Modelling the loss and retention of contacts in social networks: the role of tie strength and dyad-level heterogeneity

Chiara Calastri*† Stephane Hess*‡ Andrew Daly*§
Juan Antonio Carrasco¶ Charisma Choudhury*‖

February 9, 2018

Abstract

Social networks have attracted attention in different fields of research in recent years and choice modellers have engaged with the discipline by looking at the role that social networks play in shaping decisions across a variety of contexts. The incorporation of the social dimension in choice models creates the need for understanding how social networks evolve over time and in particular which social contacts (alters) are retained over time by an individual (ego). Existing work fails to capture the full extent of ego-level and ego-alter level heterogeneity in these processes. We propose the use of a hybrid model framework which is based on the notion of latent strength of relationship. The resulting model allows for heterogeneity in the latent strength both across individuals and across their different relationships. In addition, we allow for heterogeneity not linked to the latent strength concept. We demonstrate the benefits of the approach using data from Chile, showing the presence of extensive variations in retention of social contacts and in strength of relationship both at the ego and ego-alter level, only some of which can be linked to observed characteristics.

Keywords: hybrid choice models; social networks; random taste heterogeneity; intra-respondent

1 Introduction

Social networks have attracted substantial attention across different research fields in recent years, looking for example at information diffusion (e.g. Bakshy et al., 2012) and

*Institute for Transport Studies and Choice Modelling Centre, University of Leeds (UK)
†tscca@leeds.ac.uk
‡s.hess@its.leeds.ac.uk
§andrew@alogit.com
¶Universidad de Concepción (Chile), j.carrasco@udec.cl
‖c.f.choudhury@leeds.ac.uk

social influence (e.g. Kempe et al., 2003). Choice modellers have engaged with this discipline by looking at the potential effects that social networks may have on decisions across a variety of contexts, including time use (Calastri et al., 2017a), telecommuting (Páez and Scott, 2007) and evacuation stategies (Sadri et al., 2017) and by modelling decisions related to social interactions (Calastri et al., 2017b). Several contributions focused on addressing the issue of how to capture social influence (Dugundji and Walker, 2005; Maness et al., 2015), with some also dealing with the issue of endogeneity that might be implied by including such effects (Walker et al., 2011).

A key characteristic of social networks is that they are not static but evolve over time. Some first attempts to model social network dynamics and their interaction with life-cycle events have been made by (Chávez et al., 2017; Sharmeen et al., 2014, 2015, 2016). While these papers have modelled the changes in social networks over time, more can be done to accommodate the extent of the heterogeneity involved in this process at the respondent and at the dyad level. Indeed, there are many possible reasons why a relationship might be maintained or lost. These have to do with circumstances related to a specific subject's and his/her social contacts' lives, and the number of these effects that we can control is rather limited. For this reason, we need to try to capture some of the heterogeneity across people and at the dyad level through random effects, as we want to test the hypothesis that people are different in what we do not observe.

Researchers in the social sciences have studied characteristics and processes inherent to social networks themselves, i.e. how they are formed and how they can be represented. The process of network formation and change over time is complex and depends on the characteristics of the different individuals involved, and its study requires adequate data. In particular, when the aim of the study is to investigate network changes over time, longitudinal data, inclusive of information about individuals and their attributes, are needed. Such data are rare and most examples in the social network literature are for mainly for small groups (e.g. Wasserman and Faust, 1994), while studies using larger groups have mainly focused on cross-sectional analyses.

Given the generally limited sample sizes, qualitative methods have often been applied to investigate the determinants of social network evolution. Interviewing a sample of 33 people from Toronto in 1968 and then in 1978, Wellman et al. (1997) analysed the change in personal networks, finding that frequent telephone interactions and social support increase the likelihood of retaining contacts over time, while changes in marital status (especially for women) may involve losing friends. A strong turnover in the network is reported, except for a stable core component. The latter finding is confirmed by Mollenhorst et al. (2014), who study the changes occurred over seven years in the social network of Dutch people aged 18 to 65 (although they use a larger sample). High average numbers of social contacts lost over time are reported in particular in association with important life course events. This is especially apparent in a study surveying young French people (Degenne and Lebeaux, 2005).

The social network analysis literature proposes different approaches to analyse social network changes over time. The stochastic actor-based models (e.g. Snijders et al., 2010) is based on the idea that when an actor (or ego) has the opportunity to make a change, he/she will select the new social composition from a given set of possible states of the network. The probability to select a given state x takes a form that is similar to a multinomial logit model, with the element corresponding to the utility function (called "objective function") defined as a linear combination of components called effects. The latter mainly refer to network characteristics, such as reciprocity and transitivity, rather than on sociodemographic and life-course information. Of course, both aspects could be incorporated for a more complete behavioural representation of the process. Moreover, a requisite for this technique is the use of complete networks. Exponential random graph models (ERGM), used to analyse the structural features of social networks, have been extended to incorporate their temporal dynamics (e.g. Hanneke et al., 2010). A subset of these models is represented by the Dynamic network logistic regression (e.g. Almquist and Butts, 2013), a model which is essentially the same as a binary logit model.

While the nomenclature is different, the underlying mathematical structures used are very similar to basic choice models. While the social network analysis literature has applied these methods only in a relatively basic way, it provided a theoretical support for their suitability for the analysis of networks. The choice modelling literature provides much more advanced tools, able to represent more complex real-life behaviour, making use of ego-centric data on top of entire network data. . This in part motivates our work in the present paper.

For these reasons, choice models represent a suitable tool for the study of how social networks evolve, and in particular to explain whether a given social contact is maintained over time. Indeed, the outcome of maintaining or removing a social connection from the network is a discrete outcome associated with certain variables. Mathematically, that type of outcome is correctly represented with such a model. From a behavioural perspective, this outcome can in some occasions be seen as a choice (e.g. if an ego decides to stop seeing an alter, or vice-versa) or as a result of a non-choice process (e.g. if an alter dies or is forgotten in the name generator). Discrete choice models can also accommodate latent constructs influencing the outcome, as we will do in this paper to measure the impact of relationship strength on the probability of retaining social contacts. Moreover, such models allow to acknowledge the presence of random heterogeneity across individuals and outcomes. Whether or not the actual process is a choice has little bearing on the mathematical suitability of the models.

As with many other decisions, there is scope for extensive heterogeneity, both deterministic and random. Crucially, this is an area where the heterogeneity may be especially strong at the individual level, so that the likelihood of retaining people in one's network varies extensively across individuals, but there is further (and possibly even larger) heterogeneity across the individual members of a network. This is line with the work on network evolution which often refers to "core" ties, identified as the ones who are emotionally closer to the surveyed individual, and which are also the ones that are generally more likely to be retained over time. These are often identified as the closest friends or family members, or through questions asking for the names of the people who provide more support. This creates the notion of relationship strength, which has been the object of studies linking it to frequency of interaction, amount of time spent together and reciprocal services (for a discussion, see Marsden and Campbell 1984).

In the present paper, we put forward the idea of modelling this notion of strength of a relationship as a latent component within a hybrid choice model. We show how this allows us to separately account for different layers of heterogeneity, both at the level of individual survey respondents (which we call egos) and across the different members of their network (alters). The incorporation of these different levels of heterogeneity, together with the application of advanced choice models to social network dynamics, represents the main contribution of this paper. We demonstrate this approach using data from a typical name generator survey conducted in Chile. Our findings can be operationalized to dynamically predict the composition of the social network as well as the strength of the ties, both of which can lead to more realistic modelling of activity, travel and other choices.

This paper is organised as follows. Section 2 introduces the data collection protocol and describes the sample used for analysis. This is followed by a discussion of the modelling methodology (Section 3) and the results (Section 4). Finally, we draw conclusions about the modelling work performed in the paper.

2 Data

2.1 Survey and data collection

For the present analysis, we use longitudinal network data collected in two waves (2008 and 2012) within the Communities in Concepción project, in the city of Concepción, Chile. The project, aimed at understanding multiple aspects of the life of the respondents, included a very rich questionnaire, a name generator and name interpreter and a time use diary.

The questionnaire, which presented only minor differences in the two waves, asked respondents to provide information about themselves and their household, their housing arrangements as well as their past education and job history.

We make use of multiple parts of the questionnaire, although the crucial elements for our study are the *name generator* and the *name interpreter*. A name generator is a survey question, usually in the form of a table, asking respondent to list the names of the people in their social network (Campbell and Lee, 1991; Marsden, 1990). Different studies use different types of stimuli to help respondents recall the relevant social network. For example, some studies are more interested in business networks, while others ask people to recall the names of those with whom they spend their free time. In the case of the present study, the instrument is based on Carrasco et al. (2008b). Two different name generators were presented to respondents. Both asked them to report names of people outside of their household (could be friends, family members, neighbours etc.), dividing those who are emotionally very close from those who are "somewhat close", although not mere acquaintances. On top of providing the names of the alters, egos were asked to enrich this list by answering questions about them: some basic socio-demographic characteristics (sex, age, occupation, residential location) as well as some information concerning the relationship between the two are asked. In particular, egos are asked to specify for how for long they had known each alter (less than a year, 1-10 years or more than 10 years) and how they would define their relationship (immediate family, other family, neighbour, friend, colleague, someone from a club or organisation). Egos were also asked to report the frequency of interaction by different modes of communication with each alter. As in most social network surveys, respondents were given a separate table to report these additional information about each alter, referred to as the name interpreter.

The name generator was not the only part of the survey related to the social environment. A social capital section listed different "types of help" (e.g. advice on important matters, borrowing small amounts of money, assisting when ill), and for each of them asked the egos to identify one or more alters to whom/from whom they could grant/receive this type of help. Several Likert-scale questions about the ego's personality traits and subjective well-being were also included.

Finally, respondents are asked to fill in a time use diary for two days, a week day and a week-end day. Start and end times as well as activity type, place and people involved had to be specified.

The data were collected by an interviewer at the respondents' homes, except for the time diary, which respondents were given instructions about and let free to complete on the chosen days.

2.2 Sample characteristics

Participants were recruited by post using their home address. This implies that a certain number of respondents from the original (2008) sample of 240 were lost due to relocation or unresponsiveness. 102 people took part in both waves. Due to the specific nature of our study, we excluded from this sample the participants who did not complete the name generator in either wave or who did not have any overlap in their network, as we assumed that this was due to severe recall issues. The usable sample for analysis was made up of 94 respondents, who named a total of 1912 alters in 2008 (20.34 each, on average).

Table 1 reports the socio-demographic characteristics of the egos as of 2008 (first wave) as well as some statistics about the life course events occurred between the two waves and network size statistics. For each level of the different statistics, we also report the number of alters recorded by egos with the given characteristics. Although the selection of the subsample for this analysis somewhat limited its representativeness, the originally collected sample matched the characteristics of the local population well (Carrasco and Cid-Aguayo, 2012).

The histogram produced using the social network size data (see 1) shows peaks around 15 and 16 contacts. It is also clear that only a handful of people name a particularly high number of connections, with only 1 person each naming 37, 43 and 44 contacts.

Table 1: Ego characteristics

	Number of egos		**Number of alters**	
	94		**1912**	
	Freq	**%**	**Freq**	**%**
Sex				
Male	33	35%	632	33%
Age				
18-30	25	27%	557	29%
31-40	16	17%	347	18%
41-50	18	19%	357	19%
51-60	17	18%	336	18%
over 61	18	19%	315	16%
Education				
Elementary School	15	16%	303	16%
Medium School	37	39%	752	39%
Technical School	8	9%	167	9%
Undergraduate Degree	25	27%	476	25%
Postgraduate Degree	9	10%	214	11%
Employment status				
Employed	49	52%	1037	54%
Unemployed	9	10%	191	10%
Student	6	6%	127	7%
Homemaker	17	18%	329	17%
Retired	8	9%	140	7%
other	5	5%	88	5%
Mobility				
Driving license	31	33%	625	33%
Communication tools ownership				
Landline	61	65%	1213	63%
Mobile phone	80	85%	1685	88%
Internet connection	51	54%	1074	56%
Life course events 2008-2012				
Went to University	6	6%	124	6%
Finished University	2	2%	39	2%
Started a new job	4	4%	70	4%
Quit a job	4	4%	68	4%
Divorced	3	3%	43	2%
Got married	12	13%	253	13%
Had a child	14	15%	268	14%
Network size				
Min	7	-	-	-
Max	44	-	-	-
Average	20.34	-	-	-
Median	19	-	-	-

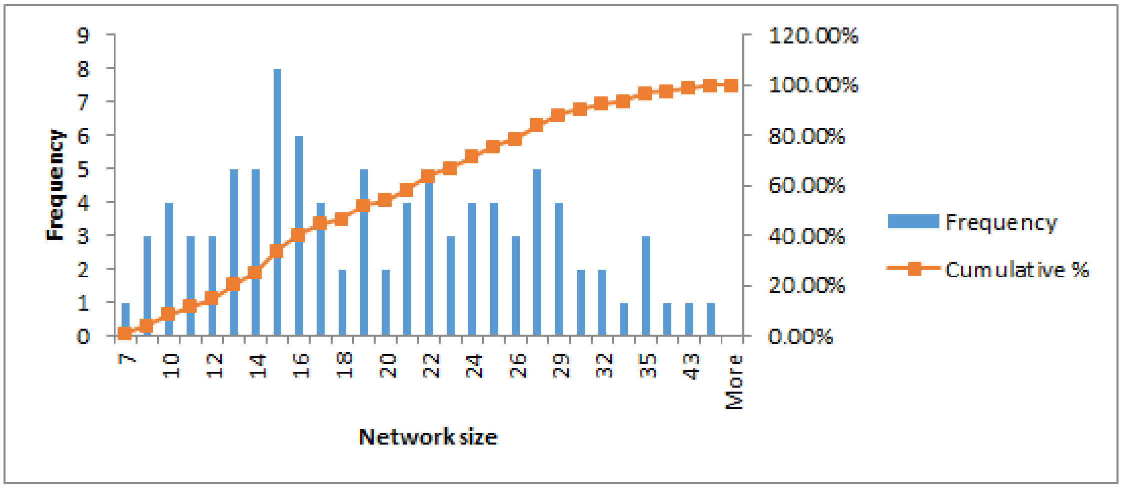

Figure 1: Social network size histogram

As it can be observed from Table 1, while in the case of the egos we can make use of both waves of data and infer life-course changes that could be relevant for social network-related behaviour, the same cannot be done when it comes to the ego-alter level variables. In fact, while we are aware of socio-demographic changes about the alters who have been retained in 2012, we have no information about those who are no longer part of the egos' network. This is indeed a potential limitation of the present work that we will come back to later on in the paper, as the outcome that we are focusing on is likely to depend not only on characteristics and events related to the ego's life, but also to those related to the alters. It is also worth acknowledging that some characteristics apply only to a few egos but are still used later on in our models. This is the case for example for egos going to university of getting divorced. However, the modelling is clearly performed at the ego-alter level, and this is what enables the estimation of the parameters, where for example in the case of ego-student we have 127 observations, while for ego-divorced we have 43.

Table 2 reports ego-alter characteristics. We can observe that for *Sex Homophily*, *Age Homophily* and *Time known each other*, the categories are of course mutually exclusive, while egos could classify a given alter as (say) both a colleague and a friend. This happens in a limited number of cases (230 out of 1912), not sufficient to create interaction effects in our models.

3 Modelling methodology

We observe E separate egos that participate in both the 2008 and 2012 survey, where for a given ego e, we have a set A_e of different alters named in 2008. The objective of

Table 2: Ego-alter characteristics

Number of ego-alter pairs	**1912**	
	Freq.	%
Sex homophily		
both male	388	20%
both female	802	42%
different sex	722	38%
Age homophily		
Under 30	390	20%
30-60	135	7%
Over 60	161	8%
Different age	1222	64%
Type of relationship		
Immediate family	369	19%
Other family	460	24%
Neighbour	352	18%
Colleague	171	9%
Club/ Organisation	107	6%
Friend	690	36%
Time known each other		
Less than 1 year	173	9%
1-10 years	1081	57%
More than 10 year	397	21%
NA	261	14%

our modelling work is to explain the retention or loss of a given alter by a given ego[1]. For this, we specify y_{a_e} to be the dependent variable of a binary model, which takes the value 1 if and only if ego e retains alter a_e in his/her network of named social contacts.

3.1 Binary models for retention of social contacts

As a first step, we model the retention of an alter as a function of the characteristics of the ego and alter in 2008 ($z_{e,a_e,2008}$) and any changes in the characteristics of the ego between 2008 and 2012 ($z_{e,2008-2012}$). Using a simple binary logit model, we would then write the utility of retention as:

$$U_{e,a_e} = V_{e,a_e} + \varepsilon_{e,a_e} = \delta + f\left(\beta, z_{e,a_e,2008}, z_{e,2008-2012}\right) + \varepsilon_{e,a_e} \quad (1)$$

[1] It is important to note that there are potential recall issues associated with the use of name generators, so although we attempt to model loss, what we can model, given the potential measurement error, is recall.

where δ is an estimated constant, β is a vector of parameters measuring the impact of $z_{e,a_e,2008}$ and $z_{e,2008-2012}$ on V_{e,a_e} and ε_{e,a_e} is an *i.i.d.* extreme value error term. The probability of the observed outcome would then be:

$$P_{y_{a_e}}(\delta,\beta)=\frac{e^{y_{a_e}\cdot V_{e,a_e}}}{1+e^{V_{e,a_e}}}, \tag{2}$$

and the likelihood of the sequence of outcomes for ego e would be given by:

$$L_e(\delta,\beta)=\prod_{a_e\in A_e}P_{y_{a_e}}(\delta,\beta), \tag{3}$$

where $\prod_{a_e\in A_e}$ is a product over all alters named by ego e in 2008.

The simple base model accounts for some of the differences across egos and across ego-alter pairs in the probability of retention by linking this to observed characteristics. However, there is clearly scope for additional unexplained variation both at the level of an individual ego (affecting his/her probability of retention equally across all alters) as well as the ego-alter level. This latter component of random heterogeneity is potentially especially important given that the retention in a social network is driven not just by the ego but also by the alter, where an added source for this is the lack of data on changes in alters' characteristics/circumstances between 2008 and 2012 (given that this is of course only available for retained alters).

Random heterogeneity at the ego level is relatively easy to deal with by making δ_e ego-specific in Equation 1, and allowing it to be distributed randomly across egos with $\delta_e\sim h(\delta_e\mid\mu_\delta,\sigma_\delta)$, using for example a Normal distribution. Equation 3 then becomes a binary mixed logit model, with:

$$L_e(\mu_\delta,\sigma_\delta,\beta)=\int_{\delta_e}L_e(\delta_e,\beta)\,h(\delta_e\mid\mu_\delta,\sigma_\delta)\,\mathrm{d}\delta_e, \tag{4}$$

where the integration is carried out at the level of an ego, recognising that we are dealing with ego level heterogeneity.

The binary mixed logit model in Equation 4 will capture differences in the probability of retention across the individual egos in the estimation sample. Such differences are likely to exist both as a result of unobserved characteristics of the ego as well as unobserved characteristics in his/her circumstances. It is also likely that some egos are more likely to name the same alters in 2012 and 2008 than others, i.e. some are more prone to forgetting some alters than others.

While ego-level heterogeneity obviously plays a role in the retention of social contacts, there is also substantial scope for ego-alter level heterogeneity - to put it colloquially *"it takes two to tango"*. Some of this heterogeneity can again be linked to the characteristics of the alter in 2008 and the differences/similarities between the ego and alter characteristics in 2008. However, there is substantial scope for additional unobserved heterogeneity. Such observation level random heterogeneity has received growing interest in choice modelling in recent years, and is typically referred to as inter-agent and intra-agent heterogeneity (Hess and Rose, 2009; Hess and Train, 2011). In our case, we refer to it as ego-level and ego-alter level heterogeneity. In the simple binary case faced here, any random heterogeneity in δ at the ego-alter level would be very difficult (or impossible) to disentangle from the extreme value error term in Equation 1[2]. This forms the motivation for the remainder of our methodological discussion in Sections 3.2.

3.2 Hybrid choice model with two layers of heterogeneity

In this section, we delve deeper into the possible reasons for retention. Specifically, we hypothesise that a key driver in the retention of an alter a_e in ego e's network is the strength of that relationship, and we show how incorporating this notion into our models allows us to have heterogeneity also at the ego-alter level.

3.2.1 Concept of latent relationship strength

We define α_{a_e} to be a latent variable which describes the strength of the relationship between ego e and alter a_e. We hypothesise that this latent strength depends on the characteristics of the ego and alter in 2008, i.e. $z_{e,a_e,2008}$. We exclude $z_{e,2008-2012}$ from the structural equation of this latent variable as it is used to explain the strength of the relationship in 2008. We then write:

$$\alpha_{a_e} = g\left(\gamma, z_{e,a_e,2008}\right) + \eta_e + \eta_{a_e} \tag{5}$$

where γ takes a role similar to β in Equation 1. This structural equation for the latent strength includes two random error terms, one distributed across egos (η_e) and one distributed across alters and egos (η_{a_e}). Both are specified to follow Normal distributions with a mean of 0, where, for normalisation, we set σ_{a_e}=1 and estimate σ_e. The rationale for two separate error terms is that we assume that the strength of a relationship varies both across egos and across alters for that ego. This means that we allow for the fact that some egos will be more prone to establishing strong relationships than others (variation across egos) and that even within a specific ego's network, certain ties will be stronger than others (variation across alters, for an ego).

3.2.2 Measurement models of latent relationship strength

The latent strength concept is explored in a number of measurement model components which use this variable to explain an ego's answers to a number of subjective questions, which we treat as indicators of the latent variable. This set of indicators I_e for ego e includes seven binary responses and one ordered response. The binary items were from survey questions where the ego had to specify the names of the alters he/she was exchanging support with, in particular: giving advice on important matters and about job opportunities, lending and borrowing small amounts of money, receiving help in terms

[2]Some weak empirical identification would only be possible thanks to differences between the distribution used for δ (say a Normal) and the extreme value distribution.

of mobility in times of need and whether or not the ego and alter participate in joint social activities; in addition, each alter could be reported in a different name generator depending on how emotionally close he/she was perceived: we use this as a measure of closeness. The ordered response related to the frequency of face-to-face contact. The selection of these statements as indicators of the latent variable was guided by studies in the field of sociology. Granovetter (1973) defines the strength of a tie as a "(probably linear) combination of the amount of time, the intimacy (mutual confiding) and the reciprocal services (...)". Some studies talk about "multiplexity", i.e. argue that the co-presence of multiple elements constitutes a strong tie (e.g Kapferer, 1969), while others consider the possibility that ties with only one content or with diffuse content might also be strong (Simmel, 1950). In our case, we did not have any measure of intimacy, but we did have measures of reciprocal services (the so-called social capital or support questions). We decided to also use the involvement in social activities as a measure of time voluntarily spent together.

Given these different and somewhat contradictory definitions, we acknowledge that the latent construct could be related to a social structure other than relationship strength, possibly linked to reciprocity. Nevertheless, the presence of the latent variable is important in our paper, and we believe that its interpretation as relationship strength is likely to be quite accurate. Whatever the source of the latent process, the mathematical framework is suitable for this analysis.

We use binary logit models to explain the answers to the first seven items, and an ordered logit model for the frequency of interaction. Each time, we estimate a parameter ζ_i that measures the impact of the latent variable on the indicator, along with a mean parameter in the binary logit model and four threshold parameters for the ordered logit model (for the 5-level indicator).

Under the error assumptions made in Equation 5, the probability of the observed set of answers to these questions for a given ego e across all his/her alters is then given by:

$$LI_e(\gamma, \sigma_e, \mu_I, \zeta_I) = \int_{\eta_e} \prod_{a_e \in A_e} \int_{\eta_{a_e}} \prod_{i=1}^{8} PI_{a_e,i}(\alpha_{a_e}, \mu_I, \zeta_I)\, \phi(\eta_{a_e})\, \mathrm{d}\eta_{a_e} \phi(\eta_e)\, \mathrm{d}\eta_e, \qquad (6)$$

where $\phi(\eta_{a_e})$ and $\phi(\eta_e)$ are Normal density functions with mean 0 and where, as mentioned above, the variance of η_{a_e} is normalised to 1. The probability $PI_{a_e,i}(\alpha_{a_e})$ for the observed response for indicator i for ego-alter pair a_e is conditional on the latent strength, with the functional form for this probability being either binary logit or ordered logit. This leads to the estimation of two vectors of parameters for the measurement models (μ_I and ζ_I) in addition to the parameters of the structural equation for α_{a_e}, namely γ and σ_e.

Equation 6 involves integration at two separate levels, leading to an ability to separate out the two layers of heterogeneity (ego and ego-alter level), albeit at a high computational cost (Hess and Train, 2011). The estimate of σ_e relative to the normalised σ_{a_e} gives an indication of the relative importance of the two layers of heterogeneity.

Two simplifications of this model arise. In the first, we allow for only ego-level random heterogeneity (but still deterministic ego-alter level heterogeneity through γ), meaning that we set $\eta_{a_e} = 0$ in Equation 5 and then normalise σ_e to 1. We then rewrite Equation 6 as:

$$LI_e\left(\gamma, \mu_I, \zeta_I\right) = \int_{\eta_e} \prod_{a_e \in A_e} \prod_{i=1}^{8} PI_{a_e,i}\left(\alpha_{a_e}, \mu_I, \zeta_I\right) \phi\left(\eta_e\right) \mathrm{d}\eta_e. \tag{7}$$

While this model is substantially easier to estimate, it led to slightly poorer performance, confirming the presence of extensive levels of random ego-alter heterogeneity.

It is similarly possible to estimate a version with only ego-alter level heterogeneity, thus setting $\eta_e = 0$ in Equation 5 and rewriting Equation 6 as:

$$LI_e\left(\gamma, \mu_I, \zeta_I\right) = \prod_{a_e \in A_e} \int_{\eta_{a_e}} \prod_{i=1}^{8} PI_{a_e,i}\left(\alpha_{a_e}, \mu_I, \zeta_I\right) \phi\left(\eta_{a_e}\right) \mathrm{d}\eta_{a_e}. \tag{8}$$

This model now has random heterogeneity only at the ego-alter level, but the estimation of this (in contrast with the binary choice model) is made possible by the presence of multiple indicators at the ego-alter level, just as in Equation 6.

3.2.3 Introduction of latent strength in binary choice model

We now exploit the concept of latent strength introduced in Section 3.2.1 to allow for ego-alter level heterogeneity in the retention model. We do this by jointly estimating the binary choice model and the various measurement models, allowing for a joint influence on both by the latent strength variable.

We first rewrite the utility in the choice model as:

$$U_{e,a_e} = V_{e,a_e} + \varepsilon_{e,a_e} = \delta_e + f\left(\beta, z_{e,a_e,2008}, z_{e,2008-2012}\right) + \tau\alpha_{a_e} + \varepsilon_{e,a_e} \tag{9}$$

This utility specification retains the randomly distributed δ_e from Section 3.1 but in addition allows for an impact of the latent strength variable through the estimation of τ.

The joint likelihood of the observed retention patterns and answers to the indicator questions for ego e with this model is now given by:

$$\begin{aligned} &LJ_e\left(\mu_\delta, \sigma_\delta, \beta, \tau, \gamma, \sigma_e, \mu_I, \zeta_I\right) \\ &= \int_{\delta_e} \int_{\eta_e} \prod_{a_e \in A_e} \int_{\eta_{a_e}} P_{y_{a_e}}\left(\delta_e, \beta, \tau, \alpha_{a_e}\right) \prod_{i=1}^{8} PI_{a_e,i}\left(\alpha_{a_e}\right) \phi\left(\eta_{a_e}\right) \mathrm{d}\eta_{a_e} \phi\left(\eta_e\right) \mathrm{d}\eta_e h\left(\delta_e \mid \mu_\delta, \sigma_\delta\right) \mathrm{d}\delta_e. \end{aligned} \tag{10}$$

The resulting model jointly explains the retention pattern for ego e across all his/her alters as well as the answers to the 8 indicator questions, again for each alter. For a given ego e, this model thus explains A_e binary outcomes of retention, along with the answers to 7 A_e binary and A_e ordered questions. This wealth of data for each ego allows us to incorporate not just detailed patterns of deterministic heterogeneity (both directly in the choice model as well as in the structural equation for the latent strength), but also random heterogeneity across egos in both the baseline retention utilities (δ_e) and the latent strength variable and across ego-alters in the latent strength variable. This model is of course much more complex to estimate than either the binary choice models or the measurement models, and again involves integration on two different levels. Figure 2 represents the modelling framework described in this section. As discussed, all the ego and ego-alter characteristics affect the utility of retention, while the latent variable is only affected by the ego and ego-alter characteristics in 2008. While the stated (emotional) closeness, participation in social activities (binary) and the frequency of face-to-face interactions (ordered) are explicitly reported in the graph, the different types of help that the ego and the alter can exchange are grouped into a single category ("social capital").

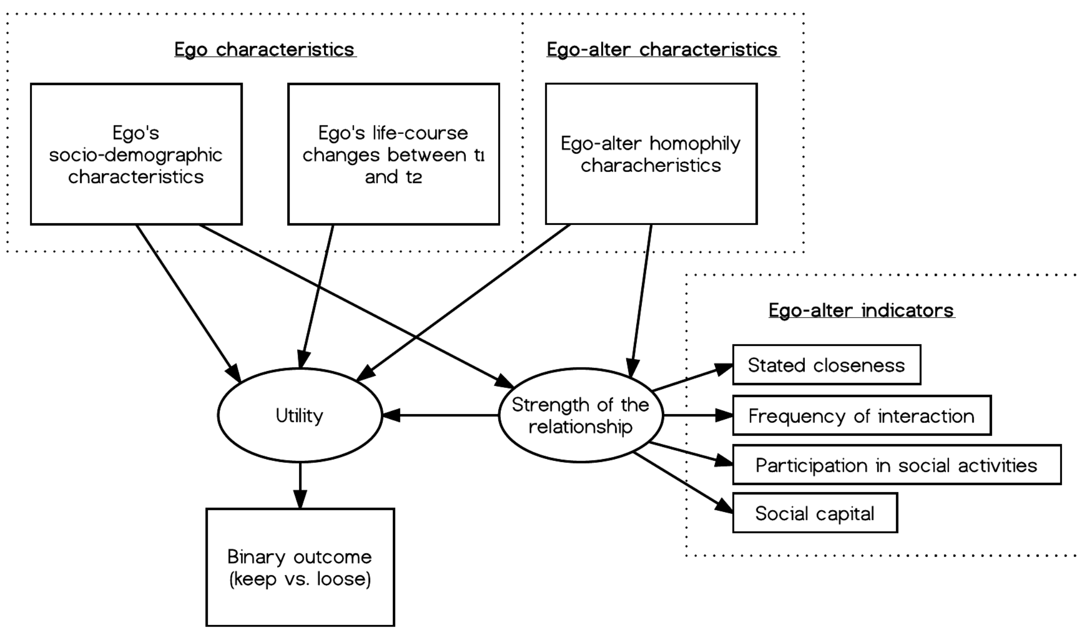

Figure 2: Modelling framework

A number of simplifications of this model are possible. Following the experience with the measurement models, it is clear that making the random heterogeneity in the latent strength purely ego-specific makes little sense, and as such a specification excluding η_{a_e} is not advisable. On the other hand, it is possible to estimate models which exclude the ego-level random heterogeneity in the latent strength (i.e. using only η_{a_e}) as well as models which exclude the additional ego level heterogeneity in the binary choice model by using a fixed δ in Equation 9, i.e. setting $\sigma_\delta = 0$.

3.3 Implementation

All the models presented in this paper were coded and estimated using R (R Core Team, 2016), and in particular the Choice Modelling Centre (CMC) estimation package[3]. To help with estimation times, parallel processing was used in model estimation, especially of the models with two layers of integration.

In all models including random heterogeneity, we made use of 200 Halton draws (Halton, 1960) for each random parameter in simulation-based estimation. We compared this to models using 100 draws and found no substantial differences. The use of 200 draws would clearly be seen as a *low* number in many applications of random coefficients models, and going beyond 200 draws would be easy for the model in Section 3.1. However, it is worth highlighting that the number of draws you use in a model with joint inter and intra-respondent heterogeneity (such as in the measurement models in Section 3.2.2 and the hybrid models in Section 3.2.3 is a different consideration from that in a standard random coefficient model. Indeed, as discussed by Hess and Train (2011), while, for a simple mixed logit model, the use of R draws means that each choice probability is calculated R times, in a model with two layers of mixing, the use of R_{ego} and $R_{ego-alter}$ draws means that each probability is calculated $R_{ego} \cdot R_{ego-alter}$ times. With 200 draws as we use, this implies $40,000$ calculations per probability item (e.g. choice, indicator probability), with the associated implications for CPU time and memory requirements. The CPU requirements should be obvious, and the memory requirements can be understood by noting that prior to the averaging across draws, the estimation software needs to maintain in memory a three dimensional matrix of probabilities with dimensions $(N, R_{ego}, R_{ego-alter})$ where N is the number of observations. In our case, this equates to a matrix with $76,480,000$ elements. In the hybrid models, there is one such matrix for each indicator on top of the choice probabilities.

One further point to briefly address is the fact that different individuals provide different numbers of observations (i.e. ego-alter pairs), as shown in Table 1. In our models, each ego-alter pair has the same weight. We avoided the use of weighted estimation, partly due to concerns about inefficiency, but also as the weighting itself might introduce bias - indeed, our unit of measurement is the ego-alter pair, and we would expect egos with more relationships to provide more information about the way in which relationships are lost or retained. We did conduct a brief comparison between unweighted models and models using weights (so that each ego has the same weight), and found no substantial or significant differences in results, but overall lower levels of significance for the weighted model[4].

[3] http://www.cmc.leeds.ac.uk/resources/software

[4] Results available on request.

4 Results

This section presents the results for the different models estimated in our study. While the main interest is in the hybrid models (results in Section 4.3), we also present results for simple binary choice models (Section 3.1) as well as measurement models for the latent strength without an accompanying choice model (Section 4.2). This is motivated by results showing that if the choice or latent strength components are modelled on their own, the explanatory drivers differ from a situation where they are modelled jointly. This suggests that only the hybrid model is able to disentangle the role that attributes play in these two components.

4.1 Binary choice models

Estimation results for the binary models are presented in Table 3, where we refer to the simple base model as $A1$ and the model with added random heterogeneity as $A2$. We see that $A2$ provides better fit according to the log-likelihood (LL) and Bayesian Information Criterion (BIC). This is also confirmed by a likelihood ratio test between the two models, where $p \sim 1.9 * 10^{-9}$. The μ_δ parameter is negative and significant in both models, indicating that egos are overall more likely to lose their social contacts over time. When estimated, σ_δ is highly significant, confirming the presence of heterogeneity in retention at the ego level.

Table 3: Estimation results for binary retention models

	Model A_1		**Model A_2**	
Final LL	-1,020.19		-1,002.19	
BIC	2,120.95		2,092.51	
	est.	**rob t**	**est.**	**rob t**
μ_δ	-1.866	-9.94	-1.9882	-9.83
σ_δ	n/a	n/a	0.58	-7.47
$\beta_{ego\,landline}$	0.4132	2.25	0.4123	2.13
$\beta_{ego\,student}$	-0.4903	-1.5	-0.5355	-1.6
$\beta_{ego\,divorced\,2008-2012}$	0.517	2.11	0.5584	2.23
$\beta_{ego-alter\,both\,male}$	0.4243	2.81	0.4827	3.12
$\beta_{ego-alter\,both\,female}$	0.4868	3.06	0.5377	3.36
$\beta_{ego-alter\,immediate family}$	1.0467	5.94	1.1385	6.14
$\beta_{ego-alter\,organisation}$	-0.5951	-2.38	-0.5466	-2.45
$\beta_{ego-alter\,known\,under\,one\,year}$	-0.4694	-1.52	-0.489	-1.7
$\beta_{ego-alter\,network\,betweenness}$	0.0185	2.9	0.0218	3.23

As explained above, in A_1 only characteristics of the ego and of ego-alter relationships are used to explain retention. For each effect, we report the estimate (est.) and the robust t-ratio (rob t). In our final specifications, we only retain those variables that have a statistically significant effect, where this list was arrived at after an extensive specification search. All available ego-level socio-demographic effects were tested in the model, in a linear fashion and, where relevant, with non-linear transformations. These included age, sex, level of education, income, different life course changes such as getting married, divorcing or having a child, as well as different social network measures such as network density, number of isolates and different measures of centrality (degree, betweenness and closeness). Ego-alter variables were also tested, including homophily measures, such as same sex or same age, as well as type of relationship between the ego and the alter.

As mentioned in Section 2.1, information about the residential location of alters was also available. Previous research in social network analysis suggests that distance is an important element when it comes to social interactions and establishing social links (e.g. Carrasco et al., 2008a). For this reason, while the change in distance between the ego's and the alters' residential locations was only available for those alters who had been retained in the network, we attempted to develop a model to infer the same information for the alters that were not retained. Unfortunately, due to the small sample and the high level of missing information, as well as to the complexities in the development of such a model, this attempt was not successful. We believe that a similar effort with more suitable data could lead to better results.

Table 3 first reports the ego-level characteristics that help explain the retention of social contacts. In particular, we find that having a landline in 2008 increases the probability of retaining social contacts, while being a student decreases it. The former effect might be related to the fact that owning means of communication helps contact retention; while latter could be related to the fact that students and their social contacts are more likely to experience life changes that might result in losing touch.

At the ego-alter level, we find that gender homophily makes relationships more likely to last over time. Gender homophily, together with other types of homophily, has been found to be important in establishing connections (McPherson et al., 2001). As expected, strong positive effect is found for immediate family members, while people met through clubs and organisations are more likely to be removed from the network. Moreover, alters who have been only known to the ego for one year or less are less likely to be retained than those who have been known for longer. Betweenness centrality is conceptualised as a measure of "control" of the flow of resources in a network (Wasserman and Faust, 1994), and in the context of social activities is normally seen as the level of contact between a node and the rest of the network. The alters with a higher level of betweenness centrality are more likely to be maintained, likely because they have a more important role in the network structure.

4.2 Measurement models

Table 4 reports the estimation results for the measurement models. As specified in the table, the first model ($B1$) only includes heterogeneity at the ego level, the second one ($B2$) only at the ago-alter level, while the last one ($B3$) includes both, so that σ_e is estimated while σ_{a_e} is set to 1, as explained in Section 3.2.1. The model including both the ego and ego-alter heterogeneity performs better than the other two. The p-value for the likelihood ratio test comparing $B3$ to $B1$ is $p \sim 6.3 * 10^{-37}$ while the comparison between $B3$ to $B2$ results in $p \sim 1.4 * 10^{-15}$. $B1$ and $B2$ have the same number of parameters, so the test cannot be performed, but both the LL and BIC are better for $B2$. This makes intuitive sense, as we would expect most of the heterogeneity in strength to be at the ego-alter level, and that serves as a motivation for attempting to include such ego-alter level heterogeneity also in the choice model, which is only possible in the hybrid structure discussed in Section 3.2.3.

For each of the binary indicators used in the measurement model, we report a mean (μ_I), estimated from the binary logit model. For example, $\mu_{I,receive\,advice\,on\,work\,opportunities}$ is negative and significant in all the models, indicating that most egos in the sample do not receive advice about new jobs from most of their social contacts. Frequency of face-to-face interaction is the only non-binary indicator used: respondents could state that they never see the alter face-to-face, that they do so at least once a year, at least once a month, at least once a week, or multiple times per week. We estimate four thresholds in an ordered logit model $t^1 - t^4_{I,face\,to\,face}$, so that the probability that a person answers "never" is the probability that the utility is less than -3.5628 (in model $B1$) and so on. The impact of the latent strength on the indicators is captured by the ζ_I parameters. We see that the latent variable positively affects all the indicators, confirming that α_{a_e} might indeed be measuring relationship strength.

After this first set of estimated parameters, we report the σ terms, which reflect the heterogeneity at the ego and at the ego-alter level, as discussed above. σ_e is significant and its value is less than 1, confirming the presence of heterogeneity in strength at the ego level, but also indicating that there is higher heterogeneity at the ego-alter level.

The set of γ reported in the lower part of the tables represent the significant coefficients of the covariates included in Equation 5. We have included these covariates to explore possible correlations, but the work in sociology looking at relationship strength has mainly relied on using indicators of this purely latent construct, rather than modelling it as a dependent variable. Nevertheless, we believe that some of the effects we obtain are intuitive.

Some ego-level socio-demographics were significant, although their effects are the most difficult to interpret. For example, the age of egos is included as a continuous variable and is negatively related to relationship strength. Older people may be lonelier and not benefit from as much social support and interactions as others, which help build stronger relationships. Being unemployed or a homemaker is also negatively related with having stronger relationships. While this result is difficult to interpret, it is possible that being at University or in employment implied further opportunities to keep in touch with others, and having the financial means to get involved in social activities can help strengthen relationships, while unemployed people might be busy looking for jobs and taking care of the household (as homemakers) and living less socially active lives. On the contrary, one could argue that homemakers have more free time than employed people and they could devote it to deepen their relationship. We believe that this effect be worth further investigation.

Having lived for many years in the same neighbourhood is related with less strong social connections, a result which is somewhat counter-intuitive. The larger the network size of an ego, the less strong his/her individual relationships will tend to be. This result is intuitive, as a larger network will require more resources such as time and other types of effort to maintain, and each node will receive less "care" so that many nodes can be active.

The two measures of network density and network centrality (also referred to as graph centrality) were found to have opposite effects on strength, the first one being negative and the second one positive. Network density is computed as the ratio between the actual connections in a network and the potential connections. Therefore, if it is equal to 1, if means that all the alters know each other. The effect is rather weak in all the models. Network centrality denotes the variations in the point centralities in the network, which in turn represent a measure indicating whether each alter is directly connected with a large proportion of network members. So in networks with high variations in the point centralities, egos are more likely to have strong social contacts, although the effect is only significant in model $B2$. In terms of ego-alter measures, we observe that if the alter is an immediate family member, the relationship is more likely to be strong. The opposite is true if the ego and the alter have known each other for less than one year or were both students in 2008. The latter effect might be due to the fact that if both were students, possibly they met at University they did not establish a strong relationship. As expected, alters with higher level of betweenness are also more likely to be strong contacts. These are the contacts with a central role in the network, so they are likely to be important for the ego. Finally, we found that a higher level of ego-alter degree centrality, or "point centrality" is associated with stronger ties. These alters are likely to be crucial nodes in the network, and therefore it is to be expected that they will be important for the ego as well.

4.3 Hybrid choice models

As described above, we report the results for different versions of the hybrid model, where we jointly estimate the binary choice model and the various measurement models, in Tables 5 and 6. The specific assumptions in terms of heterogeneity in each of the models are specified in the top 3 rows of the table. In the first 3 models, all the heterogeneity is linked to the latent variable α. In C_1, this heterogeneity is only at the ego level, in C_2 it is only at the ego-alter level, while in C_3 it is in both. The likelihood ratio test comparing C_3 to C_1 shows a strongly significant improvement ($p \sim 9.3 * 10^{-51}$) as well as in the case where C_3 is compared to C_2 ($p \sim 1.3 * 10^{-15}$), confirming the need to accommodate the heterogeneity in strength both at the ego and at the ego-alter level. Models C_4 to C_6 include additional ego level heterogeneity in retention not linked to the latent variable, while in order from C_4 to C_6, they use the same specification of the latent variable heterogeneity as in C_1 to $C3$. C_6 is therefore the most complex model,

Table 4: Estimation results for measurement models for latent strength

	Model B_1		**Model B_2**		**Model B_3**	
ego-level heterogeneity	yes		no		yes	
ego-alter-level heterogeneity	no		yes		yes	
Final LL	-7,830.01		-7,781.25		-7,749.43	
BIC	15,894.24		15,796.74		15,740.64	
	est.	**rob t**	**est.**	**rob t**	**est.**	**rob t**
$\mu_{I,give\,advice\,on\,important\,matters}$	-0.4526	-1.16	-0.361	-0.98	-0.4397	-1.22
$\zeta_{I,give\,advice\,on\,important\,matters}$	0.3314	3.39	0.6947	5.11	0.6225	4.19
$\mu_{I,receive\,advice\,on\,work\,opportunities}$	-2.1817	-3.14	-2.1761	-4.29	-2.2758	-4.61
$\zeta_{I,receive\,advice\,on\,work\,opportunities}$	0.3588	1.54	0.7555	3.11	0.6619	2.28
$\mu_{I,give\,emergency\,financial\,support}$	-0.5574	-0.69	-0.8247	-1.23	-0.8731	-1.19
$\zeta_{I,give\,emergency\,financial\,support}$	0.6946	3.01	1.154	4.56	1.2056	3.52
$\mu_{I,receive\,emergency\,financial\,support}$	-1.2196	-1.64	-1.2793	-1.52	-1.4578	-1.84
$\zeta_{I,receive\,emergency\,financial\,support}$	0.6128	3.4	1.4702	5.33	1.3282	4.16
$\mu_{I,receive\,emergency\,transport\,support}$	-1.2216	-1.58	-1.3585	-2	-1.4967	-2.18
$\zeta_{I,receive\,emergency\,transport\,support}$	0.6113	2.76	1.2273	4.65	1.1434	2.98
$\mu_{I,stated\,closeness}$	1.1524	3.11	1.5186	3.61	1.2431	3.29
$\zeta_{I,stated\,closeness}$	0.3723	4.66	0.8574	4.5	0.6677	3.99
$\mu_{I,conduct\,joint\,social\,activities}$	-0.22	-0.89	0.0033	0.01	-0.1209	-0.47
$\zeta_{I,conduct\,joint\,social\,activities}$	0.2305	3.54	0.5708	4.49	0.4687	3.79
$t^1_{I,face\,to\,face}$	-3.5768	-11.54	-3.8853	-11.23	-3.7675	-11.43
$t^2_{I,face\,to\,face}$	-2.1291	-8.26	-2.4106	-8.47	-2.2988	-8.48
$t^3_{I,face\,to\,face}$	-1.1329	-4.67	-1.3777	-5.19	-1.2731	-5.11
$t^4_{I,face\,to\,face}$	0.4352	1.94	0.2642	1.09	0.3546	1.59
$\zeta_{I,face\,to\,face}$	0.1844	3.37	0.4467	4.9	0.3719	3.83
σ_e	1	-	n/a	n/a	0.5344	4.96
σ_{a_e}	n/a	n/a	1	-	1	-
$\gamma_{ego\,age}$	-0.0179	-2.02	-0.013	-2.94	-0.0122	-2.4
$\gamma_{ego\,homemaker}$	-1.0585	-2.62	-0.6498	-3.44	-0.6535	-3
$\gamma_{ego\,unemployed}$	-0.5057	-1.48	-0.3681	-2.26	-0.345	-1.84
$\gamma_{ego\,years\,in\,neighbourhood}$	-0.9752	-1.95	-0.6033	-2.45	-0.6004	-2.05
$\gamma_{ego\,network\,size}$	-0.0841	-3.97	-0.0414	-4.15	-0.048	-4.01
$\gamma_{ego\,network\,density}$	-1.3174	-1.18	-0.6553	-1.1	-0.6848	-1.06
$\gamma_{ego\,network\,centrality}$	1.004	1.06	0.8042	1.8	0.7858	1.53
$\gamma_{ego-alter\,immediate family}$	2.0165	4.58	1.1383	6.01	1.2157	5.74
$\gamma_{ego-alter\,known\,under\,one\,year}$	-0.6766	-1.76	-0.5189	-3.01	-0.4389	-2.1
$\gamma_{ego-alter\,both\,students}$	-1.2855	-2.24	-0.9442	-2.33	-0.8258	-2.24
$\gamma_{ego-alter\,network\,betweenness}$	0.0461	3.85	0.0287	4.59	0.0294	4.83
$\gamma_{ego-alter\,network\,degree\,centrality}$	0.1161	3.02	0.0607	3.91	0.065	3.13

including both ego and ego-alter heterogeneity in the latent variable, as well as ego level heterogeneity in retention. We again see that the most complex model, C_6, offers significant improvements over C_4 and C_5, with p-values of the likelihood ratio tests of $p \sim 7.9 * 10^{-53}$ and $p \sim 7.2 * 10^{-15}$, respectively. We can also compare the models that have the same pattern of heterogeneity in the latent variable but differ in terms of heterogeneity in the retention part of the model. Model C_4 provides a significant improvement over C_1 ($p \sim 3.2 * 10^{-7}$), as does C_5 over C_2 ($p \sim 4.2 * 10^{-10}$) and C_6 over C_3 ($p \sim 2.4 * 10^{-9}$). Finally, we always see that the using ego-alter level heterogeneity in the latent variable is better than ego-level heterogeneity ($C1$ and $C2$, and $C4$ and $C5$). These results show, as expected, that allowing for additional random heterogeneity in both retention and latent strength, both at the ego and at the ego-alter level, significantly improves model fit.

The overall log-likelihoods of the hybrid models are clearly more negative than those of the simple binary choice models as they also incorporate the likelihood of the measurement model component. It is possible, as we have done in Table 5 to factor out the component of the log-likelihood that relates to the choice model component only. Some caution is required to not over-interpret these values however. As discussed at length by Vij and Walker (2016), a hybrid choice model is not theoretically able to offer better fit to the choice model component than an equally flexible choice model estimated on its own. A slightly different picture applies in the context of our application, as the simple binary models in Table 3 only incorporate ego-level heterogeneity. As such, it is not contradictory that a slightly better fit to the choice component alone is obtained by models C_4 to C_6. However, this improvement is negligible - the real benefit of the model is that it can disentangle the different layers of heterogeneity which is behaviourally appealing. On the other hand, it is worth discussing why a lower fit to the choice component is obtained by models C_1 to C_3. The key reason here is that these three models force all random heterogeneity (whether ego-level or ego-alter level) to be in the latent strength variable. As this variable needs to also explain the heterogeneity in the measurement model (on top of the choice model), it is then not surprising that it offers a poorer explanation of the choices alone. Additionally, it should be noted that the utility specification differs between the simple choice models and the choice component of the hybrid model, such that no formal comparison of fit is possible anyway.

Table 5 reports the results related to the retention part of the model (although of course the two parts of the model were jointly estimated). Where present, the heterogeneity in retention in the choice model (σ_δ) is significant, meaning that there are differences across egos in retention independent of the latent strength variable, even when the latent strength incorporates heterogeneity at the ego level. The τ parameters represents the effect impact of the latent variable α_{a_e} on the outcome of retention. The estimated coefficient is positive and significant in all the specifications, although it is worth noting that it is weaker (the coefficient is smaller in absolute value and less significant) in models C_1 and C_4, i.e. where we do not allow for ego-alter heterogeneity in strength. Given the discussion in Section 4.2 about the interpretation of the latent variable, a positive τ

means that a higher value in the latent variable has a positive impact on the likelihood of contact retention. The actual sign of this can only be interpreted together with the findings from the measurement model in Table 6. We see that all ζ parameters are again positive, confirming the directionality of the latent variable as a positive strength. This shows that a higher latent strength increases the likelihood of retaining a social contact in the choice model (positive τ). In terms of random heterogeneity introduced by the latent strength into the choice models, models all models except C_1 and C_4 introduce significant heterogeneity at the ego-alter level, something that was not possible with the binary choice models alone.

The socio-demographic and network characteristics also need to be jointly studied in the two model components, given that for numerous variables, a significant effect arises in both[5]. Some effects matter only in the choice model. This includes a positive impact on retention for egos who have a landline in 2008 (although this is only weakly significant in model C_2), a negative impact for students in 2008 or those who went to university between 2008 and 2012 and a positive impact if both ego and alter are male. The first two effects were also found in models A_1 and A_2. Students who went to University between the two waves are likely to have removed some connections to form new ones, as suggested by the literature associating life course events and social network changes (e.g. Degenne and Lebeaux, 2005). As mentioned above, gender homophily is confirmed to be an important factor for the choice of contacts.

Similarly, some effects are only present through the latent strength variable. This includes reduced relationship strength for older respondents, those not employed and homemakers, as also found for models B_1-B_3. Network size and density have a negative impact on latent strength, while ego network centrality has a positive impact, as has ego-alter network betweenness. There is increased strength if both ego and alter are aged over 60 or if both are professionals, with reduced strength for newer acquaintances. The first two effects confirm the importance of homophily effects (McPherson et al., 2001), importantly not only for retention but also for strength. We also find it reasonable that people who have known each other for a short time have not had the time to deepen their relationship.

The most interesting findings arise when looking at those variables present in both model components. Here we need to look at the sum of $\beta+\tau\gamma$. While we see positive signs for both β and γ for female gender homophily and for immediate family (as expected), the same is not the case for three other measures. We see that, like in the measurement models alone, having lived in a neighbourhood for longer reduces the latent strength, and this outweighs the positive sign for β in the choice model alone, for all the models except C_1.

If both ego and alter are students, this reduces the latent strength, but the final impact on the utility of retention remains positive ($1.0205 + 0.858 \cdot -0.7946 = 0.3387$ for

[5] With the exception of changes in ego characteristics between 2008 and 2012, which should not affect latent strength in 2008, we tested the impact of all measures both in the choice model and the structural equation for the latent variable, thus reduce the risk of misattribution of effects (Vij and Walker, 2016)

model C_6). Finally, the rather strong positive effect of ego-alter network centrality in the latent variable is dampened in the choice model by the less strong β, so that the effect is weaker in the choice model. These results validate our approach of testing the role of these variables in both models where the inclusion in only the structural equation would have assumed a common impact on latent strength and retention, while the inclusion in only the choice model would have prevented us from understanding the drivers of latent strength.

The ζ_I parameters for the measurement models showed in Table 6 are weaker (the coefficients are smaller in absolute value and less significant) in models C_1 and C_4, i.e. where we do not allow for ego-alter heterogeneity in strength. The signs are still as expected, so the interpretation does not change. σ_e, when estimated (in models C_3 and C_6) is significant and smaller than 1, suggesting that there is higher heterogeneity at the ego-alter level than at the ego level.

5 Conclusions

Our study investigated social networks dynamics over time, with a particular focus on the retention in the social network (vs. loss) of social contacts over a four-year interval. Our results unveil interesting insights, showing that both ego socio-demographics and life-course changes, as well as ego-alter characteristics, have a significant impact on retention.

The key contribution of our work comes in attempts to disentangle sources of heterogeneity and in particular to allow for ego-alter level random heterogeneity. This is important due to a key limitation of the type of data collected for social network evolution. Indeed, the vast majority of such data are collected from the point of view of the ego, and the analyst can only rely on information provided by one of the *dancers*, if we use the metaphor *"it takes two to tango"*. This already creates significant scope for heterogeneity at the ego-alter level. As an example, a relationship might have ended not because the ego decided to interrupt it, but because the alter was no longer interested in it. Explaining such an outcome on the basis of ego-level only data is problematic and creates clear scope for the type of heterogeneity we introduce in our models.

We accommodate this ego-alter level random heterogeneity through a hybrid framework in which we develop a latent variable for relationship strength, which varies both across egos and across ego-alter pairs. We use this to explain a number of indicators of relationship strength as well as accommodating heterogeneity in the choice model (on top of independent heterogeneity).

Our findings on a typical name generator dataset underline the relevance of random ego-level heterogeneity in retention, as well as both ego and ego-alter level heterogeneity in strength. The former means that some egos will retain more social contacts than others, while the latter implies that some egos will be more prone to establishing strong relationships than others (variation across egos) and that even within a specific ego's network, certain ties will be stronger than others (variation across alters, for an ego).

This paper has focused on understanding the process of retention of social contacts.

Table 5: Estimation results for hybrid models (retention part)

	model C_1		**model C_2**		**model C_3**		**model C_4**		**model C_5**		**model C_6**	
ego-level heterogeneity in α	yes		no		yes		yes		no		yes	
ego-alter-level heterogeneity in α	no		yes		yes		no		yes		yes	
ego-level heterogeneity in δ	no		no		no		yes		yes		yes	
Final LL	-8,830.10		-8,749.79		-8,717.84		-8,817.05		-8,730.30		-8,700.04	
LL of the choice model	-1,007.72		-1,019.76		-1,006.80		-1,000.26		-1,001.02		-1,001.52	
LL of the measurement model	-7,819.33		-7,776.80		-7,744.30		-7,819.60		-7,776.44		-7,743.85	
BIC	18,000.21		17,839.60		17,783.25		17,981.66		17,808.18		17,755.20	
	est.	**rob *t***	**est.**	**rob *t***	**est.**	**rob *t***	**est.**	**rob *t***	**est.**	**rob *t***	**est.**	**rob *t***
μ_δ	-0.782	-1.62	-0.4825	-0.98	-0.5054	-1.08	-0.7147	-1.42	-0.3246	-0.58	-0.325	-0.62
σ_δ	n/a	n/a	n/a	n/a	n/a	n/a	0.5436	6.69	0.669	5.95	0.6725	6.68
τ	0.2837	3.59	0.7886	6.9	0.7112	6.13	0.3166	4.04	0.9342	8.11	0.858	7.8
$\beta_{ego\,landline}$	0.31	1.63	0.3465	1.66	0.3191	1.54	0.2812	1.49	0.2992	1.32	0.2873	1.33
$\beta_{ego\,years\,in\,neighbourhood}$	0.2716	0.95	0.3609	1.14	0.3455	1.07	0.2249	0.79	0.304	0.81	0.3275	0.92
$\beta_{ego\,student}$	-0.8008	-2.68	-1.0329	-2.98	-1.0144	-3.06	-0.8674	-2.97	-1.1596	-3.14	-1.1755	-3.38
$\beta_{ego\,went\,to\,university\,2008-2012}$	-0.8455	-3.21	-0.9183	-3.36	-1.0997	-4.27	-0.8599	-2.48	-1.0593	-3.23	-1.1846	-3.55
$\beta_{ego-alter\,both\,male}$	0.4475	2.86	0.459	2.84	0.4574	2.77	0.5092	3.29	0.5331	3.17	0.5397	3.17
$\beta_{ego-alter\,both\,female}$	0.3681	2.46	0.3791	2.18	0.3655	2.12	0.3981	2.62	0.4238	2.3	0.3993	2.21
$\beta_{ego-alter\,immediate\,family}$	0.435	1.83	0.2246	1.12	0.237	1.22	0.4869	2	0.2086	0.94	0.2243	1.04
$\beta_{ego-alter\,both\,students}$	0.6747	1.9	0.7463	1.88	0.7806	1.86	0.7793	1.85	1.0393	2.03	1.0205	1.95
$\beta_{ego-alter\,network\,degree\,centrality}$	-0.0301	-2.68	-0.0395	-3.65	-0.0383	-3.41	-0.0203	-1.25	-0.0314	-1.87	-0.0308	-1.82

Table 6: Estimation results for hybrid models (latent strength and measurement model part)

	model C_1		model C_2		model C_3		model C_4		model C_5		model C_6	
	est.	**rob t**	**est.**	**rob t**	**est.**	**rob t**	**est.**	**rob t**	**est.**	**rob t**	**est.**	**rob t**
$\mu_{I,give\,advice\,on\,important\,matters}$	-0.4041	-0.95	-0.3765	-0.95	-0.4404	-1.22	-0.2914	-0.66	-0.3294	-0.83	-0.3657	-0.97
$\zeta_{I,give\,advice\,on\,important\,matters}$	0.3364	3.87	0.7184	5.63	0.6376	4.83	0.3501	3.86	0.7166	5.52	0.6445	4.74
$\mu_{I,receive\,advice\,on\,work\,opportunities}$	-2.1157	-3.41	-2.2139	-4.64	-2.2692	-5.06	-2.0147	-2.83	-2.1632	-4.46	-2.2004	-4.73
$\zeta_{I,receive\,advice\,on\,work\,opportunities}$	0.3759	1.64	0.7641	3.2	0.6924	2.49	0.3823	1.61	0.765	3.16	0.6907	2.45
$\mu_{I,give\,emergency\,financial\,support}$	-0.5585	-0.67	-0.9119	-1.35	-0.9276	-1.36	-0.367	-0.4	-0.8339	-1.22	-0.7961	-1.08
$\zeta_{I,give\,emergency\,financial\,support}$	0.6564	3.05	1.116	4.65	1.1308	3.96	0.6747	2.92	1.1222	4.69	1.1599	3.89
$\mu_{I,receive\,emergency\,financial\,support}$	-1.1688	-1.48	-1.3495	-1.54	-1.4543	-1.78	-1.0378	-1.25	-1.2581	-1.44	-1.3206	-1.6
$\zeta_{I,receive\,emergency\,financial\,support}$	0.6107	3.35	1.5206	5.4	1.4433	4.4	0.6023	3.18	1.4953	5.47	1.3969	4.48
$\mu_{I,receive\,emergency\,transport\,support}$	-1.2058	-1.81	-1.4364	-2.21	-1.5105	-2.43	-1.0406	-1.38	-1.3525	-2.06	-1.389	-2.17
$\zeta_{I,receive\,emergency\,transport\,support}$	0.5859	2.67	1.2003	4.6	1.1327	3.31	0.6008	2.63	1.213	4.58	1.1539	3.21
$\mu_{I,stated\,closeness}$	1.1868	3.17	1.5186	3.6	1.2993	3.56	1.2833	3.51	1.5344	3.72	1.3512	3.76
$\zeta_{I,stated\,closeness}$	0.3689	4.9	0.8892	5.03	0.7149	4.32	0.3737	5.05	0.8615	5.1	0.7018	4.45
$\mu_{I,conduct\,joint\,social\,activities}$	-0.1682	-0.6	-0.021	-0.07	-0.0918	-0.34	-0.1125	-0.4	0.002	0.01	-0.0579	-0.21
$\zeta_{I,conduct\,joint\,social\,activities}$	0.2399	3.87	0.5785	4.75	0.5011	4.3	0.2404	3.99	0.5658	4.72	0.4884	4.25
$t^1_{I,face\,to\,face}$	-3.6119	-12.44	-3.8805	-11.38	-3.8159	-12.02	-3.658	-11.96	-3.9158	-11.6	-3.8584	-12.17
$t^2_{I,face\,to\,face}$	-2.1623	-9.26	-2.4029	-8.7	-2.34	-9.24	-2.2092	-8.73	-2.4374	-8.81	-2.3825	-9.28
$t^3_{I,face\,to\,face}$	-1.1647	-5.24	-1.3683	-5.18	-1.3086	-5.49	-1.212	-5.04	-1.402	-5.33	-1.3509	-5.61
$t^4_{I,face\,to\,face}$	0.4062	2.01	0.2765	1.14	0.3285	1.54	0.359	1.62	0.2444	1.01	0.2871	1.32
$\zeta_{I,face\,to\,face}$	0.1873	3.51	0.4582	5.16	0.4004	4.43	0.1891	3.58	0.4582	5.3	0.3999	4.44
σ_e	1	-	n/a	n/a	0.4979	5.78	1	-	n/a	n/a	0.5023	5.62
σ_{a_e}	n/a	n/a	1	-	1	-	n/a	n/a	1	-	1	-
$\gamma_{ego\,age}$	-0.0254	-2.43	-0.0157	-3.18	-0.0155	-2.86	-0.0285	-2.95	-0.0158	-3.17	-0.0167	-2.92
$\gamma_{ego\,homemaker}$	-1.1279	-2.93	-0.5961	-3.29	-0.6301	-3.13	-1.1971	-3.25	-0.623	-3.42	-0.6644	-3.3
$\gamma_{ego\,unemployed}$	-0.6561	-1.95	-0.3771	-2.34	-0.4003	-2.25	-0.7791	-2.47	-0.3898	-2.38	-0.4407	-2.48
$\gamma_{ego\,years\,in\,neighbourhood}$	-0.8526	-1.46	-0.5503	-2.13	-0.5098	-1.74	-0.9225	-1.7	-0.5458	-2.09	-0.528	-1.73
$\gamma_{ego\,network\,size}$	-0.0905	-3.85	-0.0396	-3.74	-0.0482	-3.94	-0.0862	-3.81	-0.0424	-4.15	-0.0482	-4.08
$\gamma_{ego\,network\,density}$	-1.9829	-1.78	-0.8733	-1.68	-1.0605	-1.84	-2.1787	-2.21	-0.8991	-1.72	-1.1368	-2.04
$\gamma_{ego\,network\,centrality}$	1.3881	1.76	0.8907	2.46	0.9484	2.38	1.5276	2.05	0.9045	2.36	1.0083	2.45
$\gamma_{ego-alter\,both\,female}$	0.4363	2.24	0.1319	1.23	0.206	1.95	0.4119	2.14	0.1382	1.28	0.2001	1.9
$\gamma_{ego-alter\,both\,over\,60}$	0.7929	1.84	0.3572	1.48	0.4244	1.81	0.7664	1.85	0.3689	1.53	0.4136	1.87
$\gamma_{ego-alter\,immediate\,family}$	2.0806	4.81	1.143	6.27	1.21	6.16	2.021	4.52	1.1403	6.23	1.1946	6.14
$\gamma_{ego-alter\,known\,under\,one\,year}$	-0.7387	-1.83	-0.5376	-3.37	-0.4513	-2.25	-0.726	-1.84	-0.5289	-3.26	-0.4546	-2.3
$\gamma_{ego-alter\,both\,students}$	-1.2313	-2.2	-0.8996	-2.33	-0.7546	-2.13	-1.2599	-2.3	-0.9247	-2.39	-0.7946	-2.26
$\gamma_{ego-alter\,both\,professionals}$	0.7338	1.55	0.2089	1.17	0.3459	1.72	0.7016	1.51	0.2088	1.15	0.3353	1.66
$\gamma_{ego-alter\,network\,betweenness}$	0.0467	3.71	0.0276	4.47	0.0279	4.65	0.0458	3.57	0.0276	4.55	0.0277	4.68
$\gamma_{ego-alter\,network\,degree\,centrality}$	0.1366	3.17	0.0653	4.32	0.0756	3.66	0.1344	3.3	0.0669	4.46	0.0756	3.72

Differently from the social networks literature, it proposed a more advanced approach to this behavioural process, accommodating different factors that are shown to have a significant effect on retention of social contacts and relationship strength. The treatment of ego and dyad-level random heterogeneity adds to this effort, capturing substantial variation at both levels and providing a more detailed picture of real-life behaviour. While these modelling techniques have been (to some extent) applied in the field of choice modelling, their application to social network evolution represents an innovation in two different fields. In addition, the work itself is also useful for the choice modelling community as it shows how the use of a hybrid model can be helpful in accounting for heterogeneity at different levels (i.e. inter and intra-person).

The findings we presented could be operationalised to forecast the composition of social networks, as well as the strength of social networks, as they both depend on socio-demographic characteristics of the ego and the alter. This could allow analysts to produce better predictions of other decisions that are connected to an individual's social networks.

While our results are in line with expectation and provide interesting insights, we acknowledge some limitations of the current work due to the data used. We made use of a two-wave dataset from Chile, where social network data were collected by means of two name generators. Name generators have raised concerns due to potential recall biases (Bell et al., 2007), and in our case we are modelling whether egos recall each alter in the second wave, not whether the alter is still in the network. Collecting reliable social network data and efforts into exploring different data sources, especially panel, is a research area where further effort is needed. Similarly, a larger sample might provide more detailed insights as well as allowing for validation with out-of-sample prediction testing. Moreover, while the application of this model is of course possible, the type and amount of information needed for accurate application might not be trivial to gather.

Nevertheless, we believe this study to be a necessary step to develop methods and tools that will be successively refined, and future work will be able to focus on their application and on the development of better techniques for data collection.

Finally, an important area for future work is looking beyond which alters leave a network and also incorporate the arrival of new alters.

6 ACKNOWLEDGMENTS

The Leeds authors acknowledge the financial support by the European Research Council through the consolidator grant 615596-DECISIONS. The fourth author acknowledges the financial support by the Chilean Research, Science, and Technology Council (CONICYT), Programa de Financiamiento Basal para Centros de Excelencia, Instituto de Sistemas Complejos de Ingeniería, FBO-16.

References

Almquist, Z. W., Butts, C. T., 2013. Dynamic network logistic regression: A logistic choice analysis of inter-and intra-group blog citation dynamics in the 2004 us presidential election. Political analysis 21 (4), 430–448.

Bakshy, E., Rosenn, I., Marlow, C., Adamic, L., 2012. The role of social networks in information diffusion. In: Proceedings of the 21st international conference on World Wide Web. ACM, pp. 519–528.

Bell, D. C., Belli-McQueen, B., Haider, A., 2007. Partner naming and forgetting: recall of network members. Social networks 29 (2), 279–299.

Calastri, C., Hess, S., Daly, A., Carrasco, J. A., 2017a. Does the social context help with understanding and predicting the choice of activity type and duration? an application of the multiple discrete-continuous nested extreme value model to activity diary data. Transportation Research Part A: Policy and Practice 104, 1–20.

Calastri, C., Hess, S., Daly, A., Maness, M., Kowald, M., Axhausen, K., 2017b. 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.

Campbell, K. E., Lee, B. A., 1991. Name generators in surveys of personal networks. Social Networks 13 (3), 203 – 221.
URL http://www.sciencedirect.com/science/article/pii/037887339190006F

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

Carrasco, J. A., Cid-Aguayo, B., 2012. Network capital, social networks, and travel: an empirical illustration from concepción, chile. Environment and Planning A 44 (5), 1066–1084.

Carrasco, J. A., Hogan, B., Wellman, B., Miller, E. J., 2008b. Collecting social network data to study social activity-travel behavior: an egocentric approach. Environment and planning. B, Planning & design 35 (6), 961.

Chávez, Ó., Carrasco, J.-A., Tudela, A., 2017. Social activity-travel dynamics with core contacts: evidence from a two-wave personal network data. Transportation Letters, 1–10.

Degenne, A., Lebeaux, M.-O., 2005. The dynamics of personal networks at the time of entry into adult life. Social networks 27 (4), 337–358.

Dugundji, E., Walker, J., 2005. Discrete choice with social and spatial network interdependencies: an empirical example using mixed generalized extreme value models with field and panel effects. Transportation Research Record: Journal of the Transportation Research Board (1921), 70–78.

Granovetter, M. S., 1973. The strength of weak ties. American journal of sociology 78 (6), 1360–1380.

Halton, J. H., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Mathematik 2 (1), 84–90.

Hanneke, S., Fu, W., Xing, E. P., et al., 2010. Discrete temporal models of social networks. Electronic Journal of Statistics 4, 585–605.

Hess, S., Rose, J. M., 2009. Allowing for intra-respondent variations in coefficients estimated on repeated choice data. Transportation Research Part B: Methodological 43 (6), 708–719.

Hess, S., Train, K. E., 2011. Recovery of inter-and intra-personal heterogeneity using mixed logit models. Transportation Research Part B: Methodological 45 (7), 973–990.

Kapferer, B., 1969. Norms and the manipulation of relationships in a work context. Social Networks in Urban Situations.

Kempe, D., Kleinberg, J., Tardos, É., 2003. Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 137–146.

Maness, M., Cirillo, C., Dugundji, E. R., 2015. Generalized behavioral framework for choice models of social influence: Behavioral and data concerns in travel behavior. Journal of transport geography 46, 137–150.

Marsden, P. V., 1990. Network data and measurement. Annual review of sociology 16 (1), 435–463.

Marsden, P. V., Campbell, K. E., 1984. Measuring tie strength. Social forces 63 (2), 482–501.

McPherson, M., Smith-Lovin, L., Cook, J. M., 2001. Birds of a feather: Homophily in social networks. Annual review of sociology 27 (1), 415–444.

Mollenhorst, G., Volker, B., Flap, H., 2014. Changes in personal relationships: How social contexts affect the emergence and discontinuation of relationships. Social Networks 37, 65–80.

Páez, A., Scott, D. M., 2007. Social influence on travel behavior: a simulation example of the decision to telecommute. Environment and Planning A 39 (3), 647.

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/

Sadri, A. M., Ukkusuri, S. V., Gladwin, H., 2017. Modeling joint evacuation decisions in social networks: The case of hurricane sandy. Journal of Choice Modelling.

Sharmeen, F., Arentze, T., Timmermans, H., 2014. An analysis of the dynamics of activity and travel needs in response to social network evolution and life-cycle events: a structural equation model. Transportation Research Part A: Policy and Practice 59, 159–171.

Sharmeen, F., Arentze, T., Timmermans, H., 2015. Predicting the evolution of social networks with life cycle events. Transportation 42 (5), 733–751.

Sharmeen, F., Chávez, Ó., Carrasco, J.-A., Arentze, T., Tudela, A., 2016. Modeling population-wide personal network dynamics using two-wave data collection method and origin-destination survey. In: Transportation Research Board 95th Annual Meeting. No. 16-3387.

Simmel, G., 1950. The sociology of georg simmel. Vol. 92892. Simon and Schuster.

Snijders, T. A., Van de Bunt, G. G., Steglich, C. E., 2010. Introduction to stochastic actor-based models for network dynamics. Social networks 32 (1), 44–60.

Vij, A., Walker, J. L., 2016. How, when and why integrated choice and latent variable models are latently useful. Transportation Research Part B: Methodological 90, 192–217.

Walker, J. L., Ehlers, E., Banerjee, I., Dugundji, E. R., 2011. Correcting for endogeneity in behavioral choice models with social influence variables. Transportation Research Part A: Policy and Practice 45 (4), 362–374.

Wasserman, S., Faust, K., 1994. Social network analysis: Methods and applications. Vol. 8. Cambridge university press.

Wellman, B., Wong, R. Y.-l., Tindall, D., Nazer, N., 1997. A decade of network change: Turnover, persistence and stability in personal communities. Social networks 19 (1), 27–50.