Similarity in transition to adulthood of friends and school mates

Nicola Barban (ISER, University of Essex);
Anette Fasang (Humboldt; WZB);
Nicoletta Balbo (Universita' Bocconi)

October 2019

Abstract

Previous research on 'linked lives' has mainly focused on similarity in life course outcomes of parents and children, between siblings or spouses, neglecting the potentially powerful impact of friends and schoolmates. This paper uses data from the National Longitudinal Study of Adolescent and Adult Health (Add Health) to investigate the similarity of life course trajectories in the transition to adulthood of a national representative sample of young women in the US. Using recent methodological innovation in sequence analysis, we first estimate the similarity in life course trajectories among friends and peers. In the second part of the paper we combine sequence analysis to causal inference to estimate the causal effect of friends' life course transitions in respondents' transition to parenthood, marriage and cohabitation. Results indicate that friend' trajectories are more similar than random school-mates but less than siblings. Although friends seem to have a direct effect on transition to adulthood, the effect is reduced once we control for previous trajectories and other confounders.

**Paper prepared for the European Population Conference 2020.
Please do not cite or quote**

1 Introduction

Life course trajectories are embedded in social contexts. However, the importance of social interaction in life course has not been coupled with a satisfactory body of empirical research at the micro level (Balbo and Barban, 2014) Because individual lives are lived interdependently, changes in one person's life patterns often lead to changes in other people's lives as well (Elder, 1985). *To what extent are life course decisions, such as family formation, divorce and separation or leaving the parental home, influenced by the behaviour of significant others?*

Previous research on 'linked lives' has mainly focused on similarity in life course outcomes of parents and children, between siblings or spouses, neglecting the potentially powerful impact of friends and schoolmates. Young adults tend to turn away from their parental homes but are not yet embedded in stable relationships and therefore spend significant amounts of time with friends and schoolmates. Particularly, during this life course stage, friends function as role models and reference points for own life choices. As a possible consequence of the second demographic transition, (Lesthaeghe and Van de Kaa, 1986) friends may be equally or more important than siblings in affecting life course decisions, mainly for two reasons. First, declining fertility has led to smaller families—that is, more singletons and fewer siblings. Siblings' roles have likely been replaced by close friends. Second, friends are freely chosen by an individual and voluntary relationships have gained in importance compared to ascribed family relationships.

Our previous work (Balbo and Barban, 2014) shows that a friend's childbearing increases an individual's risk of becoming a parent. The effect is short-term and curvilinear: an individual's risk of childbearing starts increasing after a friend's childbearing, reaches its peak approximately two years later, and then decreases. We extended the previous study in three ways. Firstly, we take an holistic approach in describing the entire trajectory of transition to adulthood by examining the similarity among friends and schoolmates. Secondly, we compare friendship similarity with sibling similarity. Thirdly, we analyze the causal effect of friendship events on the respondents' trajectories to adulthood by adopting a matching design that takes into account both time-fixed confounders and pre-treatment trajectories, combining sequence analysis and propensity score matching.

We use data from the National Longitudinal Study of Adolescent and Adult Health (Add-Health) we will investigate to what degree friends share similar life course trajectories and influence each other in life course decisions. We use recent methodological developments in sequence analysis for dyadic data (Raab et al., 2014; Balbo and Barban, 2014) to describe similarity in life course trajectories and to identify peer effect in life course transitions. Furthermore, we will examine whether demographic trajectories of high school friends are more similar among the lower educated, since they do not move to college and form new friendship networks. Understanding the role of social interactions in shaping individual life course decisions may help in identifying successful policy interventions.

The first part of the paper investigates if lifecourse trajectories of friends are more similar to randomnly drawn respondents from the sample or previous schoolmates. We then extend the analysis and compare friends' similarity to siblings. The second part of the paper investigates the effect of friends' life course events (transition to parenthood; first marriage and first cohabitation) on the respondents life course trajectories.

2 Data

We use data from the National Longitudinal Study of Adolescent Health (Add Health) in the United States,nationally representative longitudinal survey of adolescents who were in grades 7 through 12 in Wave I (1995). The Add Health cohort (born between 1976 and 1982) was followed into young adulthood with four in-home interviews (Wave I in 1995, Wave II in 1996, Wave III in 2001 to 2002, and Wave IV in 2008 to 2009), at the end of which the sample was between 26 and 33 years old. Add Health provides an opportunity to combine three different types of information: longitudinal data on respondents' socioeconomic characteristics; information on life course events and trajectories; and data on friendship networks.

Friendship information have been collected as follows. In Wave I, in-school network information was collected and up to 10 friendship ties for each respondent were identified. In Wave III, a follow-up of the Wave I network module was administered to 3,572 respondents, who were in 7th or 8th grade at Wave I. In the *friends module* of Wave III, respondents were asked a battery of questions about their current relationship with 10 former schoolmates. These 10 people were selected into a respondent's questionnaire by a name generator based on the probability of remaining friends with that respondent. Every schoolmate selected was also part of the study sample. Following our previous work (Balbo and Barban, 2014), we used information on friendship status at Wave III to defined two typologies of network relationship: peers (i.e., former schoolmates who have never been friends) and friends (i.e., former schoolmates who became friends during high school and remained so over time).

The friend module of Add Health contains information on 27,803 dyads, 4,611 of which are listed as current friends in Wave III, 3,859 are former friends and 9,220 are pairs of individuals that are listed by the computer name generator and are not friends. Once merged with the life course trajectories of transition to adulthood, the sample reduced to 3,100 female respondents.

3 Friends' similarity

3.1 Sequence definition

Life course trajectories are represented by monthly combinations of union and childbearing states from age 15 until wave IV (age 28-32). We designed the state space to take six possible values: Single (S); Single Parent (SP); Cohabiting (C); Cohabiting Parent (CP); Married (M) and Married Parent (MP) (Barban, 2013). The timing in which family events take place are defined using retrospective questions. Sequences are calculated only for female respondents for two main reasons: first, women experience family life course events earlier than men; second, women are more reliable on the timing of life course events (e.g. pregnancy) in the Add Health sample.

A life course trajectory can be described as the representation, over the course of an individual's age (or any other alternative time reference), of an ordered sequence of life course events. The concept of trajectory derives from the representation of the life course paradigm proposed by Elder (1985), in which trajectories are based on the occurrence of events in multiple life domains. For example, one may want to describe the evolution of residential independence, marital status, and childbearing over an individual's life course.

Trajectories can be envisioned as sequences of transitions that are enacted over time. A life course transition is a discrete life change or event within a trajectory (e.g., from single to married) often accompanied by socially shared ceremonies and rituals, such as a graduation or a wedding ceremony. On the other hand, a trajectory is a long-term pathway, with age-graded patterns of development in major social institutions such as the family. A primary objective of life course analysis is to study the entire development of life trajectories for different groups of individuals. In particular, researchers are often interested in the *timing* (at what age different life transitions happen), the *quantum* (what and how many transitions happen), and the *sequencing* (which transition comes first and which after) of life course events.

A convenient representation of life course trajectories is to depict them as sequences of linked states within a conceptually defined range of behaviors or experiences. In statistical terms, this can be translated into a categorical time series. For each individual i we can then associate a variable s_{it} that indicates her/his life course status at time t. As one can assume that s_{it} takes a finite number of values, trajectories are then represented as categorical time series.

In other terms, trajectories can be represented as strings or sequences of characters, with each character denoting one particular state. The state-space, (i.e the set of possible states from which sequences are constructed) has a finite number of elements and represents all the possible combinations of events that an individual can take in each time period.

For instance, the marital status trajectory of a woman who is single for four years since the start of our observation (e.g., age 18), then starts a cohabitation lasting three years and then marries and remains married for seven years can be described as follows:

`SSSSCCCMMMMMMM`

More formally, we can define the life course trajectory of length T of individual i as the set of realizations of a discrete-time stochastic process $S_t : t \in T$ with state-space $\Sigma = \{\sigma_1, \ldots \sigma_K\}$. The trajectory of the individual i can be then described by the sequence $s_i = \{s_{i1} \ldots s_{iT}\}$.

Life course sequences can be described with graphical representation (i.e. chronograms, see Figure 3) or by studying the occurrence, timing and order of particular events (Barban, 2013).

3.2 Sequence similarity

We merged the life course trajectories with the *friends module* included in the Wave III of Add Health. We obtained information on 27,803 dyads that include friendship relationship, former friendship (respondents who were friends in the past, but discontinued their relationships) and peers (respondents nominated by the name-generator in Wave III).

Friends share more similar characteristics than peers such as geographical proximity, same racial background and family characteristics than peers (see Table 3).

Most often sequence analysis is used to quantify distances between categorical time series, i.e. life course trajectories.

Optimal Matching algorithm (OM) is the most used algorithm that has been applied within social sciences. Basically, OM expresses distances between sequences in terms of the minimal amount of effort, measured in terms of edit operations, that is required to change two sequences such that they become identical. The OM dissimilarity measures are derived from the measure originally proposed

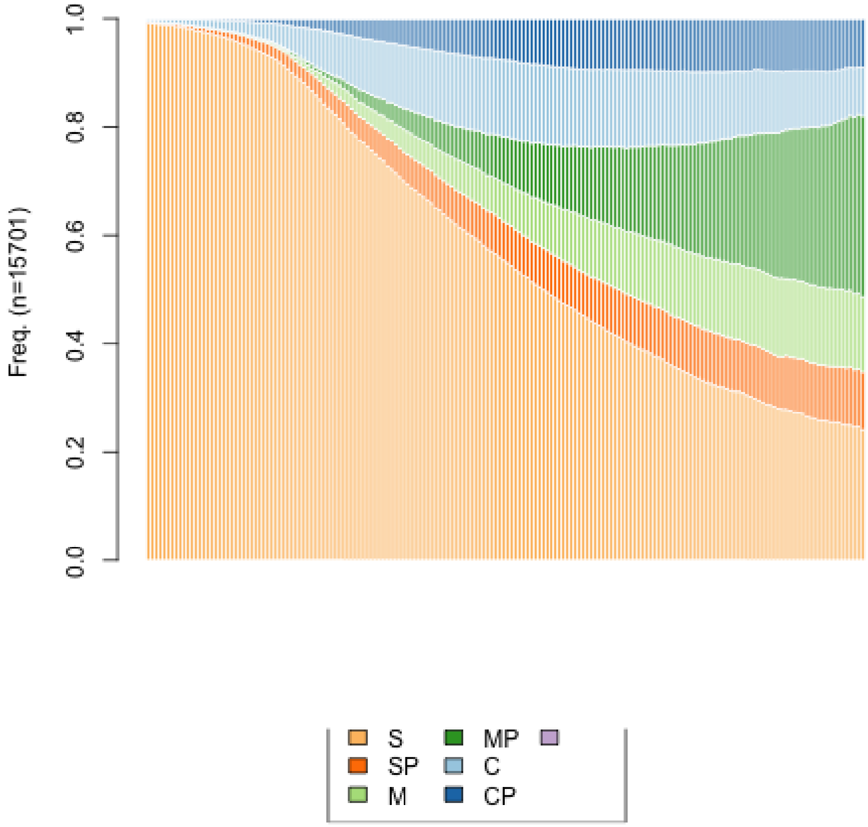

Figure 1: Distribution plot

Dyad Type	Close Friends	Former Friends	Peers
Avg. Distance in miles at WI	7.28	7.29	7.99
Prop. same race	0.82	0.78	0.71
Prop. living in same state at WIII	0.79	0.74	0.77
Prop living in same county at WIII	0.52	0.45	0.51
Prop. living in same census tract at WIII	0.16	0.11	0.09
Prop. living in same block at WIII	0.08	0.05	0.03
N	1,254	1,385	5,185

Table 1: Descriptive Statistics. Friends, Former Friends and Peers

in the field of information theory and computer science by Vladimir Levenshtein (Levenshtein, 1966) and later adapted to the social sciences (Abbott and Hrycak, 1990; Abbott, 1995). hree basic operations on sequences are used: $\Omega = \{\iota, \delta, \sigma\}$, where ι denotes *insertion* (one state is inserted into the sequence), δ denotes *deletion* (one state is deleted from the sequence) and σ denotes *substitution* (one state is replaced by another state into the sequence). To each of these elementary operations $\omega_k \in \Omega$, a specific cost can be assigned using a cost function $c(\omega) : \Omega \to \mathcal{R}^+$. If K operations must be performed to transform one observed sequence $\mathbf{s}_1$ into another $\mathbf{s}_2$ such that

$$\mathbf{s}_2 = \omega_1 \circ \omega_2 \circ \cdots \circ \omega_K(\mathbf{s}_1) = \omega_.(\mathbf{s}_1),$$

then the transformation cost is defined as $\sum_{j=1}^{K} c(\omega_j)$. The distance between two sequences can thus be defined as the minimum cost, independent of the order of the operators and of transforming one sequence into the other one:

$$\mathcal{D}_s(\mathbf{s}_1, \mathbf{s}_2) = \min_{\omega_.}\{\sum_{j=1}^{K} c(\omega_j) \text{ s.t. } \mathbf{s}_2 = \omega_.(\mathbf{s}_1)\}.$$

The choice of the operations' costs determines the matching procedure and influences the results obtained.

Several other dissimilarity measures have been proposed to compare life course trajectory. An exhaustive review can be found in (Studer and Ritschard, 2016).

We use Optimal Matching Sequence Analysis to derive a measure of life course similarity across all respondents. We obtain a dissimilarity matrix, i.e. a matrix of dimension $N \times N$ containing all pairwise distances calculated with the sequence analysis algorithm.

Following the approach of (Raab et al., 2014), we use the information from the dissimilarity matrix to compute the sequence similarity for each respondent in the *friend module*. To avoid multiple sampling from the same person (respondents have different number of friends and peers), we randomly selected one listed friend, one peer (nominated from the computer generator) and one random respondent from the sample. We then use this information to compare how similar are life

Dyad Type	Close Friends	Peers	Random Assigned
Avg. OM distance	0.801	0.898	0.945
SD OM distance	0.470	0.464	0.455
95% CI	(0.766–0.836)	(0.874–0.920)	(0.923–0.967)
Diff from random	**-15.2%**	**-4.9%**	–
N	708	1,538	1,706

Table 2: Average OM sequence dissimilarity across dyadic type

Dyad Type	Twins	Full Siblings	Half Siblings	Friends
Avg. OM distance	0.721	0.755	0.920	0.801
SD OM distance	0.477	0.42	0.397	0.470
95% CI	(0.685–0.754)	(0.732–0.778)	(0.891– 0.949)	(0.766–0.836)
Diff from random	**-23.7%**	**-20.1%**	**-2.6 %**	**-15.2%**
N	741	1,251	442	708

Table 3: Average OM sequence dissimilarity across dyadic type

course of the selected pairs.

Table 3 and show the average OM distances between friendship dyads. Close friends have on average more similar trajectories than peers and randomly assigned dyads. In particular, their OM distance is 15% smaller than a random dyad, while peers have a reduction of 5% conmpared to random couples of respondents.

Table 3, repeates the analysis for family dyads. In particular, we show the average OM distances of Full Siblings, Half Siblings, and Twins. Twins and Full siblings have life course trajectories more similar than friends, with a average reduction of 24% and 20% in distance compared to random pairs. Both categories have smaller distances than friends. Half siblings, however, seem to have bigger differences than friends, comparable to peers.

Figure 2 summarized the results.

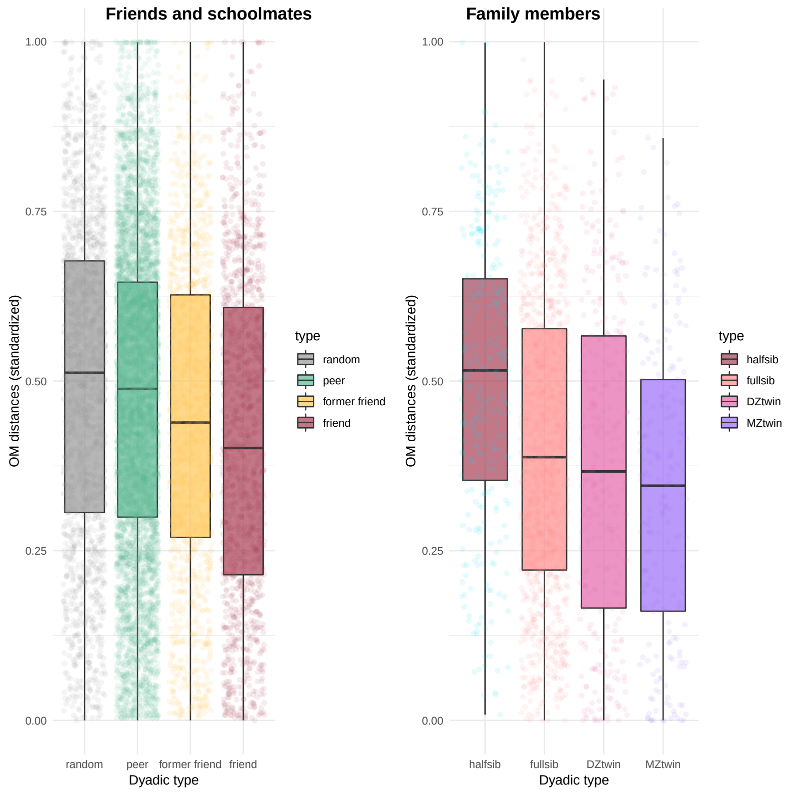

Figure 2: Optimal Matching Distance distribution across dyadic types

3.2.1 Multivariate analysis

We than expand the descriptive analysis by regressing the OM distance on several dyadic characteristics. In particular, we estimated the following equation:

$$D_{ij} = Friends_{ij} + FormerFriends_{ij} + X_{ij} + \mu_i + \epsilon_{ij} \tag{1}$$

Where D_{ij} are OM distance between respondent i and j. $Friends_{ij}$, $FormerFriends_{ij}$, $Peers_{ij}$ are a series of dummy variables indicating the nature of dyadic relationship, where the reference categories are peers. We include a series of dyadic characteristics: the natural logharithm of geographic distance between respondent i and j in Wave I, a dummy indicating same parental education (College vs. no college); a dummy indcating if respondents where raised by both biological parents, a dummy indicating that respondents where both born in the US, and have the same race (White non Hispanic vs. Other). We then included a seriues of dummies indicating the respondents' education achieved. Finally all regression include fixed effects μ_i, as the same respondent may be included in multiple dyads.

Results in Table 4 indicate that Friends, are more similar then peers, even after controlling for a series of dyadic characteristics and individual fixed effects. Education seems to play an effect on life course distances. On average, life trajectories of high educated respondents are more similar. This may reflect the age-structure of education and the fact that individuals who graduate from college delay family transition after finishing education, thus reducing variability in life course trajectories.

The complete distribution of OM distances by education is depicted in Figure 3. Both Peers and Friends with college education have more similar life course trajectories. However, most of the differences between peers and friends seem to arise among college educated respondents, while there is no substantial differences among non-college educated respondents.

	(1) Model 1	(2) Model 2
Dyadic Type: Ref= Peer		
Close Friend	-0.0749***	-0.0562***
	(0.0143)	(0.0145)
Former Friend	-0.0180	-0.0120
	(0.0132)	(0.0143)
Dyadic characteristics		
Log distance Wave I		0.00277
		(0.00284)
Same Parental Education		-0.0209*
		(0.0107)
Same Family Structure Wave I		-0.0294***
		(0.0105)
Both Born in US		-0.0332***
		(0.0119)
Same Race		0.000229
		(0.0137)
Education		
Both college educated		-0.392***
		(0.0197)
One College educated		-0.0829***
		(0.0125)
Constant	0.904***	1.042***
	(0.00564)	(0.0165)
Observations	7,824	7,824
R-squared	0.464	0.451
Individual Fixed effects	YES	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: OLS regression Optimal matching distances by dyadic characteristics

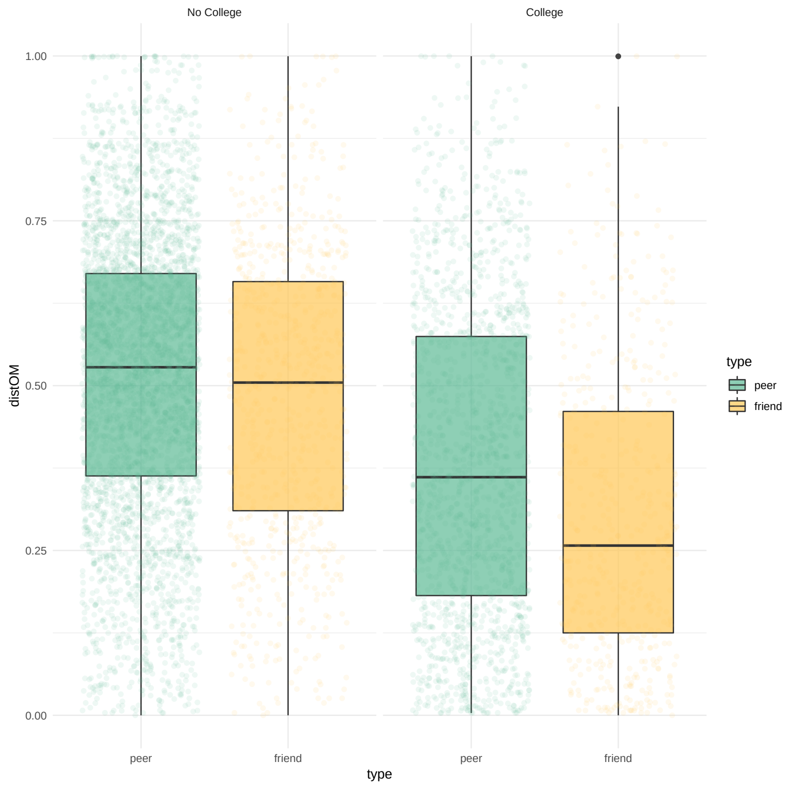

Figure 3: Optimal Matching Distance distribution across dyadic types

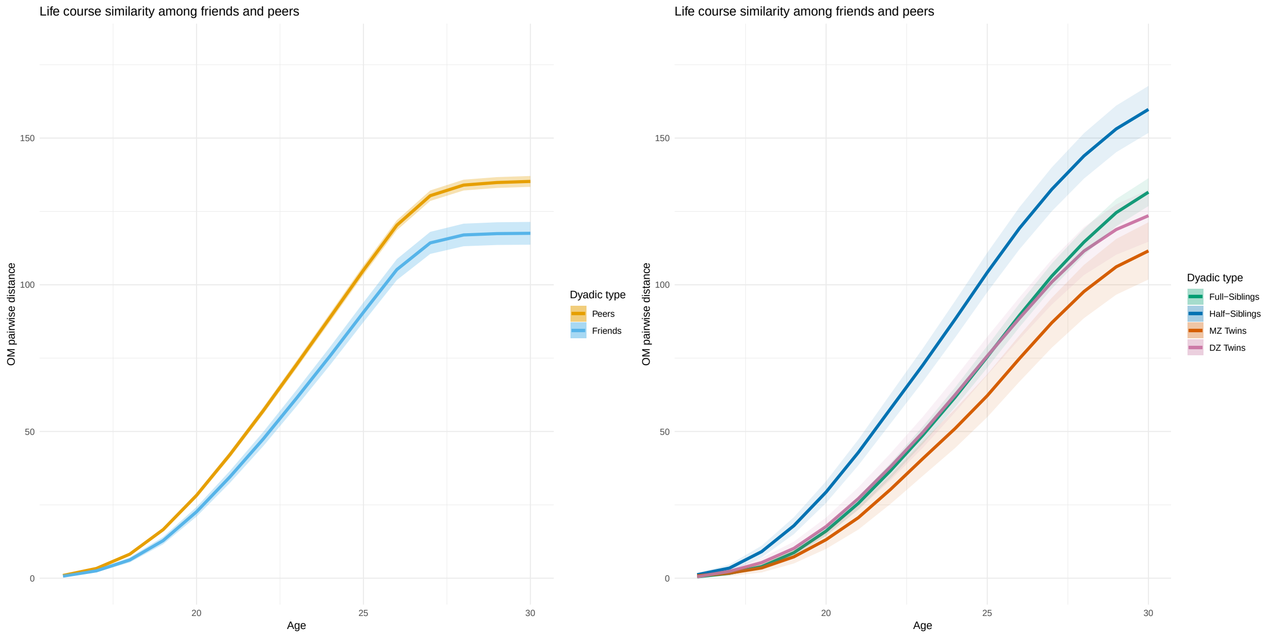

Figure 4: Optimal Matching Distance distribution across dyadic types

4 Life course dissimilarities during the life course

When do differences among friends and family become more relevant? To answer this question, we recalculated the OM distances at different age intervals, starting from age 15 until age 30, using a series of annual steps.

Figure 4 shows how differences in trajectories increase over time. From age 15 to age 20, trajectories are quite similar and and the increase in difference is relatively small. Between age 20 and 25, there is a steep increase in differences across individuals. After age 25 distances between life trajectories seem to stabilize.

5 Friends influences in life course events

5.1 Matching life course sequences

In the second part of the paper, we attempt to assess the causal effect of friends' events. In particular, we analyze the effect of having a high school friends experiencing childbearing, marriage and cohabitation for the first time. Respondents have multiple friends and their effect could be cumulative. For this preliminary analysis we restrict our interest to the first friend who experienced the event. In other words, respondents become at risk of childbearing (marriage; cohabitation) once a listed friend experience the same event. Using the terminology common in the literature of causal inference, a respondent who has a friends experiencing the event before her, is called a *treated individual*. On the other hand, respondents who do not have any friends who experienced the events or those who experienced the events before their friends are called *controls*. Treatment time depends therefore on both respondents' and friends' life course trajectories.

Our analytical approach can be described as follows:

1. Starting from dyadic data (27,803 friendship pairs), we identify for each respondent i the earliest time when a friend j experienced childbearing (marriage, cohabitation). For example, if a respondent has two friends who had a child when she was 19 and 21, she become at risk of "friend influence" starting at age 19. In other words, her treatment time is age 19.

2. We identify all the controls, i.e. all respondents that cannot be influenced by friends because they had a child (marriage, cohabitation) before their friends

3. We calculate the *propensity score* of being a treated individual by estimating a *probit* regression.

4. We identify the *pre-treatment sequence*, that is the sub-sequence representing the life-course trajectory antecedent treatment time.

5. For each treated, we compare the pre-treatment sequence with all the controls by calculating Optimal Matching sequence dissimilarities.

6. Similarly, we compare the propensity scores by calculating the distances in propensity scores.

7. For each treated, we select the control who had the minimum combined distance (OM dissimilarity + propensity score)

8. Finally, we compare the proportion of individuals with children (married, cohabiting) at time $t+1, t+2, \ldots, t+10$.

5.2 Sequence analysis Matching

We adopt the approach described in Barban et al. (2017) to match respondents who share similar pre-treatment sequence. We use Optimal Matching dissimilarity matrix with transition costs derived by the inverse of transition probabilities, i.e. transitions that are less frequent have higher costs in the Sequence analysis algorithm.

5.2.1 Example

Respondent i is 23 when the first high school friend (listed in the friend module) has a child. She has no children and she started a cohabitation when she was 20. At age 22 she married. We know from extensive literature on family demography that her "risk" of becoming a mother is sensibly higher than a single individual. Therefore, it would be misleading to compare this respondent with someone who is single. Ideally we want to "search" in the data if there is any other respondent who had the identical (or the most similar) life-course. In addition, we know that i was born in 1982, she is Hispanic White, her family had income above the median, her parents are High School educated and so on. By combining sequence matching to propensity score, we make sure that the comparison between treated and controls takes into account all these characteristics.

5.3 Propensity score Matching

We base our propensity score matching on the following variables, measured before year t corresponding to the year of retirement for the treatment group:

1. Year of Birth
2. Household income at Wave I (Dummy variable: 1 above median; 0 below median income)
3. Parental education (less than high school; HS or equivalent;some college;college or more;unknown)
4. Family type at Wave I (Dummy variable: 1 Living with both biological parents; 0 otherwise)
5. Race/Ethnicity (White Hispanic; Black; White Asian; White Non-Hispanic)
6. Born in the US (Dummy variable)
7. State fixed effect

Propensity scores and sequence analysis distances are normalized and than summed to obtain a unique measure of similarity among controls and treated respondents. In this particular case, we assigned equal weights to the two methods and we selected only one control as matched respondent. However as discussed in

6 Preliminary Results

Once obtained the matched controls, we investigate the difference in proportion of parents after the occurrence of a friend's childbearing. Figure 5 shows the difference in the proportion of parents after the friend's event before and after matching. The graph shows a higher proportion of respondents who became mothers after their friends (*treated*) compared to the rest of the sample (*controls*). The difference is highly significant and would induce to assign a direct "friend"effect on the basis of these results. However, once we apply our matching procedure based on pre-treatment sequence and time-fixed propensity score, the "friend" effect seems to disappear completely.

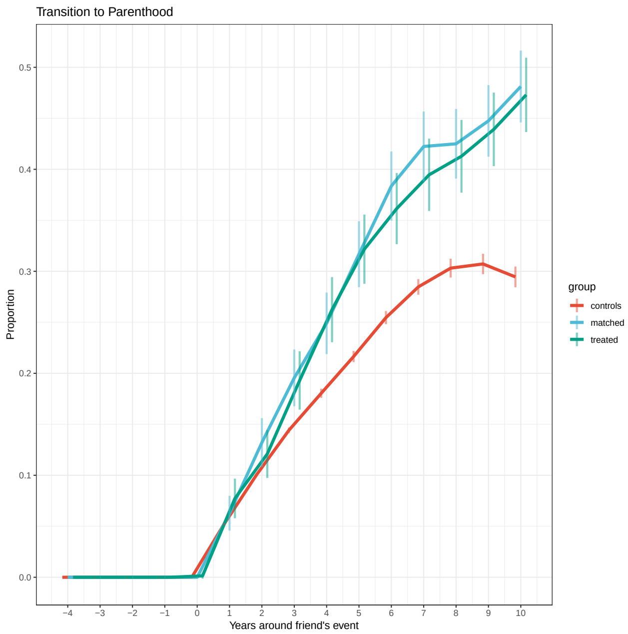

Figure 5: Proportion of Mothers after friends' childbearing

Figure 6 reports the same analysis for marriage and cohabitation. The results is similar for all the outcome with the possible exception of cohabitation, where a "negative friend effect" seems to be present in the years immediately after friends' cohabitation.

Figure 7 looks at the social interaction effect of having children out of the wedlock. Individuals with friends who are unmarried mothers have higher probability to have children while unmarried. However, after controlling for *pre-treatment* trrajectories and time-fixed charateristics, this difference seems to disappear.

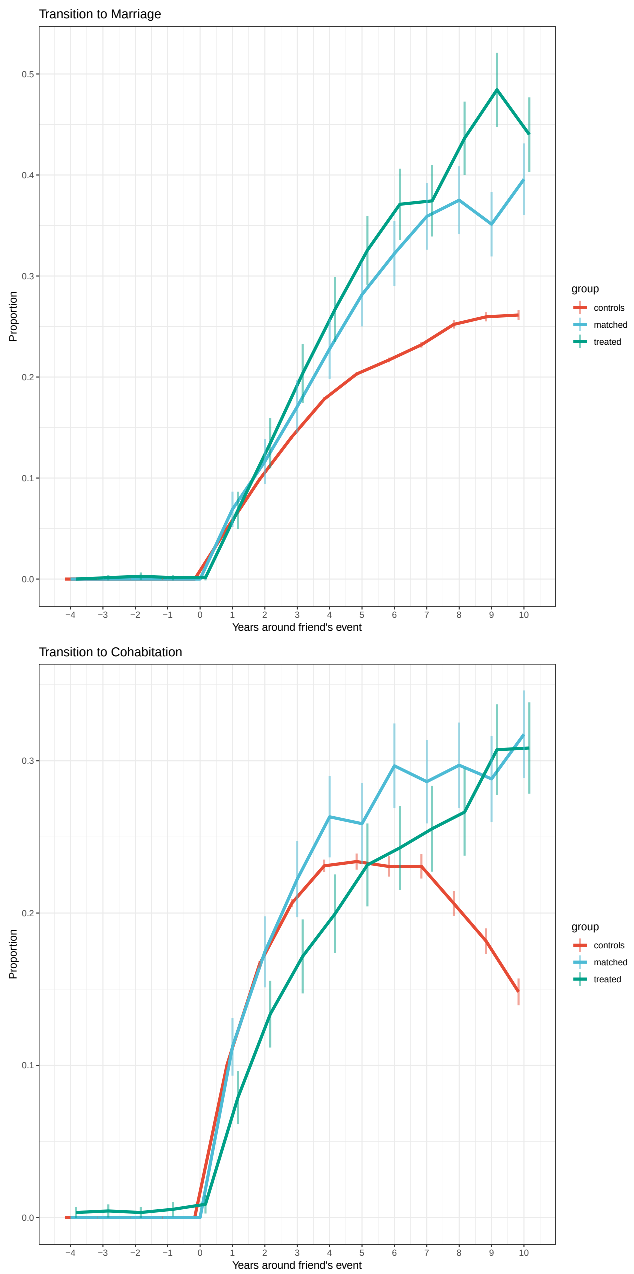

Figure 6: Proportion of married and cohabitiong men after friends' marriage

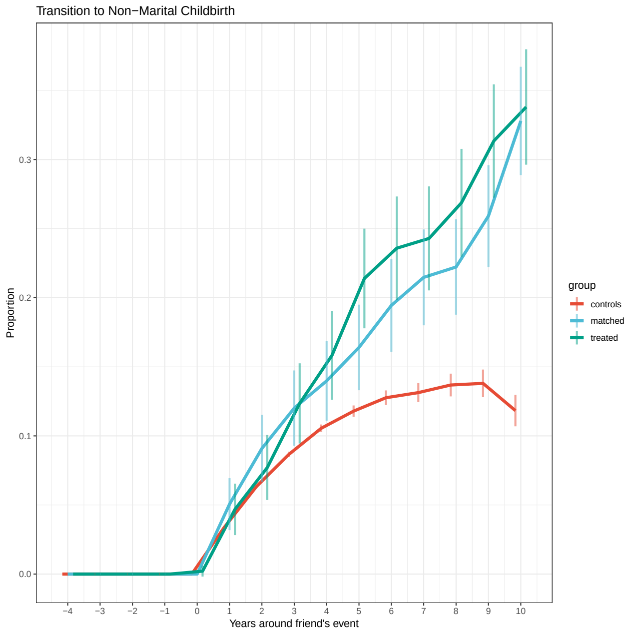

Figure 7: Proportion of Unmarried Mothers after friends' child out of wedlock

7 Discussion

Friends play an important role in the transition to adulthood. Friends have similar life course transitions, as they progress trough adulthood. Understanding the degree of interconnection of life course decisions such as entering a union or having a child, is important to understand how policies can be implemented during this age. Our work shed a light on the similarity of life-course transition using a unique representativ dataset on US women as they progress from adolescence into young adulthood. As noted elsewhere, this is a period demographically very dense where many life-changing events happen. The timing, the order of these events has strong links with socio-economic opportunities, later well-being and health in later life. Keenan and Grundy (2018)

The first part of the paper is dedicated to analyze if there is significant similarity among friends and other young adults. Friends are important relationships that are chosen by individuals. Therefore, friends often share similar background and similarity, i.e. homogamy and are affected by the similar influences such as same school, neighbourhood etc. We use a distinctive feature of the Add-Health questionnaire to distinguish between friends that are in contact and may exert an influence on life course decisions and former school mates (peers) that have been exposed to similar background but may not have a direct "friend" effect on life course. We use the methodology presented by (Barban et al., 2017), and in particular sequence analysis to represent and measure life course trajectories in a holistic manner. Rather than focusing on single life course events, we take the entire trajectory as a whole. The analysis shows that friends are more similar than peers in their life trajectories. The timing, the occurrence of life course events is more similar than peers or than random individuals.

We then, extend the analysis to a sub-sample of siblings and step-siblings in the survey. We find that siblings have more in common than friends and generally have higher life course similarity. However, this is not true for step-siblings, as they lay somewhere between friends and peers (school-mates).

In the second part of the paper, we attempt at measuring a direct effect of friends in three life course events (childbearing, marriage and cohabitation). A first analysis, shows that when a friend experience first one of these life-changing event, her friends have higher probability of experiencing as well. However, life course events are interrelate and often the results of multiple decisions that occurred in the past. To control for each individual trajectory, we adopted a recent method that combines the literature in sequence analysis and causal inference to identify the possible causal effect of life course events. We designed a matching framework where individuals are matched on time-fixed covariates (propensity score matching) and life course trajectories (sequence analysis matching). Our preliminary results show that all the "friend" effect is masked by similar life course trajectories and other characteristics.

We aim to extend this work by looking more in dept at the variety of life course events (e.g. single parenthood, compared to parenthood in a stable union) and by stratifying the results by event at different age (i.e. teenage pregnancy). In addition, the current analysis does not take into consideration socio-economic stratification. Friends from high school may be more relevant to individuals who did not go to college and did not move far from their parents. We aim to extend the analysis by including appropriate measures of socio-economic background to identify the possible differential effect of friendship across educational groups.

References

Abbott, Andrew. 1995. "Sequence analysis: new methods for old ideas." *Annual Reviews in Sociology* 21:93–113.

Abbott, A and A Hrycak. 1990. "Measuring Resemblance in Sequence Data: An Optimal Matching Analysis of Musicians' Careers." *American Journal of Sociology* 96:144–185.

Balbo, Nicoletta and Nicola Barban. 2014. "Does Fertility Behavior Spread among Friends?" *American Sociological Review* 79:412–431.

Barban, N. 2013. "Family trajectories and health: A life course perspective." *European Journal of Population/Revue européenne de* .

Barban, N, X de Luna, E Lundholm, I Svensson, and F C Billari. 2017. "Causal Effects of the Timing of Life-course Events: Age at Retirement and Subsequent Health." *Sociol. Methods Res.* 0:0049124117729697.

Elder, Glen H. Jr.;. 1985. "Perspectives on the life course." *Life course dynamics: Trajectories and transitions* 1:23–49.

Keenan, Katherine and Emily Grundy. 2018. "Fertility History and Physical and Mental Health Changes in European Older Adults." *Eur. J. Popul.* .

Lesthaeghe, Ron and Dirk J Van de Kaa. 1986. "Two demographic transitions?" *DJ van de Kaa y Ronald Lesthaeghe (coords.), Population Growth and Decline, Deventer, Van Loghum* pp. 9–24.

Levenshtein, VI. 1966. "Binary Codes Capable of Correcting Deletions, Insertions and Reversals." *Soviet Physics Doklady* 10:707.

Raab, Marcel, Anette Eva Fasang, Aleksi Karhula, and Jani Erola. 2014. "Sibling Similarity in Family Formation." *Demography* 51:2127–2154.

Studer, Matthias and Gilbert Ritschard. 2016. "What matters in differences between life trajectories: A comparative review of sequence dissimilarity measures." *Journal of the Royal Statistical Society. Series A: Statistics in Society* 179:481–511.