol{\beta x_n} + \tau_T * \alpha_n$.

It is worth noting that in the estimation of this model, we do not estimate all the parameters defined in this section for identification purposes. In particular, we can write the individual model utilities as follows:

$$\begin{aligned}
t_n^* &= \beta_0 + \boldsymbol{\beta x_n} + \tau_T \cdot \alpha_n + \epsilon_T \\
U_{f2f} &= \boldsymbol{\delta_{f2f} x_n} + \tau_{f2f} \cdot \alpha_n + \epsilon_{f2f} \\
U_{call} &= \boldsymbol{\delta_{call} x_n} + \tau_{call} \cdot \alpha_n + \epsilon_{call} \\
U_{text} &= \boldsymbol{\delta_{text} x_n} + \tau_{text} \cdot \alpha_n + \epsilon_{text} \\
U_{email} &= \boldsymbol{\delta_{email} x_n} + \tau_{email} \cdot \alpha_n + \epsilon_{email} \\
U_{OSN} &= \boldsymbol{\delta_{OSN} x_n} + \tau_{OSN} \cdot \alpha_n + \epsilon_{OSN} \\
U_{ranking} &= \boldsymbol{\theta x_n} + \tau_R \cdot \alpha_n + \epsilon_R
\end{aligned}$$

If we rewrite these equations by replacing α_n with the expression in Equation 6, we get:

$$\begin{aligned}
t_n^* &= (\beta_0 + \boldsymbol{\beta} + \boldsymbol{\gamma}\tau_T)\boldsymbol{x_{(n,\alpha_n)}} + \epsilon_T + \tau_T \eta_n \\
U_{f2f} &= (\boldsymbol{\delta_{f2f}} + \boldsymbol{\gamma}\tau_{f2f})\boldsymbol{x_{(n,\alpha_n)}} + \epsilon_{f2f} + \tau_{f2f}\eta_n \\
U_{call} &= (\boldsymbol{\delta_{call}} + \boldsymbol{\gamma}\tau_{call})\boldsymbol{x_{(n,\alpha_n)}} + \epsilon_{call} + \tau_{call}\eta_n \\
U_{text} &= (\boldsymbol{\delta_{text}} + \boldsymbol{\gamma}\tau_{text})\boldsymbol{x_{(n,\alpha_n)}} + \epsilon_{text} + \tau_{text}\eta_n \\
U_{email} &= (\boldsymbol{\delta_{email}} + \boldsymbol{\gamma}\tau_{email})\boldsymbol{x_{(n,\alpha_n)}} + \epsilon_{email} + \tau_{email}\eta_n \\
U_{OSN} &= (\boldsymbol{\delta_{OSN}} + \boldsymbol{\gamma}\tau_{OSN})\boldsymbol{x_{(n,\alpha_n)}} + \epsilon_{OSN} + \tau_{OSN}\eta_n \\
U_{ranking} &= (\boldsymbol{\theta} + \boldsymbol{\gamma}\tau_R)\boldsymbol{x_{(n,\alpha_n)}} + \epsilon_R + \tau_R\eta_n
\end{aligned}$$

It is clear from the equations above that the net effect of the characteristics $\boldsymbol{x_{(n,\alpha_n)}}$ on the utilities of the indicators is now composed of the model-specific parameters ($\boldsymbol{\beta}$, $\boldsymbol{\delta}$, $\boldsymbol{\theta}$) combined with the scaled effect of relationship strength ($\boldsymbol{\gamma}$). It is not possible to identify these different sets of parameters, as the measurable effect is given by the mixture of the two. Only the τ parameters can be correctly identified through the magnitude of the error terms η_n. For this reason, we fix the $\boldsymbol{\theta}$ parameters to zero, so that $\boldsymbol{\gamma}$ captures the specific effect of strength on the utility of the ranking, while the $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ parameters constitute shifts from these. Only by doing this, i.e. by using one of the models as a "base", can we identify the remaining parameters.

The overall likelihood of the model is equal to:

$$L_n = \int_n P_n(\alpha_n)\phi(\eta_n)d\eta_n, \tag{7}$$

where

$$\begin{aligned} P_n(\alpha_n) = & T_{t_n}(\sigma, \boldsymbol{\beta}, \tau_T, \alpha_n) \cdot F_{f_{n,f2f}}(\mu_{f2f}, \boldsymbol{\kappa}, \boldsymbol{\delta}, \tau_{f2f}, \alpha_n) \cdot F_{f_{n,call}}(\mu_{call}, \boldsymbol{\kappa}, \boldsymbol{\delta}, \tau_{call}, \alpha_n) \\ & \cdot F_{f_{n,text}}(\mu_{text}, \boldsymbol{\kappa}, \boldsymbol{\delta}, \tau_{text}, \alpha_n) \cdot F_{f_{n,email}}(\mu_{email}, \boldsymbol{\kappa}, \boldsymbol{\delta}, \tau_{email}, \alpha_n) \\ & \cdot F_{f_{n,OSN}}(\mu_{OSN}, \boldsymbol{\kappa}, \boldsymbol{\delta}, \tau_{OSN}, \alpha_n) \cdot R_n(\boldsymbol{\theta}, \tau_R, \alpha_n). \end{aligned} \tag{8}$$

3.5 Model specification

We started our analysis by estimating individual deterministic models for the components described above and then linking them via the LV. Our dataset includes a range of *ego* and *ego-alter* characteristics that have been tested as determinants of the choices studied in this paper. As described in Section 2, respondent socio-demographics were collected in the background survey, while *alter*-level information was gathered through the name generator.

The models described in Section 3 are estimated at the *ego-alter* level, so for each decision we estimate one model for each *ego-alter* pair in the data. As respondents in our sample name 1 to 40 *alters*, we estimate a minimum of 6 (as the ranking model is not estimated if only one name is listed in the name generator) and a maximum of 279 models per *ego*. As it would not have been possible to estimate models at the *ego* level[1], the coefficients only differ for each "*alter* type" (i.e. the nature of the relationship between the *ego* and each *alter*) instead of differing for each *alter* in the data. These "*alter* types" are friend, sibling, partner, parent, colleague, other relative and other acquaintance. A number of homophily measures were tested, such as whether *egos* and *alters* were of the same gender, belonged to the same age group, lived in Leeds or West Yorkshire or elsewhere in the UK as well as the distance between the residential locations of *ego-alter* pairs. We initially also tested a range of *ego*-specific characteristics, such as income,

[1]Estimating each model at the *ego* level, i.e. having for example a specific parameter for *ego* m, type of *alter* m and attribute p, would have implied a higher number of parameters than observations.

marital status and driving license, but these resulted in insignificant parameters in the joint model.

As stated in the Introduction, we want to test whether the factors affecting each of the three decisions examined are significantly different from those entering the equation of relationship strength (Equation 6). As an example, someone might have acquaintances who interact with them on online social networks, but who will not be strong contacts in an emotional way. In order to assess this, we started our investigation by estimating separate models for leisure time use, frequency of interaction by each mode and ranking in the name generator. We then compared the coefficients of these models, and when they were not significantly different from each other, we only included the respective $\boldsymbol{x_n}$ in the equation of the latent variable, i.e. measuring their impact via $\boldsymbol{\gamma}$. In the limited cases where we found the coefficients to be different for one or more separate models, we included the independent variables in the utility of the respective individual models. This approach aims at reducing the number of parameters and understanding if the concept of strength could be measured via only one of the individual choices.

4 Results

In this section, we present the results of our final linked model[2]

We estimated the parameters of the log-likelihood of the joint model in Equation 7 using the Apollo software (Hess and Palma, 2019) in R (R Core Team, 2016). We used 500 random draws to approximate Equation 7 using the MLHS sampling approach (Hess et al., 2006). The results are presented in Table 2 and 3, where only significant coefficients (as shown by the robust t-ratios) are retained, with a few exceptions that are discussed below.

Table 2 contains three sets of estimated coefficients. Following the notation described in Section 3, the β and δ coefficients measure the impact of homophily characteristics in the Tobit and ordered models, respectively. The corresponding coefficients entering the latent variable (γ) are reported in Table 3. As described in Section 3.5, we include the majority of effects in the latent variable and only include these for the individual

[2] While from a behavioural standpoint the joint model presented in this section is sufficient to highlight our findings, an anonymous reviewer suggested we add a note on the comparison across the estimated models. The table below shows the Log-likelihood and Bayes Information Criterion (BIC), calculated both on the basis of the number of respondents and the number of observations, for the three model types, their sum and the joint model presented in this section. As the table shows, when the BIC, correcting for the number of parameters, is computed, this highlights that the joint model is preferable to the sum of the three individual ones.

Individual models	N. parameters	Log-Likelihood	BIC (n. respondents)	BIC (n. observations)
Leisure time use	23	-2,013.15	4,166.45	4,246.53
Social interactions	62	-5,0045.58	100,468.96	100,684.84
Ranking in name generator	21	-8,150.11	16,428.19	16,501.31
Sum of individual models	106	-60,208.84	121,063.59	121,432.67
Joint model	66	-60,306.29	**121,014.76**	**121,244.56**

models when they are significantly different. In particular, only two such parameters were retained in the models for text and email interactions, as explained below.

As shown in Table 3, we find that female respondents tend to have a stronger relationship strength with their female siblings ($\gamma_{\text{both F, sibling}}$) and their mothers ($\gamma_{\text{both F, parent}}$). The positive sign of $\gamma_{\text{both F, friend}}$ indicates that female friends are also closer (with respect of friends of different gender), although the effect is not strongly significant. These results are in line with existing findings in the social network literature. In a cross-cultural study based on online social network data, David-Barrett et al. (2015) found that women tend to create more intimate dyadic relationships with other women, while men prefer larger cliques.

We also look at age homophily, and find that when *egos* are between 18 and 29 year old, they have stronger relationships with friends and acquaintances of the same age. The age group 30-39 displays a different pattern: we observe a lower strength (as opposed to that between pairs of different age) between friends, sibling and relatives. The pattern is similar for *egos* who are over 50. This could be due to the fact that those in the older group interact with younger members of their family and younger colleagues, as they will tend to be the most senior. The distance between the residential location of *egos* and *alter* also proved to be related to strength of the relationship: as expected, relationships between pairs who live further away prove to be weaker, with a particularly pronounced effect on other relatives.

Model	**Cofficient name**	**Estimate**	**Robust t-stat**
	β_0	-0.6455	-15.59
Time use	σ	0.4756	20.87
(Tobit model)	$\boldsymbol{\tau}$ **parameter**		
	τ_T	0.2177	17.92
	$threshold_{\mathrm{k11}}$	-0.2966	-3.75
	$threshold_{\mathrm{k12}}$	0.4870	6.48
	$threshold_{\mathrm{k13}}$	1.2586	15.04
	$threshold_{\mathrm{k14}}$	2.0651	22.12
	$threshold_{\mathrm{k15}}$	3.1045	30.63
	$threshold_{\mathrm{k16}}$	3.5038	31.28
	$threshold_{\mathrm{k21}}$	0.2922	3.50
	$threshold_{\mathrm{k22}}$	1.0963	13.12
	$threshold_{\mathrm{k23}}$	1.9528	20.01
	$threshold_{\mathrm{k24}}$	3.0952	26.34
	$threshold_{\mathrm{k25}}$	4.5310	27.15
	$threshold_{\mathrm{k26}}$	5.9680	28.37
	$threshold_{\mathrm{k31}}$	-0.8278	-9.02
	$threshold_{\mathrm{k32}}$	-0.0881	-1.13
	$threshold_{\mathrm{k33}}$	1.0520	13.04
	$threshold_{\mathrm{k34}}$	1.8264	20.33
	$threshold_{\mathrm{k35}}$	3.3598	29.67
	$threshold_{\mathrm{k36}}$	3.9653	29.71
	$threshold_{\mathrm{k41}}$	0.2424	2.82
	$threshold_{\mathrm{k42}}$	0.9328	10.69
	$threshold_{\mathrm{k43}}$	1.8916	22.02
	$threshold_{\mathrm{k44}}$	2.5038	25.37
Frequency of interaction	$threshold_{\mathrm{k45}}$	3.7005	24.50
(ordered logit models)	$threshold_{\mathrm{k46}}$	4.2004	19.51
	$threshold_{\mathrm{k51}}$	-1.3184	-13.83
	$threshold_{\mathrm{k52}}$	-0.5581	-7.51
	$threshold_{\mathrm{k53}}$	0.4137	5.27
	$threshold_{\mathrm{k54}}$	1.0365	11.39
	$threshold_{\mathrm{k55}}$	2.3227	18.95
	$threshold_{\mathrm{k56}}$	2.9125	20.65
	μ_{f2f}	5.2883	20.38
	μ_{call}	2.6494	32.50
	μ_{text}	2.8832	35.25
	μ_{email}	1.7814	32.21
	μ_{OSN}	2.1124	35.32
	$\delta_{\mathrm{email,\ both\ 18\text{-}29,\ friend}}$	-0.9147	-2.86
	$\delta_{\mathrm{text,\ both\ over\ 50,\ partner}}$	-1.6946	-4.18
	$\boldsymbol{\tau}$ **parameters**		
	τ_{f2f}	1.0090	17.46
	τ_{call}	0.9865	19.33
	τ_{text}	0.9810	21.90
	τ_{email}	0.4361	12.78
	τ_{OSN}	0.4014	10.27
Ranking	$\boldsymbol{\tau}$ **parameter**		
(exploded logit model)	τ_{R}	0.4222	12.12

Table 2: Model results

Model	Cofficient name	Estimate	Robust t-stat
Coefficients in latent variable	γ_{sibling}	1.1164	6.13
	γ_{partner}	4.6654	27.96
	$\gamma_{\text{colleague}}$	0.9086	4.54
	γ_{parent}	1.3053	9.96
	$\gamma_{\text{other relative}}$	0.7264	4.65
	$\gamma_{\text{both F, friend}}$	0.1161	1.09
	$\gamma_{\text{both F, parent}}$	0.6620	4.80
	$\gamma_{\text{both F, other relative}}$	0.0366	0.20
	$\gamma_{\text{both 18-29, friend}}$	0.9090	7.30
	$\gamma_{\text{both 18-29, other acquaintance}}$	1.9853	3.81
	$\gamma_{\text{both 30-49, friend}}$	-0.0437	-0.35
	$\gamma_{\text{both 30-49, sibling}}$	-0.5612	-2.48
	$\gamma_{\text{both 30-49, other relative}}$	-1.1616	-5.34
	$\gamma_{\text{both over50, sibling}}$	-0.9187	-3.26
	$\gamma_{\text{both over50, partner}}$	0.0127	0.06
	$\gamma_{\text{both over50, other relative}}$	-1.3809	-4.32
	$\gamma_{\text{dist, friend}}$	-0.0528	-4.00
	$\gamma_{\text{dist, sibling}}$	-0.0669	-2.32
	$\gamma_{\text{dist, parent}}$	-0.0659	-2.93
	$\gamma_{\text{dist, other relative}}$	-0.1293	-4.71

Table 3: coefficients in LV

At the top of Table 3 we show constants for each type of contact, using "friend" as a base. These constants represent the strength of relationship for each *alter*-type when not captured by subsequent shifts, so for example γ_{sibling} represents the strength between an *ego* and his/her sibling(s) when they are of different gender or both male and they are of different age or both aged between 18 and 29. For this reason, the magnitude of these coefficients is not immediately interpretable, but it gives the general idea that family members, partners and colleagues tend to have a stronger relationship with *egos* than friends.

The results just described are more readily understood by summing up the coefficients for each *alter*-type, so that the cumulative effect of different traits of the relationship between an *ego* and an *alter* can be directly quantified. Table 4 uses colours to highlight the overall strength of the relationship between *egos* and *alters* on the basis of the deterministic effects of age and gender homophily. A darker shade of green identifies a stronger relationship, while a darker shade of orange highlights a stronger negative strength, i.e. a lower strength than with the respective base category. General expectations are confirmed: the relationship between *egos* and their partners is the strongest irrespective of their age, followed by the relationship of female *egos* with their mothers and with acquaintances for younger people.

As expected, the parameters of the different individual models, prior to including the LVs, were generally in line with each other, confirming our hypothesis that these different behaviours have some underlying commonalities. This meant that most effects could be captured in the LV. As shown in Table 2, no Tobit-specific β parameters were needed aside from a constant β_0. The standard deviation σ of the Normal distribution is also estimated (cf. Equation 1). For the frequency of interaction models, we estimate a set of 6 thresholds for each of the 5 ordered models and a set of constants μ used in the binary logit models (interaction Vs. no interaction). The *threshold* parameters in the ordered models (6 per model, as the dependent variables are in 7 levels) are simply the utility cutoffs corresponding to these different levels (c.f. Section 7.4 of Train, 2009). The μ parameters are all positive and highly significant, implying that each pair in the sample is more likely to interact via each of the modes rather than not. The particularly high value of μ_{f2f} highlights the prominence of face-to-face interaction between most pairs. The only parameters that we found to have a significantly different value in one of the choice models and in the equation of the latent variable is the effect of *ego-alter* pairs being friends and both belonging to the age group 18-29. While $\gamma_{\text{both 18-29, friend}}$ is positive, indicating a higher strength between such pairs, $\delta_{\text{email, both 18-29,friend}}$ is negative, implying a less likely email communication between such pairs with respect to others. This result is in line with existing literature showing that younger people do not use email to communicate with their peers (Agosto et al., 2012; Calastri et al., 2017). In addition, while the coefficient relating the status of partners over 50 to strength ($\gamma_{\text{both over50, partner}}$) was not significant in this final model, the same effect is significant and negative in the text-interaction ordered model ($\delta_{\text{text, both over 50, partner}} = -1.6946$), showing that partners over 50 are less likely to communicate via text messages.

For all the models, we estimate τ parameters, measuring the impact of the LV on the utilities of the different choices. As clearly shown in Table 2, these are all strongly significant, confirming the presence of an underlying construct linking these decisions, that we interpret as relationship strength. A stronger relationship therefore implies a higher share of leisure time spent together, more likely communication by each mode (although we see a stronger effect on face-to-face, call and text) and a higher ranking in the name generator, suggesting that stronger contacts are the first to come to mind.

5 Conclusions

An increasing number of scholars in travel behaviour research have advocated the importance of better understanding patterns of leisure travel and time allocated to leisure and social activities, given their prominent role in people's lives and their implications in terms of demand for transport and infrastructure (Aguiar and Hurst, 2007; Schlich et al., 2004; Tilahun and Levinson, 2017)

	Friend	Parent	Sibling	Colleague	Partner	Other relative	Other acqu.
	Both female		Both male or different gender				
same age (18-29)	1.03	1.97	1.12	0.91	4.67	0.73	1.99
same age (30-49)	0.07	1.97	0.56	0.91	4.67	-0.44	0
same age (over 50)	0.12	1.97	0.20	0.91	4.67	-0.65	0
different age	0	0	0	0	0	0	0

Table 4: Deterministic effect of age and gender homophily on strength

A number of studies in this area have adopted a social network perspective, in line with the recognition of the fact that the people involved are the main drivers of such activities and travel (Carrasco and Miller, 2009; van den Berg et al., 2013). The present study places itself within that limited but growing literature, and provides some important contributions to the understanding of the behavioural processes at play. Differently from most of the literature, we model different choices jointly. This is particularly important in the present field of analysis, where processes such as interactions and time spent together are clearly related, and looking at these processes independently can mask important interconnections.

Moreover, this study aims to model strength of social relationships, a concept that is potentially pivotal in explaining travel and social activities. We obtain results in line with expectations, underlying a higher strength between partners and mothers and daughters.

Differently from previous research where a stated measure of closeness was provided by respondents, we treat strength as an unobserved latent factor, as suggested in the sociology literature, which explains a number of indicators. This is in line with the notion, emerging from work on this topic (e.g. Marsden and Campbell, 1984), that strength cannot be observed or measured, only its impact can be captured. While two of the indicators used, frequency of interaction and time spent, have been explored in the sociology literature, we propose the use of the ranking in the name generator as an additional indicator, and successfully demonstrate its strong correlation with the other two aspects. This has specific relevance as it offers future analysts grounds to make use of the ranking in the name generator as a proxy for relationship strength, a measure that is collected via simply completing a name generator and which does not pose additional burden on respondents.

Our work, like similar studies in this area, aims at gaining further insights on the dynamics of such decisions, with a view to being able to incorporate relevant aspects in activity based models and the ultimate goal to produce better insights and forecasts of travel behaviour. So while this study may not seem to directly shed light on specific travel choices, it hopes to constitute a building block for future more specific analyses and offer a better understanding of the underlying processes.

As is generally the case, our study could be improved in different ways. Research has shown that the social network approach provides important insights to explain leisure and social activities, but reliable approaches to collect social network data are still lacking, as name generators are known to be subject to recall and other biases. In addition, given the overall length of our survey, the name interpreter was somewhat limited. Future studies with a precise focus on leisure travel from the social network perspective could aim to investigate the impact of other homophily measures, such as similarity in level of education and occupation, which are known in the literature to correlate with interactions. Moreover, over time, further indicators of strength have been suggested in the sociology literature, and it would be interesting to incorporate these in a quantitative model, albeit that this would lead to increased model complexity.

References

Agosto, D. E., Abbas, J., Naughton, R., 2012. Relationships and social rules: Teens' social network and other ict selection practices. Journal of the American Society for Information Science and Technology 63 (6), 1108–1124.

Aguiar, M., Hurst, E., 2007. Measuring trends in leisure: the allocation of time over five decades. The Quarterly Journal of Economics 122 (3), 969–1006.

Amemiya, T., 1984. Tobit models: A survey. Journal of econometrics 24 (1-2), 3–61.

Calastri, C., dit Sourd, R. C., Hess, S., 2018a. We want it all: experiences from a survey seeking to capture social network structures, lifetime events and short-term travel and activity planning. Transportation, 1–27.

Calastri, C., Hess, S., Daly, A., Carrasco, J. A., Choudhury, C., 2018b. Modelling the loss and retention of contacts in social networks: The role of dyad-level heterogeneity and tie strength. Journal of choice modelling 29, 63–77.

Calastri, C., Hess, S., Daly, A., Maness, M., Kowald, M., Axhausen, K., 2017. Modelling contact mode and frequency of interactions with social network members using the multiple discrete–continuous extreme value model. Transportation Research Part C: Emerging Technologies 76, 16–34.

Carrasco, J.-A., Miller, E. J., 2009. The social dimension in action: A multilevel, personal networks model of social activity frequency between individuals. Transportation Research Part A: Policy and Practice 43 (1), 90–104.

David-Barrett, T., Rotkirch, A., Carney, J., Izquierdo, I. B., Krems, J. A., Townley, D., McDaniell, E., Byrne-Smith, A., Dunbar, R. I., 2015. Women favour dyadic relationships, but men prefer clubs: cross-cultural evidence from social networking. PloS one 10 (3), e0118329.

De Cannière, M. H., De Pelsmacker, P., Geuens, M., 2010. Relationship quality and purchase intention and behavior: The moderating impact of relationship strength. Journal of Business and Psychology 25 (1), 87–98.

Department for Transport, 2018. Why people travel.
URL www.gov.uk/government/statistical-data-sets/nts04-purpose-of-trips

Frei, A., 2012. Networks, geographies and travel: Travel between infrastructure and social structure. Ph.D. thesis, ETH, Eidgenössische Technische Hochschule Zürich.

Friedkin, N., 1980. A test of structural features of granovetter's strength of weak ties theory. Social networks 2 (4), 411–422.

Gilbert, E., Karahalios, K., Sandvig, C., 2008. The network in the garden: An empirical analysis of social media in rural life. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '08. pp. 1603–1612.

Gimenez-Nadal, J. I., Sevilla, A., 2012. Trends in time allocation: A cross-country analysis. European Economic Review 56 (6), 1338–1359.

Granovetter, M., 2018. Getting a job: A study of contacts and careers. University of Chicago press.

Granovetter, M. S., 1977. The strength of weak ties. In: Social networks. Elsevier, pp. 347–367.

Hess, S., Palma, D., 2019. apollo: Tools for Choice Model Estimation and Application. R package version 0.0.6.
URL https://CRAN.R-project.org/package=apollo

Hess, S., Train, K. E., Polak, J. W., 2006. On the use of a modified latin hypercube sampling (mlhs) method in the estimation of a mixed logit model for vehicle choice. Transportation Research Part B: Methodological 40 (2), 147–163.

Jones, J. J., Settle, J. E., Bond, R. M., Fariss, C. J., Marlow, C., Fowler, J. H., 2013. Inferring tie strength from online directed behavior. PloS one 8 (1), e52168.

Kim, J., Rasouli, S., Timmermans, H. J., 2017. Investigating heterogeneity in social influence by social distance in car-sharing decisions under uncertainty: A regret-minimizing hybrid choice model framework based on sequential stated adaptation experiments. Transportation Research Part C: Emerging Technologies 85, 47–63.

Kowald, M., 2013. Focussing on leisure travel: The link between spatial mobility, leisure acquaintances and social interactions. Ph.D. thesis, ETH, Eidgenössische Technische Hochschule Zürich.

Kroesen, M., Chorus, C., 2018. The role of general and specific attitudes in predicting travel behavior–a fatal dilemma? Travel behaviour and society 10, 33–41.

Maness, M., 2017. A theory of strong ties, weak ties, and activity behavior: Leisure activity variety and frequency. Transportation Research Record 2665 (1), 30–39.

Marsden, P. V., Campbell, K. E., 12 1984. Measuring Tie Strength. Social Forces 63 (2), 482–501.

Mokhtarian, P. L., Salomon, I., Handy, S. L., 2006. The impacts of ict on leisure activities and travel: a conceptual exploration. Transportation 33 (3), 263–289.

Office for national statistics, 2015. Leisure time in the uk: 2015.
URL www.ons.gov.uk/economy/nationalaccounts/satelliteaccounts/articles/leisuretimeintheuk/2015

Ohnmacht, T., Götz, K., Schad, H., 2009. Leisure mobility styles in swiss conurbations: construction and empirical analysis. Transportation 36 (2), 243–265.

R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL https://www.R-project.org/

Resource Systems Group, R., 2017. rmove. http://rmove.rsginc.com/index.html.

Schaefer, C., Coyne, J. C., Lazarus, R. S., 1981. The health-related functions of social support. Journal of behavioral medicine 4 (4), 381–406.

Schlich, R., Schönfelder, S., Hanson, S., Axhausen, K. W., 2004. Structures of leisure travel: temporal and spatial variability. Transport Reviews 24 (2), 219–237.

Stanko, M. A., Bonner, J. M., Calantone, R. J., 2007. Building commitment in buyer–seller relationships: A tie strength perspective. Industrial Marketing Management 36 (8), 1094–1103.

Stauffacher, M., Schlich, R., Axhausen, K. W., Scholz, R. W., 2005. The diversity of travel behaviour: motives and social interactions in leisure time activities. Arbeitsberichte Verkehr-und Raumplanung 328.

Tilahun, N., Levinson, D., 2017. Contacts and meetings: Location, duration and distance traveled. Travel Behaviour and Society 6, 64 – 74.
URL http://www.sciencedirect.com/science/article/pii/S2214367X16300102

Train, K. E., 2009. Discrete choice methods with simulation. Cambridge university press.

van den Berg, P., Arentze, T., Timmermans, H., 2013. A path analysis of social networks, telecommunication and social activity–travel patterns. Transportation Research Part C: Emerging Technologies 26, 256–268.