Learning About Social Networks from Mobile Money Transfers*

Adam Sheridan†

Abstract

The increasing popularity of mobile money transfer apps is generating population-scale data on real-world social ties through the social exchange of money. I show how data from these apps provides researchers with an opportunity to better comprehend social networks and their role in social and economic behavior. To do this, I construct a social network from the near universe of person-to-person mobile money transfers in Denmark, based on records from a dominant app used by 80% of the population. Exploiting complete data on socio-economic indicators, family structure and institutional attachments from government registers, I detail the network's structural properties and their striking accordance with those of other large-scale social networks, like Facebook. I also provide novel insights on the extent of segregation and integration in social networks according to economic status and country-of-origin. To show how these data can be used to understand causal social influences on behaviours, I link the network to income, balance sheet and bank account transaction data and explore the effects of income shocks to friends on individuals' spending decisions. Individuals exposed to a friend losing their job cut back their expenditure in the same month as their displaced friend and their spending remains depressed over the estimation window. The reduction in spending includes social expenditures, such as food away from home and travel, and personal durable expenditures, including home improvements, home furnishings and purchases at department stores. The timing and composition of spending reductions suggest that friends' adverse experiences may affect individuals' perceptions of their own economic security.

*Acknowledgments: I thank Asger Lau Andersen, Amalie Sofie Jensen, Niels Johannesen, David Dreyer Lassen, Andreas Bjerre Nielsen, Tue Lehn-Schiøler, Benjamin Ly Serena, and Johannes Stroebel for helpful comments and suggestions, along with participants at SODAS Data Sessions and NABE TEC2019. The activities of the Center for Economic Behavior and Inequality (CEBI) are financed by a grant from the Danish National Research Foundation. This research was facilitated by Adam Sheridan's Industrial PhD project, jointly financed by Danske Bank, which provided data access, and Innovation Fund Denmark. More details on Innovation Fund Denmark's Industrial PhD programme are available at `https://innovationsfonden.dk/en/programmes/industrial-researcher`. Support from Danske Bank and Innovation Fund Denmark is gratefully acknowledged. Danske Bank did not review the conclusions of this paper before circulation and the opinions expressed are those of the author alone and do not represent the views of Danske Bank. Support from the Danish Finance Institute (DFI) is gratefully acknowledged.

†University of Copenhagen, CEBI, and DFI. Email: adam.sheridan@econ.ku.dk

1 Introduction

A large body of academic work is concerned with understanding humans foremost as social beings. Central to this approach, researchers have sought to understand the structure of networks formed by social ties and the pervasive importance of interactions through social networks on social and econonomic behavior. These investigations require data on real world social networks that reflect their diverse and dynamic nature and such data are difficult and costly to obtain. Consequently, researchers have, in the main, relied on self-reported data on social networks that are limited in scale and accuracy. This reliance has been a barrier to scientific progress.

In this paper, I examine an alternative approach to measuring real world, population-scale social networks. I base my approach on the observation that the exchange of money can be highly social (Zelizer, 1996). Asking who a household would turn to when they need a small amount of funds, or who they would be willing to give money to, is a common way to map social networks (e.g., Banerjee et al., 2013). Furthermore, although shared experiences with friends may feel priceless, they rarely are, and money changes hands to settle shared costs, from a cup of coffee to a weekend away.

Innovative solutions have emerged that easily facilitate social money transfers using mobile phones. Examples include the application (app) Venmo in the U.S. and Facebook Messenger's payment option, both of which are explicit in their social focus, and SMS-based M-PESA, originally from Kenya and Tanzania. These solutions are increasingly popular, transitioning the social exchange of money from cash to electronic transfer. Records of these exchanges leave a digital trace of social ties and interactions, from which researchers can learn about social networks.

I take this approach to learning about social networks to the data. I construct a population-scale social network from the near universe of person-to-person mobile money transfers in Denmark. I document the structural properties of social interactions in this network, their accordance with other known social networks and their discordance with networks of purely transactional relationships. Further, I link population-wide individual-level data to the network, including information on family ties and institutional attachments, past and present. Using this novel data set, I investigate social interactions and their influence in three separate parts, each analyzing an important network phenomenon and, in combination, highlighting the value of data from mobile money transfer apps for learning about social networks. First, I quantify the evolving importance of family and institutional ties in individuals' social circles over the life cycle. Second, I analyze how individual attributes shape interaction patterns in two ways: exploring immigrant networks and the correlates of segregation and integration, and measuring network stratification by economic status. Third, I turn from studying the determinants of social interactions to their effects on behavior, estimating the contagious effects of job loss shocks on expenditure.

Specifically, I harness data from MobilePay (`https://mobilepay.dk`), a Danish mobile money transfer app that launched in May 2013. Like other mobile money apps in developed countries, MobilePay's launch mission was to make money transfers among family and friends easier, and payments between private individuals for goods or services are prohibited.[1] Three

[1]MobilePay has, since launch, introduced a separate functionality for making payments at stores and to

features of the Danish setting are unique. First, smartphone penetration is very high and electronic payments have been dominant for many years. Second, the app is a near monopoly in the Danish market for mobile money transfers. Combined, these two conditions meant that app adoption and usage was rapid and widespread: as of December 2018, over 80 percent of the population aged 13 and over were users, and my data contain the full history of transfers using the app from its launch until December 2016. Third, I am able to combine the data with several government and private administrative registers. These registers contain complete family, workplace, address and education histories for the population, as well as high-frequency data on a host of socio-economic indicators. The combined network data are unprecedented in capturing ties across multiple population-scale networks - the social network based on mobile money transfers and institutional networks - and for the richness of the linked individual data.

I begin by documenting the structure of the social network derived from mobile money transfers. I transform the raw transfer data, containing 450 million person-to-person transfers between May 2013 and December 2016, into a network representation that captures how people interact. I connect two individuals with an undirected link if a transfer has occurred between them and I preserve three features of their transfer history as measures of tie strength, or relationship intensity: the number of transfers they have engaged in, the Danish kroner (DKK) value of their exchanges, and whether the relationship is reciprocal, that is, whether money flowed in both directions. The resulting network contains 3.1 million individuals and 63 million links. I show that the structural properties of the network are strikingly similar to other large-scale social networks (e.g., the Facebook network as studied by Bailey et al. 2018; Ugander et al. 2011): the distribution of "popularity" has fat tails, friends-of-friends are often friends themselves, and people tend to befriend others who are similarly, or more, popular. This last point, a positive correlation in number of connections between linked individuals, is a characteristic that is distinct to social networks (Jackson and Rogers, 2007) and rules out the possibility that the network measures purely transactional relationships.[2]

In the remainder of the paper, I turn to exploring what these data reveal about interactions in social networks and their effects on behavior. I organize this section in three parts, each presenting an analysis of an important social network phenomenon.

In the first part, I study family and institutional ties. An individual's social circle, measured here as the people they link to through mobile money transfers, reflects their choices over who to interact with and constraints imposed by geography and social location. Institutional attachments (chiefly schools, universities, and employers) and family are important determinants of social location and a large sociological literature has explored their role as a key constraint in social network formation (e.g., Kossinets and Watts, 2006, 2009; Fischer and Oliker, 1983). A motivation for this literature is that access to the social network available at certain institutions can convey benefits, and these networks and their benefits might persist, entrenching privilege among those who get in. These mechanisms interact with a range of social and economic policies, such as those targeting school choice (e.g., Rothstein, 2006). Distinguishing the role of

merchants. I do not consider these transfers in this analysis.

[2]The finding that monetary exchange can be highly social in nature supports Zelizer's (1996; 1994) assertion that money can have a social meaning, rather than solely operating as a means to remove the social element from interactions (Weber, 1922; Simmel, 1907).

choice versus constraints in social network formation is methodologically challenging, but even before approaching this challenge there is an absence of large-scale evidence on the importance of family and institutional attachments in individuals' social circles (McPherson, Smith-Lovin, and Cook, 2001).

I overcome this gap in knowledge by quantifying the importance of institutions in the social circles of individuals in the network derived from mobile money transfers. I focus on individuals who were born in Denmark since 1975. This sample consists of 1.48 million individuals with 49.96 million links in the network. For these individuals, I can construct complete immediate and extended family networks, and lifetime networks of institutional ties through schools, universities, employers, and shared addresses. My first key finding is that family and institutional ties - the people an individual is connected to through the family and institutional networks - are hugely important, constituting almost 60 percent of the average individual's social circle. Moreover, this percentage rises to almost 75 percent when accounting for tie strength, with family ties becoming much more significant when tie strength is measured as the DKK value of transfers, likely reflecting family's important insurance role (Andersen, Johannesen, and Sheridan, 2019). My second key finding is that there appears to be persistence in institutional ties, with educational ties continuing to account for 10 percent of the average individual's social circle throughout the 30s. However, the majority institutional group in the average individual's social circle reflects their current, or recent, institutional attachments. The observed importance and persistence of institutional ties relates to the literature on assignment mechanisms to allocate institutional access and the possibly significant redistributive effects such mechanisms could bring about (Abdulkadiroğlu, Agarwal, and Pathak, 2017; Fryer Jr. and Loury, 2013).

In the second part, I turn to analyzing the role of individual attributes, immigrant status and, separately, economic status, in shaping interaction patterns in the network. Starting with immigrant status, homophily - the tendency of individuals to associate with similar others - has been identified as ubiquitous in social networks, especially so with respect to race and ethnicity (Jackson, Rogers, and Zenou, 2017; McPherson, Smith-Lovin, and Cook, 2001). Homophily has an important opposite when discussing immigrants: social integration. The extent of immigrant integration is a major policy issue and an important research agenda is designing integration measures. Typically, as in studies of homophily, these measures are based on surveys (e.g., Harder et al., 2018).

I contribute with the first population-scale analysis of the individual and economic correlates of homophily and integration among immigrants that is based on observational social network data. Approximately 12 percent of the Danish population are immigrants or the descendants of immigrants ("non-natives"), and this group are well represented among app users. I measure individual non-natives' share of their non-family social circle that consists of individuals who are also non-native, who have the same country of origin, and who have the same global region of origin. I scale these shares by the population representation of each "group" (non-natives, country of origin, region of origin) in a measure of excess, or inbreeding, homophily (Currarini, Jackson, and Pin, 2009, 2010). I also calculate a measure of integration as the share of natives (Danes) in non-natives' social circles.

I find considerable variation in excess homophily and integration across non-natives and by group definition. I show that this variation has two important correlates: geography of origin (country and region), and individual age on arrival to Denmark. With respect to the latter, excess homophily is relatively flat as arrival age increases up to the mandatory school starting age and then increases past this point. With respect to the former, there is large variance in homophily and integration across country-of-origin and these differences correlate strongly with economic integration, measured by average annual incomes.[3] Moreover, non-natives form community structures in the network that correlate with linguistic, geographic and, perhaps, cultural features of their country of origin. For example, individuals of Spanish and Spanish-speaking Latin-American origin form a community. These results contribute to research on constraints to integration based on arrival age (Bleakley and Chin, 2010) with suggestive evidence that access to schooling is important, and the literature studying the joint determinants of immigrants' networks and their economic success (Dustmann et al., 2016).

Turning to economic status, a nascent literature is examining the role of social networks in shaping inequality and social mobility (see Jackson, Rogers, and Zenou, 2017; DiMaggio and Garip, 2012, for references). A central question is the extent to which individuals of different economic status interact socially, a dimension of homophily. A related but more established literature has asked similar questions concerning *intergenerational* correlations in economic status (Chetty et al., 2014; Chetty and Hendren, 2018). I turn this literature's empirical methodology to estimating the joint distribution of individual income, a measure of economic status, and average income among individuals' social circles, a measure of the economic status of one's friends. I find a strong correlation in percentile ranks: moving 10 percentile ranks up in the individual income distribution is associated with moving 5 percentile ranks up in the distribution of average income among individuals' social circles on average, and 7 percentile ranks at the median. Moreover, I find that the joint distribution has less variance at the top of the individual income distribution, suggesting more concentrated networks among the rich. Thus, I contribute to a nascent literature by quantifying stratification by income in a population-scale social network, shedding light on one possible channel through which inequality propagates.

In the third, and final, part, I investigate social influence on spending decisions. The things that people buy are often highly visible, consumption experiences are frequently shared experiences, and people base their spending decisions on expectations and information about the future of the economy. These features make it likely that individual consumption decisions are subject to influence from others, with consequences for the macroeconomy. Motivated by this, I estimate the contagious effects of job loss shocks on spending decisions through interactions in social networks. I combine the network with high-frequency data on spending, income and job loss events. I derive the spending data from the complete transaction records of a major Danish retail bank, building a spending measured based on ATM cash withdrawals, bill and card payments (like, e.g., Baker, 2018). The income and job loss data are from population-wide monthly payroll and social transfer records. I use an event study approach to estimate the path of (household) spending around the month that a nonfamily network connection experiences a job loss shock, ruling out non-causal explanations through careful choice of the peer comparisons

[3]This correlation is useful for researchers studying immigrant integration who have access to income data but not social network data.

in the estimation and by presenting visual evidence on the precise timing of joint responses.

I find clear evidence of contagion: households reduce their monthly nondurable spending by around 12.5 USD (83 DKK) in the 6 months following a peer's job loss relative to their spending before the event. This is equivalent to a drop of 0.4 percent of average monthly spending prior to the network connection's job loss shock. The dynamics of the response suggest that the mechanism is through expectations: on losing their job, individuals tell their friends, these friends become more concerned about their own job security (Alt et al., 2017) and they cut back on spending. A recent literature has identified social influence on consumption, driven by status concerns (De Giorgi, Frederiksen, and Pistaferri, 2018) and interactions through networks affecting expectations (Bailey et al., 2018), and the implications for social insurance and macroeconomic stabilization policies. My findings contribute to this literature with the first event-study estimates of the contagious effects of job loss shocks on total household expenditure using observational social network data.

Overall, my paper relates to the literature on the measurement of social networks (Breza et al., 2017; Chandrasekhar and Lewis, 2016; Marsden, 1990) and a series of recent papers in computer science and economics using new sources of large-scale social network data (e.g., Onnela et al., 2007; Eagle, Pentland, and Lazer, 2009; Bailey et al., 2018). Of this second group, my paper is most similar in terms of approach to the study by Bailey et al. (2018), who explore the use of Facebook data to learn about the economic effects of social networks. A key difference is that their study is based on a data set that is explicitly recorded as a virtual social network, whereas I derive a social network from implicitly social real world interactions. In terms of data, I relate most closely to the study by Blumenstock, Eagle, and Fafchamps (2016), who measure networks using airtime credit transfer through mobile phones, and Jack, Ray, and Suri (2013), who survey users of the M-PESA mobile money solution in Kenya in order to learn about informal credit. As noted by Suri (2017), mobile money is widely used in many developing countries but there has been no research using transfer-level data to learn about social networks. Although a different setting, the evidence I present here - especially on the consequences of different tie strength measures - and the methodologies I apply can be useful for future researchers who gain access to these data in developing countries.

My main contribution to this literature is in highlighting the value of naturally occurring data on social ties and interactions from mobile money transfer apps for furthering research on social networks. Combining the insights from my investigation of the properties of the network and each of three analyses I perform, I identify 5 advantages of data from mobile money transfer apps for measuring social networks. These advantages go beyond the general benefits associated with using observational, rather than self-reported, social networks data (see Marsden, 1990, for a discussion of the biases and errors in survey data). Moreover, they demonstrate that, while being a complement to data from social networking sites (e.g., Ugander et al., 2011; Bailey et al., 2018) or communication records (e.g., Onnela et al., 2007; Eagle, Pentland, and Lazer, 2009) for understanding social networks, data from mobile money transfer apps have a number of advantageous features.

First, because of increasing popularity of mobile money transfer apps and network externalities in adoption that lead to one or few service providers, networks derived from mobile money transfers are measurable at population-scale. Population-scale social network data can challenge received wisdom about social networks based on smaller-scale data (Park, Blumenstock, and Macy, 2018). Second,networks derived from mobile money transfers are real world, offline networks capturing multiple types of relationships, unlike, for example, the virtual network measured by Facebook. The pursuit of network measures based on physical closeness in time and space (Crandall et al., 2010) highlights the importance attached to understanding real-world connections.

Third, networks derived from mobile money transfers are unconstrained: within app users, there are no restrictions on who one can link with or how many links one can have. Constraints on measured networks - like top coding, present even in Facebook data – can affect network inference (Chandrasekhar and Lewis, 2016). Fourth, networks derived from mobile money transfers are dynamic: social networks evolve constantly and monitoring this evolution is critical for empirical networks research (Choi, Kariv, and Gallo, 2016). Finally, networks derived from mobile money transfers have simple, meaningful measures of relationship intensity embedded. Again, Facebook is a useful example. Despite evidence that people interact with very few of their Facebook "Friends" (Wilson et al., 2009) and that interaction is highly predictive of real world friendship (Jones et al., 2013), interaction measures are relatively complex and many studies do not include them in their analysis (e.g., Bailey et al., 2018; Ugander et al., 2011). I show how simple measures of tie strength in mobile money transfers networks have clear implications for the types of ties considered. These insights are useful for future research continuing work on the importance of tie strength for social processes (e.g., Granovetter, 1973; Park, Blumenstock, and Macy, 2018) and for researchers working with similar, naturally occurring network data.

The rest of the paper proceed as follows. In Section 2, I discuss the rising popularity of mobile money transfer apps and their use for social payments. I introduce the data I use here, and I document the structural properties of the measured social network. In Sections 3, 4, and 5, I take each of the three network phenomenon I study in turn, presenting my empirical approach and results. In Section 6, I conclude with a discussion of some limitations of these data and, based on these, directions for future research.

2 Measuring Social Networks Using Data from Mobile Money Transfer Apps

2.1 Context: Mobile Money Transfer Apps and Social Payments

In the U.S., and much of Europe and Asia, mobile money transfer services, also referred to as mobile payment and mobile money, typically take the form of applications (apps) downloaded onto smartphones. Users link their payment card to the app and make person-to-person money transfers to other app users, identifying them by phone number, QR code or user ID. Most apps also facilitate payments to merchants but this is through a separate functionality and payment flow, making them easily detectable. I focus only on person-to-person transfer services.[4]

Mobile money transfer apps are increasingly popular. PayPal's Venmo app in the U.S.

[4]As mentioned, mobile money transfers are hugely popular in many developing countries. Suri (2017) provides a review of mobile money and its impact in developing countries.

demonstrates this popularity: in July 2017, the app had over 10 million unique monthly users (Hwong, 2017) and over 17 million users in total by the end of the year (Perez, 2018). In China, Tencent's WeChat Pay, a service integrated into the WeChat messenger app, has over 900 million monthly active users (Jacobs, 2018).

An important feature of mobile money transfer apps is that they have an explicit focus on facilitating "social payments": transfers between friends, family and other social acquaintances for the purposes of splitting shared costs like restaurant bills, a leading example, or gifting small amounts. Venmo, for example, has a live social feed, where individuals can see the payments their friends are making to each other in a form of social media (Read, 2017), and Venmo forbids both peer-to-peer transfers between individuals who do not personally know each other and transfers for sale of goods or services.[5]

2.2 Mobile Money Transfers in Denmark: MobilePay

MobilePay, a near monopoly in the Danish market for mobile money transfer apps, shares the same focus as other mobile money transfer apps of facilitating easy social payments. MobilePay was released in Denmark[6] in May 2013, with the purpose of making it "just as easy to transfer money as sending a text message" (Jesper Nielsen, Danske Bank, as cited by Dilling (2013)).

Danske Bank, the largest Danish retail bank, developed MobilePay but customers of any bank have always been able to use the app by registering with their social security (CPR) number and linking their payment card.[7] A consortium of other Danish banks initially offered an alternative mobile money transfer app, but MobilePay's success in building a large user base led to them collaborating with Danske Bank and transitioning users to MobilePay (Sixøj, 2016). The consolidation of the Danish market onto a single provider of person-to-person mobile money transfers highlights that this service is a network good, with a tendency towards a small number of firms offering the service in a given market (see Björkegren, 2018, for a recent discussion of network goods). The network good property of mobile money transfer apps results in the generation of population-scale data.

Person-to-person transfers using MobilePay are easy and quick, usually taking place as an instant bank transfer. In order to make a transfer a user must sign in to the app, enter the amount they wish to transfer, select a recipient from their contacts or enter a phone number, optionally add a message and/or a photo, and swipe to transfer. I present screenshots of the user flow to make a transfer in Figure A.1 in the appendix. Users can send up to 10,000 DKK (1,500 USD) per day and there is no limit on how much a user can receive in a day. It is possible to request money from another user and the app has a group-payment functionality called WeShare through which groups of users can submit shared costs, such as for a weekend

[5]The terms and conditions state (capital text original): "Venmo is designed for payments between friends and people who know and trust one another. [...] DO NOT USE VENMO TO TRANSACT WITH PEOPLE YOU DON'T PERSONALLY KNOW, ESPECIALLY IF THE TRANSACTION INVOLVES THE PURCHASE OR SALE OF A GOOD OR SERVICE [...]." See: `https://help.venmo.com/hc/en-us/articles/217532097-Can-I-use-Venmo-to-buy-or-sell-merchandise-goods-or-services-`. As reported by Miranda (2017), Venmo monitors transfers to enforce this rule.

[6]MobilePay is also the most popular mobile money transfer app in Finland. I focus only on Danish users, for whom I have data.

[7]A CPR number is given to every Danish resident at birth or on arrival to Denmark. Almost every Danish resident has a bank account because it is a requirement for receiving government transfers and salary.

trip, and have balances settled automatically. In order to use MobilePay individuals must be at least 13 years-old. Users are not allowed to use the person-to-person transfer functionality for selling goods or services.

2.3 Data: MobilePay Transfers and Administrative Registers

Mobile Money Transfers

I use data on every person-to-person mobile money transfer using MobilePay in Denmark over 3.5 years, from the launch of the app in May 2013 to December 2016. These data exclude transfers involving a small number of individuals identified as non-compliant merchants, that is, individuals operating as a merchant using the person-to-person transfer system. The final data contain over 450 million transfers.

Each transfer record includes a unique sender identifier and receiver identifier, both of which correspond to Danish social security (CPR) numbers. Moreover, each transfer record includes the amount sent in Danish kroner (DKK) and the date of the transfer.

Constructing a Network from Mobile Money Transfers

I construct a social network from the raw transfers by connecting two app users with an undirected link if a transfer has occurred between them. This transformation compresses all transfers between two individuals i and j into a single network entry recording that there has been a mobile money transfer between i and j. The final network contains 3.1 million individuals and 63 million links between individuals.

For each linked pair, I record a number of features of their transfer history. These features are different ways to measure tie strength or relationship intensity. First, I record the total DKK value of transfers between individual i and individual j, DKK_{ij}. Second, I record the total number of transfers between i and j, NUM_{ij}. I also record whether i sends money to j and also receives money from j, that is, I construct an indicator for whether the tie is reciprocated, REC_{ij}.

As an example, suppose the raw mobile money transfers data contains 5 transfers from individual A to individual B, each of which for a value of 10 DKK, and 1 transfer from B to A of 100 DKK. The entry in the undirected network would contain a link $A - B$ (equivalently,$B - A$, but only one link appears in the network representation). Attributes of this link include $DKK_{AB} = 150$, the total value of transfers between the pair, and $NUM_{AB} = 6$, the total number of transfers between the pair, and $REC_{AB} = 1$, signifying that A sent money to B and B sent money to A.

Combining the Network with Population Administrative Registers

In addition to the population-scale networks data, I also have access to a number of population-wide administrate registers from the Danish statistics agency, Statistics Denmark. In order to link individuals in the network to their records at Statistics Denmark, I sent an encrypted version of the mobile money transfers data containing true social security (CPR) numbers to

Statistics Denmark.[8] Technicians at Statistics Denmark decrypted the data and converted the true CPR numbers into the de-identified personal identifiers used in the administrative data. I have access to a secure server at Statistics Denmark containing the de-identified mobile money transfers data and administrative data and all of my analysis takes place on this secure server.

The administrative registers contain information for every resident of Denmark between 1980 and 2015 at an extraordinary level of detail. For each individual, the data contain a range of socio-demographic information: age, sex, country of origin and date of arrival for immigrants and their descendants, address, family ties including non-marital partners, and full education histories. The administrative registers also include rich information on income, employment and wealth. From these data, I use measures of total annual labor market income, business income, and social transfer income, net wealth, and employer identifiers (CVR numbers).

Mapping Family and Institutional Networks

A central contribution of my paper is that I am the first to observe both observational social network data, the network from mobile money transfers, and family and institutional networks, which I construct from the administrative registers, at such a large scale.

I construct family networks and three types of institutional network - education, workplace, and address – for every individual in the network derived from mobile money transfers. My definition of family networks is broad: I include close family, a full history of partners, any recorded parents of shared children (in the case the parents were never registered as partners), stepfamily, and extended family, including grandparents, cousins, aunts and uncles. In education networks, I include all individuals at the same educational institution in a given year. I construct workplace networks as all individuals working for the same employer, identified by company identifiers (CVR numbers), in the same year. Finally, I build address networks as all individuals living at the same address in a given year who are not partners, capturing roommates.[9] Table 1 contains definitions of each network.

The primary purpose for constructing these networks is to learn about the real world relationships that the links in the network based on mobile money transfers represent. Therefore, the networks I construct are relatively broad, and deliberately so. Comparing my family and institutional networks with those constructed by Alt et al. (2017) is a good way to see this. These authors study contagious unemployment expectations through family, workplace and education networks. In order to maximize the likelihood that the ties they identify represent real world social ties, Alt et al. (2017) put restrictions on the set of ties within families and institutions that they consider. For example, they only consider workplace ties within small organizations, where it is more likely that individuals they assign to the same network are real social ties. In contrast, I construct broader family and institutional networks with the aim of identifying the most likely real world venue wherein a tie in the network derived from mobile money transfers was formed; a more suitable approach for my purposes.

[8] I received approval for the data transfer from the Danish Data Protection Agency and followed strict security protocols throughout the transfer.

[9] Address networks could also capture homosexual couples who are not registered as such.

2.4 Adoption and Representativeness

Representativeness, at the individual and the link level, is a concern when working with this app data: for correct inference, it is important to map as much as possible of the full, latent social network (Chandrasekhar and Lewis, 2016).

Figure 1 shows the number of individuals first using the app and the share of the population who are users in each month since the introduction of the app in May 2013 and until December 2016. I define an individual's first use as the month they first use the app to make a transfer. I measure the share of app users in the population in two ways. First, as a share of all individuals eligible to use the app: everyone over 13 years old. Second, as a share of all individuals aged between 13 and 55 years old.

The evidence in Figure 1 suggests that individuals in the app data are representative of population: app adoption and usage was rapid and quickly very widespread, with almost 80 percent of the population aged 13 to 55 using the app by end of year 2016, and over 60 percent of the eligible population. Although not in the data I use here, app usage has continued to grow, with over 80 percent of the eligible population users by December 2018.[10]

I am able to use the administrative registers to investigate how the user sample compares to the full population. Early app users were more likely to be between 18 and 45 years old, higher educated, concentrated in the capital region, and higher earners than the average Dane. Table 2 compares the age, sex, education, and the geographic, wealth and income distributions in the app user sample and the full population by the end of 2016, 3.5 years after the app was introduced. App users are less likely to be in the 65+ age group but there are still over 250,000 such users in the data; unusual for digital trace data. The app users are more likely to be female, a group that other sources of financial digital trace data typically underrepresent (Baker, 2018; Gelman et al., 2014). The app users are slightly less likely to be immigrants or descendants of immigrants than their representation in the population.[11] Overall, the sample are broadly representative of the Danish population, especially so among the sample of 13-54 year olds (Column 2), reflecting the large share of the population who were app users by the end of 2016.

2.5 Properties of a Population-Scale Social Network

I turn to investigating the structural properties of interactions in the measured social network. I start by summarizing the size of individuals' social circles and interaction patterns. As throughout, I define an individual's *social circle* as the individuals they have a link to in the network. As in the networks literature, I define an individual's *degree* as the number of individuals that

[10]To give some perspective of the enormous scale of this app for measuring social networks, survey based estimates suggest that Facebook is used by 68 percent of the U.S. adult population (Smith and Anderson, 2018). Networks based on mobile money transfers have the potential to measure real world social interactions at a scale as large, if not larger, than Facebook, which is itself a source of network data that does not guarantee social interaction in the real world.

[11]I investigated the representation of Danes and immigrants and their descendants in the data in more detail and found a Spearman's rho correlation coefficient of 0.996 between the population shares of individuals with each country-of-origin (including Denmark) and the app user sample shares. This suggests that sample of immigrants and descendants using the app is representative of the population in terms of the mix of individuals of different country of origin.

a person links to in the network, equivalently the size of their social circle.

Table 3 presents summary statistics on interactions in the network for all users and for the 13-55 year old sample. Focusing on Column 2, the average number of individuals that a person links with in the mobile money transfer network, their degree, is almost 47. Thus, individuals have links with a meaningful number of people and they use the app intensively: the average user is counterparty to over 170 transfers throughout their time using the app, amounting to a DKK value of transfers that is equivalent to 16 percent of one year of gross annual income on average.

Table 3 also provides the first evidence on how patterns of interaction can measure tie strength. In many social settings, stronger ties are reciprocal: information, confiding or services flow in both directions (Granovetter, 1973). I present statistics on the average number of reciprocated links that individuals' have in the network, the average percent of reciprocal links in all links, and the average percent of reciprocal links in all links when weighting by the DKK value transferred and the number of transfers. Reciprocated links are a considerably higher percentage of all links when weighting links than when not, especially so for weighting by the DKK value of transfers between a pair.

Reciprocity is more ambiguous as a stand-alone measure of tie strength when defining links by the exchange of money than in other settings. For example, the parent-child relationship is, often, a strong tie but many children will likely not send money to their parents. Therefore, rather than limiting my analysis to reciprocated links, like in other studies exploring digital trace network data (e.g., Onnela et al., 2007), I preserve unidirectional links and the valuable information about social ties that they might contain. However, to the extent that reciprocity correlates with true tie strength in this setting, the results in Table 3 suggest that the weights I have constructed based on the number and value of transfers are revealing of the intensity of relationships.

In the next step of this analysis, I investigate three structural properties of social networks that have been regarded as "empirical regularities" (Jackson and Rogers, 2007). I investigate each in the full undirected mobile money transfer network, containing 3.1 million individuals and 63 million links.

Degree Distribution

The first structural property I study is the degree distribution, moving beyond the averages reported in Table 3 to study the full distribution of degrees in the network. A network's degree distribution is important because it can influence patterns and speeds of diffusion of information and influence (Jackson, Rogers, and Zenou, 2017). For example, hubs – highly connected individuals – have a unique position in terms of their ability to quickly spread information through the network, or influence large numbers of other individuals, and hence it is critical to understand their relative presence in social networks.

I plot the degree distribution of the network on a log-log scale in Panel A of Figure 2. I use logarithmic binning in order to emphasize the shape of the distribution up to very high degrees; the details of this procedure are in the figure footnotes. As in many social networks (Jackson and Rogers, 2007; Barabási, 2016), the degree distribution is fat tailed: the variation in degree across individuals is considerable, with a mass of individuals with low degrees and some individuals with very high degrees. The median individual has a degree of around 30, the 95th percentile of the degree distribution is around 110, and there are some individuals with degrees of over 1,000.[12]

Degree Correlation

The second structural property I consider is the extent to which low degree individuals connect to other individuals with low degrees, and likewise for high degree individuals. The degree correlation, often referred to as assortativity, has been widely studied because, like the degree distribution, it can influence the speed of contagion: if high degree individuals link to other high degree individuals then information or influence can travel faster, relative to when such links are absent (Jackson and Rogers, 2007).

I summarize degree correlations in the network in Panel B of Figure 2. I find a positive relationship between the degree of an individual and the average degree of the people they link with. For individuals with only 1 connection, the average degree of their connections is just above the average degree in the whole network. For individuals with 100 connections, the average degree of their connections is approximately the same as themselves; almost 2.5 times the average degree in the whole network and close to the 95th percentile of the degree distribution.

It is important to note that the positive degree correlation that I document in the network is a structural property that is found only in social networks (Jackson and Rogers, 2007; Newman and Park, 2003). If networks formed from mobile money transfers recorded trade between individuals, such as the sale and purchase of goods and services, then we would expect a negative degree correlation in line with empirical evidence on trade networks (e.g., Bernard, Moxnes, and Ulttveit-Moe, 2018).[13] Thus, the positive degree correlation is accordant with social networks and discordant with a network of transactional interactions, providing evidence for the widespread use of mobile money transfer apps for social payments among family, friends and acquaintances.

Clustering

The final structural property that I explore is the extent of clustering. Clustering refers to the likelihood that A and C are connected to each other if there is a link between A and B and between B and C. In other words, clustering is the density of triangles in a network. Again, this property has been widely studied because it can influence diffusion through networks and information exchange.

In order to investigate clustering in the network, I calculate the clustering coefficient for each individual. For a given individual, the clustering coefficient is the proportion of all possible pairs

[12] I am not able to report exact numbers for the percentiles in order to preserve anonymity. The numbers I report here are rounded averages of at least 5 individuals around each percentile of interest.

[13] An intuitive example of why this is the case is to think about shopping for groceries and the network of interactions formed between buyers of groceries and supermarkets. Buyers of groceries only form links to supermarkets and hence have low degrees. Supermarkets link to all or most buyers of groceries and hence have high degrees. The degree correlation – high degree supermarkets linking to low degree buyers of groceries – is negative.

of that individual's social circle that are connected to each other, that is, the proportion that also have links between them. For the average individual, the clustering coefficient is 0.105: 10.6 percent of the possible links between any two of her neighbors exist in the network. In Panel C of Figure 2, I plot the average clustering coefficient for each degree in the network. The negative correlation that I document is a common feature of social networks (Jackson and Rogers, 2007): it is less likely that neighbors of a high degree individual are linked to each other than it is for the neighbors of a lower degree individual.

Comparison With Other Population-Scale Social Networks

When assessing a new source of social network data, it is instructive to compare its structural properties with other similar networks. In the Appendix, I provide a detailed comparison of the network formed from mobile money transfers with the Facebook network, as documented by Bailey et al. (2018) and Ugander et al. (2011). I choose this comparison because the Facebook network is one of the few purely social networks where measurements exist for each of the properties I study here. However, an important difference between these network measures is that, unlike networks formed from mobile money transfers, the network based on Facebook friendships does not necessarily measure real world social interactions.

The key takeaway from the comparison is that the network derived from mobile money transfers is strikingly similar to the Facebook network with respect to these structural properties. This is especially true for the degree correlation and patterns of clustering, where measured statistics are within a number of decimal places of each other. These similarities hint at universal structural properties of population-scale networks and fundamental processes governing their formation. In one respect, the comparison suggests an advantage of using mobile money transfers to measure social networks and learn about network formation (Chandrasekhar and Lewis, 2016): unlike Facebook, there are no constraints on the number of links a user can make, meaning that the degree distribution is perhaps more representative of the latent structure of degrees.

2.6 Summary

This section highlights the value of Denmark as a laboratory for harnessing digital trace data: the population are tech-savvy and, having used electronic payment and banking solutions for many years, adoption of the money transfer app was rapid and widespread. Moreover, the Danish setting facilitates combining data with the population administrative registers. This setting is not, at present, possible to recreate precisely elsewhere. However, the growing popularity of mobile money transfer apps in other countries, like Venmo in the U.S., and the increasing usage of electronic payment methods and financial aggregator apps (Gelman et al., 2014; Baker, 2018) suggest that mobile money transfer app data could soon be similarly representative in other settings. Moreover, with access to limited demographic data on users, such as name, email address, telephone number, or address, it is possible to combine these data with external administrative data, as Bailey et al. (2018) do with Facebook data. Thus, this study has value and relevance beyond Denmark for researchers looking for new ways to learn about social networks.

In sum, the evidence I have presented suggests that data from mobile money transfer apps have *potential* as a source of data from which to learn about social networks in many settings: mobile money transfer apps focus on facilitating social payments, their usage is rising and, as a result, data from them is increasingly available at a scale that is representative of full populations. The rest of the paper explores this potential in three analyses.

3 The Importance of Institutions: Family and Institutional Ties Over the Life Cycle

The extent to which an individual's social circle reflects unconstrained choices versus constraints imposed by geographic and social location is a widely studied question in sociology (e.g., Kossinets and Watts, 2006, 2009; Fischer and Oliker, 1983; McPherson, Smith-Lovin, and Cook, 2001). An important dimension of social location are institutional attachments, chiefly schools, universities, and companies. Inside an institution, there is a finite number of social relationships one can form and, due to this, individuals might select a particular institution for the social circle it provides or have their social circle constrained by the limited social possibilities therein. Individuals can also have multiplex relations: social ties related to multiple institutions at once, for example, a cousin or school friend who one also works with on weekends. Moreover, at any point in time, an individual can have multiple social ties that span their institutional attachments at present, but also maintain ties to their institutional attachments from the past.

These properties of social networks, institutional attachments, multiplexity and multiple types of ties, are widely considered important for explaining network structure, patterns of diffusion, and social inequality. For example, parents might select schools due to the peer group their child will have access to, and if peer groups persist and have lasting effects on educational and economic achievement then the extent of freedom in school choice has distributional consequences (Rothstein, 2006). More generally, and related to the overall architecture of social networks, an individual with social ties at different institutions can bridge across long social distances if individuals at those institutions would not otherwise be friends, or even be distant friends-of-friends. These types of ties form "wormholes" linking long-range social distances, and could be important for the speed at which information flows through social networks (Park, Blumenstock, and Macy, 2018).

Despite this importance, an understanding of institutional attachments in social networks has been elusive. The key barrier is the absence of network data, especially at a population-scale, that tracks both social ties and institutional attachments (McPherson, Smith-Lovin, and Cook, 2001). I exploit the combined mobile money transfer network and administrative data on family and institutional ties to quantify the importance of different social institutions in individuals' social circles over the lifecycle. This analysis cannot identify the role of choice versus constraints in social network formation. However, it can help to reveal the importance and persistence of institutions in social networks, which are, as discussed, critical to understand for a range of social and economic policies.

I start by examining the aggregate overlap of the network formed from mobile money transfers with family, education, employer and address networks. I refer to individuals' links within these family and institutional networks as their institutional ties (as defined in Table 1). I focus on the 1.48 million individual app users born in Denmark from 1975 onwards. For these individuals I observe their full family network and lifetime of institutional networks. This group has 49.96 million links in the network formed from mobile money transfers, of which 47.4 percent correspond to an institutional tie. Put another way, for this sample of individuals there are 23.68 million links in the network formed from mobile money transfers that are also a link in the family and/or institutional networks.

The fraction of links that correspond to family or institutional ties might understate the importance of these ties for the average individual due to the presence of individuals with very high degrees. In order to assess the importance of these ties more accurately, I collapse the data to the individual level and calculate the share of individuals' social circles, measured by mobile money transfers, that are rooted in each type of institution. To make this transformation, I need to assign each link to a single institution, removing multiplexity. For example, there are 2.62 million links corresponding to both an education and an employer tie - links between people who do (or have) worked and studied together - and I need to choose which tie to record. For these multiplex relationships, I give priority to family ties and I assign other institutional ties to the earliest institutional tie formed. If both ties were formed in the same year then I define the tie according to a hierarchy that favors address ties over education ties and education ties over employer ties. For example, I record a pair of individuals who started studying together and working together in 1995 as an education tie.

In Table 4, I report statistics on the overlap of the network formed from mobile money transfers with real-world networks of institutional ties. In Panel A, I report the percentage of individuals with at least one link in the network with an individual from each type of institutional tie and any institutional tie. Every individual has at least one link to an institutional tie and the vast majority have a link to a family tie, an employer tie and an education tie. In Panel B, I report the percentage of individuals' social circle, the individuals they link with by mobile money transfer, that are also a family, education, employer, or address tie, or any one of these family and institutional ties. The majority of individuals' links in the network formed from mobile money transfers are with real-world institutional ties and this rises to almost 75 percent of links when weighting by the DKK value of transfers (Column 2) or by the number of transfers (Column 3), or when considering reciprocated links (Column 4).

In Figure 3, I show the importance of institutions in individuals' social circles over the lifecycle. In Panel A, I plot the average degree, number of real-world institutional ties in total, and the number of family, education, employer, and address ties by age. The average degree increases through the teens until it peaks at 18 at which point it decreases from around 54 to around 40 by the age of 40. The total number of real-world institutional ties follows a very similar pattern. However, the composition of institutional ties changes considerably. I show this in the first figure of Panel B: I plot the average share of individuals' total social circles belonging to any and each institutional tie. From a peak at 18, education ties become less important as a share of all links and employer ties become more important, overtaking education ties at the age of 28. Despite most individuals having completed their education by the age of 30, education ties account for around 10 percent of all links through the decade up to the age of 40.

The last two figures in Panel B of Figure 3 show the impact that weighting by tie strength – the DKK value of transfers and the number of transfers – has on the average institutional composition of individuals' social circles over the lifecycle. Weighting by the DKK value of transfers massively increases the emphasis placed on family ties, increasing the total share of family and institutional ties, and the relative emphasis on employer and education ties falls somewhat. This might reflect the important role of family, rather than unrelated social ties, in providing monetary insurance against adverse income shocks (Andersen, Johannesen, and Sheridan, 2019). Weighting by the number of transfers reduces the emphasis placed on family ties, relative to weighting by the DKK value of transfers, and increases the emphasis on education and employer ties to very similar levels as when links are unweighted. In every case, the share of links unrelated to an institutional attachment increases with age.

Overall, this evidence demonstrates the power of data from mobile money transfer apps for measuring *real-world* ties, and the importance of institutions in social networks. The lifecycle variation that I exploit is only cross-sectional, but the evidence is highly suggestive of social ties changing over the lifecycle along with institutional attachments, but there is also evidence of some persistence in ties over time. Moreover, the increase over the lifecycle in individuals' links to others with whom they do not share an institutional attachment, relative to their institutional ties, suggests that other organizations, such as sports clubs, or friends-of-friends become an increasingly important source of social ties over the lifecycle.

4 The Importance of Attributes: Immigrant Networks and Economic Status

Homophily is the tendency of individuals to associate with similar others. Evidence of homophily is uniquitous in social networks (McPherson, Smith-Lovin, and Cook, 2001). I have already presented some evidence of homophily in the network formed from mobile money transfers: individuals tend to associate with others of a similar degree to themselves, and also with whom they share institutional attachments. I now focus on exploring homophily and tie structure with respect to two characteristics of individuals: whether they are immigrants or the descendants of immigrants, and economic status, measured by income.

4.1 Immigrant Networks: Homophily, Integration, and their Correlates

Approximately 12 percent of the Danish population aged 13 years and older are immigrants or the descendents of immigrants - a group I will refer to as non-natives - and these individuals are fairly well represented among users of the mobile money transfer app (see Table 2). I investigate the social networks of these individuals, focusing on non-natives between the ages of 13 to 55 years old for whom I observe country-of-origin. After removing family links from the networks of these individuals, I am left with approximately 290,000 individuals.

A definition of homophily requires some conception of "groups", such that network links can be labelled as connecting two individuals of the same group (in-group ties) or of different groups (out-group ties). As noted by Currarini, Jackson, and Pin (2010) in their analysis of racial homophily, but not explored in their paper, the choice of grouping is crucial to measured homophily. To the extent that homophily among non-natives is driven by their preferences for socializing with similar individuals, the relevant concept of similarity could include all other non-natives, or more limited groups such as those with origin in same global region or the same country. The relevant conception of similarity could transcend specific geographies, with language, culture, or religion serving as the key sources of similarity. Moreover, the extent of homophily and integration is likely to vary across non-natives, and even within different groups defined by country or region of origin.

I focus my investigation on exploring how measured homophily, and social integration, varies across non-natives and by definition of in-group ties. To this aim, I adapt the measures of homophily and excess, or inbreeding, homophily employed by (Currarini, Jackson, and Pin, 2009, 2010). I assume that N individuals in the network can be partitioned into groups $g \in G$. An individual i belonging to group g has s_{ig} links to individuals who belong to the same group g and d_{i-g} links to individuals belonging to other groups. I define the homophily index of individual i belonging to group g as h_{ig} where

$$h_{ig} = \frac{s_{ig}}{s_{ig} + d_{i-g}}.$$

Hence, I define the homophily index of an individual as the share of their social circle, all the people they link to in the network, who are from the same group as themselves. Consequently, I define the homophily index for the entire group g as the average of the homophily indices of the individuals belonging to group g

$$h_g = \frac{1}{N_g} \sum_{i \in N_g} h_{ig}$$

where N_g is the number of individuals in group g.

I want to be able to assess whether a given level of homophily is above or below what I would expect if links were formed at random. To this aim, I define w_g as the share group g individuals in the population. If $h_{ig} - w_g = 0$, then individual i displays baseline homophily: the share of their social circle who are the same group as themselves is exactly the same as their group's prevalence in the population. If $h_{ig} - w_g > 0$, then, by Currarini, Jackson, and Pin's (2009; 2010) terminology, individual i displays inbreeding homophily: they over-represent others from the same group as themselves among their social circle, relative to their prevalence in the population. Further, I define the inbreeding homophily index of individual i as ih_{ig} where

$$i_{ig} = \frac{h_{ig} - w_g}{1 - w_g}.$$

This index normalizes the level of inbreeding homophily $h_{ig} - w_g$ by the maximum possible level $1 - w_g$. Individual i displays inbreeding homophily only if $ih_{ig} > 0$.[14] Again, I define the inbreeding homophily index for group g as a whole as the average of the inbreeding homophily

[14]Further, individual i displays complete inbreeding homophily if $ih_{ig} = 1$, baseline homophily if $ih_{ig} = 0$, and inbreeding heterophily if $ih_{ig} < 0$.

indices of the individuals belinging to that group[15]

$$ih_g = \frac{1}{N_g} \sum_{i \in N_g} ih_{ig}.$$

The final notation I introduce is regarding the measurement of social integration. I define the integration of non-native individual i belonging to group g as dk_{ig}

$$dk_{ig} = \frac{r_{idk}}{t_i}$$

where r_{idk} is the number of native Danish individuals that individual i has in their social circle and t_i is individual i's total degree (the total number of individuals in i's social circle). Thus, I am measuring integration as the share of natives in non-natives' social circles. Again, I define group level integration dk_g as the average of dk_{ig} across all individuals in group g.

I present results on homophily among non-natives in Panel A of Table 5.[16] I consider three definitions of groups: non-natives as a whole, global region of origin, and country of origin. By the first group definition, a non-natives's in-group ties are those connecting them with individuals who are also non-native. The first row in Column 2 of Table 5 reports the average homophily index by this definition: on average, 47 percent of non-natives' social circle are other non-natives. Given that non-natives account for less than 11 percent of the population (Column 1), this suggests a high level of inbreeding homophily, as reflected by the average inbreeding homophily index in Column 5 of 0.4. In Columns 3 and 4, and Columns 6 and 7, I report versions of the homophily and inbreeding homophily indices where I weight links by the DKK value and number of transfers respectively. In all cases, weighting increases measured homophily of non-natives.

According to the second and third group definitions, in-group ties are connections with individuals sharing the same region- and country-of-origin respectively. I provide the country and region of origin definitions in Table A1. Rather than reporting each group-level homophily index for each country-of-origin and region-of-origin, I report the weighted average of the group-level homophily indices across all groups for each group definition. These results are in the second and third rows of Panel A of Table 5. The key takeaway is that measured homophily is still significant but lower when considering these narrower groupings, reflecting that non-natives form social ties that cross the geographic boundaries defined by the regions and countries

[15]The definition of the homophily index and inbreeding homophily index used by Currarini, Jackson, and Pin (2009, 2010) differs slightly from mine. These authors define group g's homophily index as the ratio of the average number of in-group ties among individuals from group g to the average total degree of these individuals. Thus, their definition of group-level homophily is a ratio of averages, which simplifies to the ratio of in-group ties to total ties for a group, and mine is an average of ratios. Presumably, these authors define it as such because they have a model where the average number of within-group links is a parameter. I modify the index for two reasons. First, because the average of ratios is less subject to outliers than the ratio of averages. I am working with a population-scale network whereas they were using school-based networks, hence, I have greater degree variance in my data. Second, I want to explore variation in homophily across individuals within groups and starting from a definition of an individual-level homophily index is useful. However, it does introduce dependencies in measured homophily within groups across individuals.

[16]In Panel B of Table 5, I report measured homophily according to other characteristics of individuals in the network. These characteristics are sex, education, and geographical province. There is significant homophily along each dimension, especially geography. I also provide some graphical evidence on age homophily in Figure A.2, showing bunching in the distribution of age differences between connected individuals at 0 years.

individuals have origin in.[17]

The averages reported in Table 5 illuminate the general patterns of homophily among non-natives but they hide considerable variation in a number of ways. First, there is variation across region and country of origin, both in the level of homophily and, relatedly, in the extent to which individuals with a certain country or region of origin link to other non-natives of a different country or region of origin to themselves. Second, there is variation along dimensions not directly related to region or country of origin but to do with the experience in Denmark of the individual immigrant, for example, their age on arrival. I explore these sources of variation in Figures 4 and 5.

In Panel A of Figure 4, I show how non-native inbreeding homophily, where in-group ties are defined as links with other non-natives, varies on average across descendants and immigrants, and within immigrants by age-on-arrival to Denmark.[18] In the figure, I limit the sample to descendants and immigrants who were 18 years old or younger on arrival, and I require that individuals are 25-55 years old at the time of measurement, that is, as of December 2016.[19] Descendants are on average more homophilous than immigrants who arrive before the age of 14. Average homophily is decreasing slightly with age-on-arrival until the age of 6, the official school starting age, at which point average homophily increases with age-on-arrival. In Panel B, I show the same plot for my social integration measure, the share of Danes (natives) in non-natives' social circles, and I find the reverse pattern: integration decreases with age-on-arrival above the school starting age.

In Panel C of Figure 4, I split the sample by global region-of-origin and I redefine in-group ties in the measure of homophily to be links with other individuals with the same region-of-origin, rather than any other non-native (as in Panels A and B). I apply a diverging color scheme based on the average level of inbreeding homophily among descendants. There are stark differences across non-natives by region of origin and these differences are stable. For example, Western Asians and Sub-Saharan Africans are the most homophilous at all points in the figure and Western Europeans and Northern Americans are always the least. In Panel D, I repeat the same plot but for integration. Again, the pattern reverses, as does the ordering of region-of-origin. In both Panels C and D, number of years in school appears to be important for explaining homophily and integration: for immigrants of almost any region-of-origin, homophily increases and integration decreases with each arrival-age to Denmark above the school starting age.[20]

[17] This result also reflects, to some extent, that certain countries of origin represent a vanishingly small share of the population.

[18] In Figure A.3, I repeat Panels A and C of Figure 4 using the basic homophily index, rather than inbreeding homophily. The results do not change.

[19] I make this restriction on age at measurement in order to reduce concerns that the results in this figure are completely explained by number of years in Denmark. However, this age requirement makes no difference to the results. Moreover, I only show plots of simple averages by age-on-arrival for clarity and brevity. However, the relationship between homophily/integration and age-on-arrival is the same if I instead plot deviations in outcomes by each arrival age relative to arriving at ages 0-1 based on a regression of these outcomes (homophily/integration) on age-on-arrival indicators, controls for sex, number of years in Denmark indicators, region-of-origin indicators, and age in years indicators. In all such specifications, I find that homophily increases and social integration decreases with increasing arrival age above the school starting age.

[20] Part of the explanation could be compositional changes in the immigrants arriving at each age, but this is less likely to be an issue for age-on-arrival below 15.

Next, I investigate the community structures formed by links between non-natives with different country of origin. This relates back to my earlier finding that homophily is greater when considering all non-natives as one group relative to groups based on individual country-of-origin, hinting at important links between non-natives that cross country-of-origin boundaries. I focus on individuals from the top 100 country-of-origin present in the network. For an individual i with country-of-origin g, I calculate their share of links to individuals with a different country-of-origin k within this sample of country-of-origin (the top 100). I define this share as

$$x_{igk} = \frac{o_{igk}}{t_i^{100}}$$

where t_i^{100} is individual i's total number of links with other non-natives with a country-of-origin within the top 100 and o_{igk} is individual i's total number of links to individuals with country-of-origin k. I then take the average of x_{igk} across individuals with country-of-origin g. This gives me x_{gk}, measuring the share, on average, that individuals with country-of-origin k represent in individuals with country-of-origin g's social circle of non-natives. In other words, x_{gk} is a measure of how "important", on average, links to individuals with country-of-origin k are to individuals with country-of-origin g, out of all of their links with other non-natives. Finally, I want to capture "importance" relative to what we would expect x_{gk} to be under random link formation, where individuals with country-of-origin k are simply represented in social circles according to their share in the non-native population. Thus, I divide x_{gk} by the share of individuals with country-of-origin k in order to arrive at a scaled measure $\overset{scaled}{x_{gk}}$, which is above 1 if individuals with country-of-origin k are over represented in the non-native social circle of immigrants with country-of-origin g (relative to random link formation).

I calculate $\overset{scaled}{x_{gk}}$ for every pair of country-of-origin. This results in a directed network, where the weights in each direction are given by $\overset{scaled}{x_{gk}}$.[21] I visualize this weighted and directed network in Figure 5. Each node is a country-of-origin, labelled with its 2-digit ISO code (see Table A1 for the mapping from ISO codes to country of origin) and I show only the strongest links (filtering to those where $\overset{scaled}{x_{gk}}$ is above 1.05). I use the ForceAtlas2 algorithm in Gephi (Jacomy et al., 2014) in order to choose the location of country-of-origin in space. Broadly speaking, the algorithm pulls together nodes with a high weight and repels nodes with a low weight, accounting for different weights in each direction between a pair. I color the nodes for each country-of-origin according to the global region.

Figure 5 shows clear community structures among non-natives by country of origin. The communities seem to correspond broadly to global geography. However, linguistic and, perhaps, cultural factors appear to be important. For example, the Sub-Saharan African country-of-origin bunch tightly in the southwest quadrant, apart from South Africa, which is located near to the Anglo-Saxon country-of-origin in the northeast quadrant. Moreover, Spain and the Spanish-speaking Latin American country-of-origin are grouped together in the eastern portion of the figure, with Brazil and Portugal slightly separated. Using a similar approach to study

[21]The network is directed because $\overset{scaled}{x_{gk}}$ is not necessarily the same as $\overset{scaled}{x_{kg}}$, where I reverse k and g. For example, individuals with Iraq as their country of origin might over represent individuals of Egyptian origin in their networks whereas individuals with Egypt as their country of origin might under represent individuals of Iraqi origin, relative to a case of random link formation.

online relationships between immigrants to the U.S. on Facebook, Herdağdelen et al. (2016) find comparable community structures. These community structures are important to understand because they might interact with patterns of social interaction and economic integration.

Finally, in Figure 6 I show the correlation between economic integration and average inbreeding homophily (Panel A) and social integration (Panel B) of non-natives by country-of-origin. I measure the economic integration of non-natives with a given country-of-origin as the average annual income, calculated over 3 years in order to avoid transitory fluctuations, of non-natives with that country-of-origin who are aged between 30 and 40 years old (that is, at prime working ages). I color each country-of-origin according to the global region-of-origin, using the same colors as in Figure 4. There is a strong correlation between economic integration and these features of non-natives' social networks. Running a regression on the data in Panel B, weighting country-of-origin by number of individuals, returns a coefficient suggesting that a 10% higher share of total links that are with Danes correlates with approximately 4,200 USD (27,430 DKK) higher average annual income. Although the direction of causality is unclear, the results provide evidence that immigrants' economic success and social integration are co-determined.

4.2 Economic Status: Social Network Stratification by Income

I turn to quantifying the joint distribution of individual income and the average income among individuals' social circles in the network derived from mobile money transfers. Considering income as a measure of economic status, this joint distribution captures the extent to which high status individuals tend to associate with other high status individuals, and the equivalent for low status individuals.

In order to perform this quantification, I draw on the intergenerational mobility literature (e.g., Chetty et al., 2014; Chetty and Hendren, 2018). This literature has established a number of guiding principles for estimating the joint distribution of parent and child income or wealth, guidelines that I will apply to my setting. First, in order to remove the impact of transitory fluctuations and lifecycle effects, it is important to measure outcomes over a number of years during prime working ages. Second, in order to further control for important life cycle growth patterns that might vary by sex, it is important to control for birth cohort and sex. Third, in order to account for individuals with no earnings and to further increase robustness to sampling or specification issues, it is preferable to transform outcomes into percentile ranks.

In light of these principles, I focus on links in the network between individuals aged 30 to 40 years old in December 2016 who are not family members. This subgraph of the full mobile money transfer network contains 578,226 individuals and 6.54 million links. For these individuals, I calculate their average annual total income over three years from 2013 to 2015. To control for birth cohort and sex effects, I regress this income measure on a fully saturated model with birth cohort indicators, a sex indicator and interactions between birth cohort indicators and sex, and I calculate the residual for each individual. For each individual, I calculate average residualised income among their social circle in the subgraph. I term this an individual's network average income. I then calculate percentile ranks of individual residualised income and network average income.

In Figure 8, I present the mean and median percentile ranks of network average income for each individual income rank, along with the 25^{th} and 75^{th} percentiles. The relationships are mostly linear and steep. The estimated slope of the conditional expectation, represented by the mean, is 0.5. At the median, moving up 10 percentiles in the indivdiual income distribution is associated with a movement of 7 percentiles up the distribution of network average income, controlling for age and sex. The joint distribution is less dispersed at the high end of the individual income distribution, shown by the smaller distance between the 25^{th} and 75^{th} percentiles. This suggests that individuals of high economic status have social circles that are highly concentrated with other individuals of high economic status, whereas low economic status individuals have somewhat more diverse social circles with respect to economic status. However, the joint distribution displays skew towards low network average income ranks at low individual income ranks, and the opposite at high individual income ranks. Thus, there are strong correlations in economic status between connected individuals in social networks, especially so for those of highest and lowest status.

My findings are similar to those of Leo et al. (2016), who study patterns of communication in a telephone network across 9 groups defined by spending behavior. I contribute with a richer quantification of the joint distribution of economic status between individuals and their social circle, in a social network defined by real-world interactions and using administrative income data. The high degree of stratification by economic status and the concentration of high economic status in the networks of the rich shed light on one possible channel through which inequality propagates (DiMaggio and Garip, 2012).

5 The Importance of Interactions: Social Influence in Spending Decisions

In the final section of the paper, I investigate social influence in the network formed from mobile money transfers. Social influence, often referred to interchangeably as contagion, is considered to be pervasively important for explaining individual behavior, but difficult to identify empirically (Aral and Walker, 2012; Aral, Muchnik, and Sundararajan, 2009).

I turn to this identification challenge in the context of social influence on spending decisions. How much and what people consume is a fundamental part of their welfare and individual consumption decisions influence outcomes for the whole economy. This impact at the macro-level means it is critical to understand what shapes individual spending choices. Several properties of consumption make it likely that social influence is important: spending is often highly visible, it is shared with others, and decisions about how much to consume today are impacted by expectations and information about how the economy will look in the future. Recognizing these properties, there are a number of theories of social influence on consumption, dating back to Veblen (1899).

I focus on job loss, which is one of the most important economic risks that individuals face and a shock that results in persistently lower spending for those affected (Ganong and Noel, 2018; Andersen et al., 2018). Alt et al. (2017) find evidence that individuals update their expectations of own unemployment risk when a social tie experiences unemployment. Bailey et al. (2018) find that interactions through social networks that affect expectations, like those identified by Alt et al. (2017), can have consequences for economic decision-making, in their case on the decision to purchase a house. This mechanism could cause job loss shocks to results in spillovers on spending through interactions in social networks. Moreover, the drop in spending by those affected by job loss might reduce spending of individuals in their social circle through a “keeping up with the Joneses” mechanism, that is, by reducing the social pressure to consume (De Giorgi, Frederiksen, and Pistaferri, 2018). I test whether job loss shocks have a causal effect on the spending of individuals in the social circle of those affected, an important question for the design of social insurance and macroeconomic stablization policies.

Spending and Income Data

In order to test for the contagious effects of job loss on spending, I need data on income, unemployment experiences and spending. I construct spending measures from the full transaction records of the largest Danish retail bank, Danske Bank (henceforth, “the bank”), which I combine with the network and administrative register data on servers at Statistics Denmark.[22] The bank categorizes account outflows into groups based on method of payment and I use this categorization to construct individual-level measures of total monthly card and bill expenditure, and ATM cash withdrawals. For card and bill payments, I exclude transactions that are not expenditure, such as tax and debt payments, using merchant category codes (MCCs) for card payments and a proprietary classification for bill payments. Moreover, I use the MCCs, an international standard for labelling card payments, and the bill classification to construct measures of specific types of expenditure, such as grocery store, home appliance, and total nondurable expenditure.[23]

These spending data cover over 30 percent of individuals in the network and the period from January 2009 to December 2016. It is important to measure spending at the household level. Hence, I aggregate spending to the household level, using household identifiers from the administrative registers, and I limit the sample to household-months where I observe every adult in a household in the spending data. Moreover, I require that households have at least 5 card, bill or ATM cash transactions in every month of the year, in order to capture total spending and minimize the probability that sample households are spending significantly from accounts at other banks. In Andersen et al. (2018), we use the same data and provide evidence on the accuracy of these data for measuring household spending, especially in Denmark where electronic payment usage is high.

For individuals in the households in the spending data set, I extract their social circle from the mobile money transfer network, excluding family ties, and I aggregate them to the household level, such that I have records linking each sample household to the individuals with which a household member has been involved in a mobile money transfer. Thus, I record all individuals in households’ social circles.

[22] Again, this combination is facilitated by social security (CPR) numbers, which are present in all data.

[23] I employ the same definition of nondurable expenditure as Ganong and Noel (2018), who also use bank transaction data. I define nondurable expenditure as all cash withdrawals, card payments excluding purchases at stores that primarily sell durable goods (e.g., home electronics or furniture stores) and tax and debt payments, and bill payments excluding tax and debt payments and payments for durable goods (e.g., vehicle purchases or construction goods).

My income and job loss data comes from monthly income records for the population of Denmark, covering the period from January 2008 to March 2015. I get these data from the Danish tax authority, who gather the information for the purposes of calculating tax liability. I use these data to construct a measure of total monthly pre-tax income (including government transfers). Moreover, these data include unique identifiers for the employer making wage payments to individuals. I use these employer identifiers to identify job loss shocks. I define a job loss shock as a month in which an individual receives their final wage payment from an employer who has paid them at least 1,500 USD (10,000 DKK) in every month of the previous 6 months and then transitions to receiving unemployment insurance benefits within two months of that date. I limit my focus to job loss shocks affecting individuals aged 25 to 50 years old in order to remove transitions in an out of studying and retirement, events that are not shocks. This leaves me with a sample of 251,612 job loss events, an idiosyncratic income shock, for 224,579 individuals.

I now have four datasets. The first contains monthly spending data for households who are active customers of the bank, my primary sample of interest. The second records the individuals that these households link with in the network formed from mobile money transfers. The third contains monthly income data for the population. The fourth contains information on individuals who experience a job loss. I exploit these data in an event study design.

Identification: Event Study Design

My objective is to estimate the dynamics of household income and spending around the time that someone in their social circle, a network neighbor, experiences a job loss shock. To do this, I estimate event study models of the form

$$y_{it} = \alpha + \sum_{m=\{-5,\ldots,6\}} \beta_m \mathbb{1}[e_{jt} = m] + \sum_n \gamma_n \mathbb{1}[a_{it} = n] + \delta_t + \epsilon_{ijt} \tag{1}$$

where i indexes households, j indexes network neighbors who experience a job loss, t indexes months, y_{it} is one of the outcomes of interest, and e_{jt} are event-month indicators recording time relative to j's job loss (where month 0 is the month of the shock, coinciding with the final wage payment, month -1 is one month prior, and so on). I include time indicators for head of household age a_{it} and calendar month fixed effects δ_t. A household can be exposed to multiple network neighbors experiencing a job loss shock in any given month - such that they appear in equation 1 multiple times in the same month - and shocked individuals can influence multiple households in the same month, and they can experience multiple events over the sample period.[24] I account for these dependencies across observations when I estimate equation 1 by double clustering the standard errors at the household and shocked network neighbor levels.

The coefficients β_m capture the dynamic response of the household outcome variable around the month that a network neighbor experiences a job loss shock. I estimate the model from 6 months prior to neighbors' shocks and up to 6 months after. I exclude indicators for month 6

[24] Although not a network setting, Lafortune, Rothstein, and Schanzenbach (2018) have a similar estimation model where units can experience multiple overlapping events. They also "duplicate-and-stack" for such units. They demonstrate that this approach performs well and that weighting observations to account for the number of times they appear has no effect, hence I do not apply a weighting scheme.

prior to a shock such that the coefficients β_m capture the relative change in monthly outcomes, measured in Danish krone (DKK), compared to that point in time. Further, I only estimate the model for households who are exposed to a network neighbor experiencing a job loss shock. This restriction means that I am identifying the coefficients β_m by comparing outcomes of households in a given month that have household heads who are the same age but that are exposed to network neighbors losing their jobs at different points in calendar time.

For equation 1 to estimate social influence, the identifying assumption I must make is that, conditional on exposure to a network neighbor experiencing a shock and head-of-household age, the precise timing of network neighbors' shocks is uncorrelated with other factors that might influence outcomes for households. The key threats to this identification assumption are common shocks. For example, a member of household i might work for the same employer as individual j, their network neighbor who experiences a job loss shock. If so, household i might change their spending around j's job loss because they also lose their job at this time, or because of bad news about the health of their employer, rather than directly through learning about j's job loss. This common shocks problem, often driven by homophily (such as working for the same firm), is the primary challenge to identifying social influence (Aral, Muchnik, and Sundararajan, 2009). In order to remove the impact of common shocks and identify social influence, I drop any households who work in the same industry as their shocked network neighbor at the time of job loss.

Another challenge to identification is endogenous tie formation: household i might meet individual j because of the shock - at a job center, for example - and this drives any observed correlations in outcomes around the event. This seems unlikely. However, in case it is a problem, I only include households and their shocked neighbors who knew each other prior to the job loss event. To do this, I use the dynamic structure of the mobile money transfers data to limit my sample to households and their shocked neighbors who were involved in a mobile money transfer prior to the shock. I also include households and individuals who are linked by an institutional tie – school, employer, or address – from prior to the shock date. Finally, I limit the sample to working age households, where the head of the household is within 5 years of age of their network neighbor who experience the job loss shock. The final estimation sample includes 232,923 households, exposed to the job loss shocks of 154,542 individuals, and 6,994,304 regression observations.

Estimating equation 1 is useful for presenting graphical evidence of social influence. In order to summarize the full effect of exposure to a network neighbor's job loss shock on household outcomes at a monthly level, I estimate

$$y_{it} = \alpha + \beta After_{jmt} + \sum_{n} \gamma_n \mathbb{1}[a_{it} = n] + \delta_t + \epsilon_{ijt} \tag{2}$$

where, rather than a full set of event time indicators, I include the indicator $After_{jmt}$, which is equal to 1 in every month from the point month that j experiences the job loss shock, and 0 otherwise. The coefficient β captures the full effect of the network neighbor's job loss shock on monthly outcomes for households.

Results

I start by presenting the results of event study designs for the shocked individuals themselves. This is equivalent to estimating equation 1 from individual j's perspective, around their own job loss event.[25] I show these results because they help with interpretation of the results for the households of interest and they demonstrate the economic impact of job loss shocks, a precondition for seeing any consequential social influence. In the top left plot in Figure 8, I show the results for the income of the shocked individuals around their job loss, plotting the estimated event time coefficients, the equivalent of the β_m, as solid lines and 95 percent confidence intervals on the estimates as dashed lines. In the top right plot, I show the results for nondurable spending as the outcome. The key takeaways are that individuals experience a persistent drop in income following job loss and a persistent drop in non-durable expenditure, and this drop in expenditure starts, on average, 2 months prior to the final wage payment in month 0. The likely driver of the anticipation effect in spending is advance notice of termination.[26]

I now turn to the evidence of social influence. I show the main results in the bottom row of Figure 8, plotting the estimated β_m from the event study design (equation 1) for household income (left) and household nondurable expenditure (right) and the 95 percent confidence intervals of the estimates. There is no change in household income around the event of exposure to a network neighbor's job loss shock. This suggests that there is no common shock. However, there is clear evidence of social influence on spending: household expenditure drops significantly 2 months prior to the job loss shock of their neighbor – precisely the same month that the neighbor themselves starts cutting back on expenditure – and remains at this new level throughout the estimation period.

I summarize the full social influence effect on household spending by estimating equation 2, excluding the months where I see anticipation. At the top of Figure 9, I present the estimated coefficient β, the relative DKK change in monthly spending from before to after the job loss shock of the network neighbor, as a percentage of the average DKK spending of households 6 months prior to their network neighbors' job loss shocks. Below this value, I report equivalent relative spending changes in percentages for 11 other spending measures. The results in Figure 9 suggest that households, when exposed to a network neighbor's job loss, cut their monthly nondurable expenditure by 0.4 percent of spending 6 months before their network neighbor's job loss on average.[27] Relative effects are slightly larger for total spending, which includes durables, and home improvements and furnishings, a major durable good. Households cut back on luxuries, such as food away from home and travel, and they reduce spending on groceries, a component of spending that is highly salient and for which, in many cases, it is possible to economize.

Overall, the evidence suggests that household spending decisions are susceptible to influence through interactions in social networks. Households reduce their spending in precisely the same month that their network neighbors begin to cut back on spending in anticipation of their upcoming job loss, and then they maintain this new level. The precise simultaneity in the initial

[25] I cluster standard errors at the individual, the network neighbor, level.

[26] In Andersen et al. (2018), we explore these patterns in more detail. The spike in income at month 0 for the shocked individuals is driven by severance payments and holiday pay.

[27] The estimated value of β is 83 DKK, equivalent to a 12.5 USD drop in monthly expenditure on average.

spending drop is suggestive of an expectations mechanism: when individuals learn of their upcoming job loss, they tell their friends, and these friends increase their expectations of their own probability of job loss and respond by cutting back on spending. Other hypotheses concerning the channel of influence are based on the act of spending itself, rather than expectations, and in light of this evidence seem less likely to be driving effects.[28] For example, if individuals attempt to "keep up with the Joneses" – reducing expenditure when their peers do as a result of lower social pressure to consume – then one would expect the time profile of spending of households to more closely match that of the shocked individual throughout the unemployment spell, dropping again after the final wage payment. In sum, my results provide evidence for the important role of social interactions on expectations and behavior, in line with the findings by Alt et al. (2017) and Bailey et al. (2018).

6 Discussion and Conclusions

In this paper, I demonstrate how the increasing popularity of mobile money transfer apps is generating population-scale data on real world social interactions, presenting researchers with an opportunity to better comprehend social networks and their role in social and economic behavior. To this end, I harness data from MobilePay, the monopoly mobile money transfer app in Denmark, containing the near universe of the population's person-to-person mobile money transfers. From these data, I construct a population-scale social network. I detail the structural properties of the social network, their striking accordance with other known social networks, and their discordance with networks of transactional relationships. This is evidence of the potentially highly-social nature of monetary exchange (Zelizer, 1994, 1996).

Further, I combine the network with population-wide data on individual-level socio-economic attributes and I overlap the network with networks of family ties and institutional attachments, derived from government administrative registers. In this uniquely rich setting, I show how social networks derived from mobile money transfers can reveal novel insights by analyzing three important social network phenomenon: the role of institutions in tie formation; the importance of individual attributes in tie formation, and the influence of interactions in social networks on behavior. I produce four sets of findings.

First, I quantify the importance of family and institutional ties in individuals' social circles over the life cycle. I find evidence to suggest that ties evolve with changing institutional attachments but that there is significant persistence in ties formed through past attachments. Given this persistence, access to beneficial instituional environments early in life is likely to have lasting effects.

Second, and turning to the importance of attributes in tie formation, I present evidence on variation in social segregation and integration across immigrants by country of origin and age on arrival. Segregation appears to increase and integration decrease with each increase in arrival age above the school starting age. Moreover, differences in the segregation and integration

[28] I do not consider some other channels of social influence. One is risk-sharing (Townsend, 1994). Under risk-sharing, households give money to each other in order to smooth consumption through shocks to resources. The primary reason I do not investigate this channel is that Denmark has a generous tax-financed social insurance system, which is expected to crowd out private risk-sharing among households (Banerjee et al., 2018). In future versions of this paper, I hope to empirically test for this channel.

of immigrants by country of origin correlates with economic integration, measured by average prime age income. These findings emphasize the joint determination of immigrants' social integration and economic success, and the possible role of factors not directly in their control - childhood age on arrival - for these outcomes.

Third, I apply techniques from the intergenerational mobility literature to measure the extent of stratification in the network by economic status. I find that the networks of the rich and those of the poor are divided, with individuals at the top of the income distribution having the most segregated networks. This finding leaves room for the role of economic stratification in social networks as a factor in perpetuating inequality.

Finally, I use high-frequency data on spending and income to identify social influence on household expenditure decisions. Households appear to cut their spending when an individual in their network experiences a job loss shock, and the mechanism seems to be driven by changing expectations. This result suggests that social influence in expenditure could contribute to the propogation of aggregate shocks.

In addition to their independent relevance, these findings highlight 5 advantages of using data from mobile money transfer apps to measure social networks, beyond overcoming the biases and limitations inherent to self-reported network measures. First, due to their increasing popularity and network externalities in adoption that lead to monopoly provision, data from mobile money transfer apps are increasingly available and can measure social networks at population-scale. Second, networks derive from mobile money transfers are real-world, offline networks capturing multiple types of relationships. Third, networks derive from mobile money are unconstrained; there are no restrictions on who one can link with or how many links one can have, overcoming problems such as top coding that bias network inference. Fourth, networks derive from mobile money are dynamic, reflecting the dynamics of real-world social networks. Fifth, and finally, networks derive from mobile money have simple, meaningful measures of relationship intensity embedded.

However, data from mobile money transfer apps have some limitations for learning about social networks, and it is important to investigate these challenges in future research. First, the Danish setting for this study is especially suited to exploiting digital trace data and it is likely that similar data in other countries will not be as representative yet. However, the global growth in usage of mobile money transfer apps suggests that these apps are generating massive sources of network data for large populations. Second, and related, in many contexts it will not be possible to combine data from these apps with such rich individual-level information. However, the analyses by Baker (2018) and Gelman et al. (2014) of data from financial aggregator apps suggest that even the limited demographic data that such apps collect can be used to correct for sampling biases. Moreover, Bailey et al. (2018) demonstrate that, with cooperation from data providers, it is possible to combine similar data with external demographic data using basic information about app or product users, such as email address, name, and telephone number.

The final, and perhaps most important, limitation of these data is that they measure only a subset of the full latent social network. This limitation is not unique to my data source: almost all sources of social network data, from survey-elicited to Facebook, capture only part of the true structure of social interactions. The non-randomness of this sampling is, however, partly particular to the definition of a social tie based on monetary exchange. For example, it seems plausible that individuals will not expect their poorer friends to pay them back for a shared meal, and the network might miss such links.[29] It is also plausible that very close social ties operate by a system of reciprocity that does not involve monetary exchange ("I'll get this one, you get the next."). On top of these sources of bias, a general problem is the slight underrepresentation of some groups in the network, such as older individuals and the very low income, implying that the data does not reflect their true social relations.

It is unclear how the different sources of bias in measuring true social structure from mobile money transfers will aggregate in the results I have presented here. It is important to recognize that this is a feature of all such digital trace data and that there is often little that can be done to learn about the directions and magnitudes of bias. However, these issues warrant attention and an important, and fruitful, avenue for research is to apply new methods to these data to infer missing links (e.g. Chandrasekhar and Lewis, 2016), exploiting the data on demographics and institutional ties. Alternatively, one could attempt to recover an implied network from panel data on behaviors in the population and compare the result with the links in the mobile money transfer data (e.g., de Paula, Rasul, and Souza, 2018). These investigations would help build an understanding of the types of social ties that monetary exchange captures and they would help also aid a nascent field of research attempting to understand how to infer social structure when population-scale social network data, as I use here, are not available.

[29] As a country with very low levels of inequality, it is likely this problem is less severe in Denmark than elsewhere.

References

Abdulkadiroğlu, Atila, Nikhil Agarwal, and Parag Pathak. 2017. "The Welfare Effects of Coordinated Assignment: Evidence from the NYC High School Match." *American Economic Review* 107 (12):3635–3689. URL `http://www.nber.org/papers/w21046.pdf`.

Alt, James E., Amalie Jensen, Horacio A. Larreguy, David D. Lassen, and John Marshall. 2017. "Contagious Political Concerns: Transmission Using the Danish Population Network."

Andersen, Asger Lau, Amalie Jensen, Niels Johannesen, Claus Thustrup Kreiner, Søren Leth-Petersen, and Adam Sheridan. 2018. "How do Households Respond to Unemployment Shocks? Lessons from Multiple High-Frequency Data Sets."

Andersen, Asger Lau, Niels Johannesen, and Adam Sheridan. 2019. "Bailing Out the Kids: New Evidence on Intra-Family Insurance from 1 Billion Bank Transfers."

Aral, S., L. Muchnik, and A. Sundararajan. 2009. "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks." *Proceedings of the National Academy of Sciences* 106 (51):21544–21549. URL `http://www.pnas.org/cgi/doi/10.1073/pnas.0908800106`.

Aral, Sinan and Dylan Walker. 2012. "Identifying Influential and Susceptible Members of Social Networks." *Science* 337 (6092):337–341. URL `http://science.sciencemag.org/content/337/6092/337`.

Bailey, Michael, Rachel Cao, Theresa Kuchler, and Johannes Stroebel. 2018. "The Economic Effects of Social Networks: Evidence from the Housing Market." *Journal of Political Economy* 126 (6):2224–2276.

Baker, Scott R. 2018. "Debt and the Response to Household Income Shocks: Validation and Application of Linked Financial Account Data." *Journal of Political Economy* 126 (4):1504–1557. URL `https://www.journals.uchicago.edu/doi/10.1086/698106`.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson. 2013. "The Diffusion of Microfinance." *Science* 341 (6144):1236498–1236498. URL `http://www.sciencemag.org/cgi/doi/10.1126/science.1236498`.

Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. 2018. "Changes in Social Network Structure in Response To Exposure To Formal Credit Markets." URL `https://web.stanford.edu/$\sim$arungc/BCDJ_change.pdf`.

Barabási, A.-L. 2016. *Network Science*. Cambridge: Cambridge University Press.

Bernard, Andrew B., Andreas Moxnes, and Karen Helene Ulttveit-Moe. 2018. "Two-Sided Heterogeneity and Trade." *The Review of Economics and Statistics* 100 (3):424–439.

Björkegren, Daniel. 2018. "The Adoption of Network Goods: Evidence from the Spread of Mobile Phones in Rwanda." *Review of Economic Studies* 0:1–28.

Bleakley, Hoyt and Aimee Chin. 2010. "Age at Arrival, English Proficiency, and Social Assimilation Among US Immigrants." *American Economic Journal: Applied Economics* 2 (1):165–192. URL `http://www.jstor.org/stable/25760197`.

Blumenstock, Joshua E., Nathan Eagle, and Marcel Fafchamps. 2016. "Airtime transfers and mobile communications: Evidence in the aftermath of natural disasters." *Journal of Development Economics* 120:157–181.

Breza, Emily, Arun G. Chandrasekhar, Tyler H. McCormick, and Mengjie Pan. 2017. "Using Aggregated Relational Data to feasibly identify network structure without network data." URL `http://arxiv.org/abs/1703.04157`.

Chandrasekhar, Arun G. and Randall Lewis. 2016. "Econometrics of Sampled Networks."

Chetty, Raj and Nathaniel Hendren. 2018. "The Impacts of Neighborhoods on Intergenerational Mobility: Childhood Exposure Effects and County-Level Estimate." *The Quarterly Journal of Economics* 133 (3):1107–1162.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility."

Choi, Syngjoo, Shachar Kariv, and Edoardo Gallo. 2016. "Networks in the Laboratory." In *The Oxford Handbook of the Economics of Networks*.

Crandall, D. J., L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. 2010. "Inferring social ties from geographic coincidences." *Proceedings of the National Academy of Sciences* 107 (52):22436–22441. URL `http://www.pnas.org/cgi/doi/10.1073/pnas.1006155107`.

Currarini, S., M. O. Jackson, and P. Pin. 2010. "Identifying the roles of race-based choice and chance in high school friendship network formation." *Proceedings of the National Academy of Sciences* 107 (11):4857–4861. URL `http://www.pnas.org/cgi/doi/10.1073/pnas.0911793107`.

Currarini, Sergio, Matthew O. Jackson, and Paolo Pin. 2009. "An Economic Model of Friendship: Homophily, Minorities and Segregation." *Econometrica* 77 (4):1003–1045.

De Giorgi, Giacomo, Anders Frederiksen, and Luigi Pistaferri. 2018. "Consumption Network Effects." URL `http://www.nber.org/papers/w22357.pdf`.

de Paula, Áureo, Imran Rasul, and Pedro Souza. 2018. "Recovering Social Networks from Panel Data: Identification, Simulations and an Application."

Dilling, Søren. 2013. "Nu kan du sende penge til dine venner over mobilen." *Politiken* URL `https://politiken.dk/oekonomi/privatoekonomi/art5445543/Nu-kan-du-sende-penge-til-dine-venner-over-mobilen`.

DiMaggio, Paul and Filiz Garip. 2012. "Network Effects and Social Inequality." *Annual Review of Sociology* 38:93–118.

Dustmann, Christian, Albrecht Glitz, Uta Schönberg, and Herbert Brücker. 2016. "Referral-based job search networks." *Review of Economic Studies* 83 (2):514–546.

Eagle, Nathan, Alex (Sandy) Pentland, and David Lazer. 2009. "Inferring friendship network structure by using mobile phone data." *Proceedings of the National Academy of Sciences* 106 (36):15274–15278.

Feld, Scott L. 1991. "Why Your Friends Have More Friends Than You Do." *American Journal of Sociology* 96 (6):1464–1477. URL `https://www.journals.uchicago.edu/doi/10.1086/229693`.

Fischer, Claude S. and Stacey J. Oliker. 1983. "A Research Note on Friendship, Gender, and the Life Cycle." *Social Forces* 61 (1):124–133.

Fryer Jr., Roland G. and Glenn C. Loury. 2013. "Valuing Diversity." *Journal of Political Economy* 121 (4):747–774. URL `http://journals.sagepub.com/doi/10.1177/0011000009334094`.

Ganong, Peter and Pascal Noel. 2018. "Consumer Spending During Unemployment: Positive and Normative Implications." *Forthcoming in the American Economic Review* .

Gelman, Michael, Shachar Kariv, Matthew D. Shapiro, Dan Silverman, and Steven Tadelis. 2014. "Harnessing naturally occurring data to measure the response of spending to income." *Science* 345 (6193):212–215.

Granovetter, M S. 1973. "The Strength of Weak Ties." *The American Journal of Sociology* 78 (6):1360–1380.

Harder, Niklas, Lucila Figueroa, Rachel M. Gillum, Dominik Hangartner, David D. Laitin, and Jens Hainmueller. 2018. "Multidimensional measure of immigrant integration." *Proceedings of the National Academy of Sciences* 115 (45):11483–11488. URL `http://www.pnas.org/lookup/doi/10.1073/pnas.1808793115`.

Herdağdelen, Amaç, Bogdan State, Lada Adamic, and Winter Mason. 2016. "The social ties of immigrant communities in the United." In *8. WebSci 2016: Hannover, Germany.*

Hwong, Connie. 2017. "Chart of the Week: Who's Using Which Mobile Payment Apps?" *Verto Analytics* URL `https://www.vertoanalytics.com/2017/08/16/chart-week-mobile-payment-apps/`.

Jack, William, Adam Ray, and Tavneet Suri. 2013. "Transaction networks: Evidence from mobile money in Kenya." *American Economic Review* 103 (3):356–361.

Jackson, Matthew O. 2017. "The Friendship Paradox and Systematic Biases in Perceptions and Social Norms." *Forthcoming in the Journal of Political Economy* .

Jackson, Matthew O and Brian W Rogers. 2007. "Meeting Strangers and Friends of Friends: How Random Are Social Networks?" *American Economic Review* 97 (3):890–915.

Jackson, Matthew O., Brian W. Rogers, and Yves Zenou. 2017. "The Economic Consequences of Social-Network Structure." *Journal of Economic Literature* 55 (1):49–95.

Jacobs, Harrison. 2018. "One photo shows that China is already in a cashless future." *Business Insider* URL `https://nordic.businessinsider.com/alipay-wechat-pay-china-mobile-payments-street-vendors-musicians-2018-5`.

Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. 2014. "ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software." *PLoS ONE* 9 (6):1–12.

Jones, Jason J., Jaime E. Settle, Robert M. Bond, Christopher J. Fariss, Cameron Marlow, and James H. Fowler. 2013. "Inferring Tie Strength from Online Directed Behavior." *PLoS ONE* 8 (1):8–13.

Kossinets, Gueorgi and Duncan J. Watts. 2006. "Empirical Analysis of an Evolving Social Network." *Science* 311 (January):88–90.

Kossinets, Gueorgi and DuncanÂ J. Watts. 2009. "Origins of Homophily in an Evolving Social Network." *American Journal of Sociology* 115 (2):405–450. URL `http://www.journals.uchicago.edu/doi/10.1086/599247`.

Lafortune, Julien, Jesse Rothstein, and Diane Whitmore Schanzenbach. 2018. "School finance reform and the distribution of student achievement." *American Economic Journal: Applied Economics* 10 (2):1–26.

Leo, Yannick, Eric Fleury, J. Ignacio Alvarez-Hamelin, Carlo Sarraute, and Márton Karsai. 2016. "Socioeconomic correlations and stratification in social-communication networks." *Journal of the Royal Society Interface* 13 (125):no pagination. URL `http://rsif.royalsocietypublishing.org/content/royinterface/13/125/20160598.full.pdf%0Ahttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed18b&NEWS=N&AN=614329918`.

Marsden, Peter V. 1990. "Network Data and Measurement." *Annual Review of Sociology* 16 (1):435–463. URL `http://www.annualreviews.org/doi/10.1146/annurev.so.16.080190.002251`.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27 (2001):415–444.

Miranda, Leticia. 2017. "People Are Getting Kicked Off Venmo For Breaking Its Very Long User Agreement." *BuzzFeed News* URL `https://www.buzzfeednews.com/article/leticiamiranda/people-are-getting-kicked-off-venmo-for-breaking-its-very`.

Newman, M. E.J. and Juyong Park. 2003. "Why social networks are different from other types of networks." *Physical Review E* 68 (3):8.

Onnela, J-P, J Saramäki, J Hyvönen, G Szabó, D Lazer, K Kaski, J Kertész, and A.-L. Barabási. 2007. "Structure and tie strengths in mobile communication networks." *Proceedings of the National Academy of Sciences* 104 (18):7332–7336. URL `http://joi.jlc.jst.go.jp/JST.Journalarchive/kyubyochu1955/51.36?from=CrossRef`.

Park, Patrick S., Joshua E. Blumenstock, and Michael W. Macy. 2018. "The strength of long-range ties in population-scale social networks." *Science* 362 (6421):1410–1413. URL `http://www.sciencemag.org/lookup/doi/10.1126/science.aau9735`.

Perez, Sarah. 2018. "Zelle forecast to overtake Venmo this year." *TechCrunch* URL `https://techcrunch.com/2018/06/15/zelle-forecast-to-overtake-venmo-this-year/`.

Read, Simon. 2017. "Chat and pay: How social media is beating the banks." *BBC News* URL `https://www.bbc.com/news/business-42237432`.

Rothstein, Jesse M. 2006. "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions." *American Economic Review* 96 (4):1333–1350.

Simmel, Georg. 1907. *The philosophy of money.* London and Boston: Routledge and Kegan Paul.

Sixøj, Mads. 2016. "Swipp er færdig: De sidste banker skifter til MobilePay." *Berlingske* URL `https://www.berlingske.dk/virksomheder/swipp-er-faerdig-de-sidste-banker-skifter-til-mobilepay`.

Smith, Aaron and Monica Anderson. 2018. "Social Media Use in 2018." Tech. Rep. March, Pew Research Center.

Suri, Tavneet. 2017. "Mobile Money." *Annual Review ofEconomics* 9:497–520.

Townsend, Robert M. 1994. "Risk and Insurance in Village India." *Econometrica* 62 (3):539–591.

Ugander, Johan, Brian Karrer, Lars Backstrom, and Cameron Marlow. 2011. "The Anatomy of the Facebook Social Graph."

Veblen, Thorstein. 1899. *The Theory of the Leisure Class: An Economic Study of Institutions.* Macmillan.

Weber, Max. 1922. *Economy and society: An outline of interpretive sociology, Vols. I and II.* Berkeley: University of California Press.

Wilson, Christo, Bryce Boe, Alessandra Sala, Krishna P.N. Puttaswamy, and Ben Y. Zhao. 2009. "User interactions in social networks and their implications." *Proceedings of the fourth ACM european conference on Computer systems - EuroSys '09* :205URL `http://portal.acm.org/citation.cfm?doid=1519065.1519089`.

Zelizer, Viviana A. 1994. *The Social Meaning of Money.* New York: Basic Books.

———. 1996. "Payments and Social Ties." *Sociological Forum* 11 (3):481–495.

Tables and Figures

Table 1: Definitions of family and institutional networks.

Network		Members	Notes
Family	*Close:*	Parents, adoptive parents, siblings, half-siblings, and partners (measured at the beginning of the year).	I can identify parents and siblings if an individul's father or mother are alive and resident in Denmark in any year between 1980 and 2015. To identify grandparents and the other links, the same must be true for the individual's parents.
	Extended and step:	Step-parents, step-siblings, half-step-siblings, grandparents, cousins, aunts, uncles, parents of common children (whom are not partners).	
Education		Primary school, secondary school, vocational education institution, and university cohorts.	I can identify every individual in the same cohort at each institution, regardless of whether they completed their education. I also observe all individuals in the same graduating cohorts for the highest education obtained if that was before 1980.
Employer		Coworkers within same employer (CVR number) (measured at the end of year).	
Address		Roommates: individuals sharing an address (measured at the beginning of the year).	

Notes: This table contains details of the construction of the family and institutional networks from the population administrative registers.

Table 2: Representativeness: comparing the user population to the full population.

	Users	Population (ages 13+)	Users, ages 13-54	Population, ages 13-54
	(1)	(2)	(3)	(4)
		Age		
13-17	7.16	6.73	9.10	10.84
18-24	15.40	10.43	19.57	16.81
25-34	18.90	14.01	24.02	22.57
35-44	18.72	14.58	23.78	23.49
45-54	18.51	16.32	23.53	26.29
55-64	12.59	14.23		
65+	8.71	23.70		
		Sex		
Male	47.68	49.47	48.49	50.71
Female	52.32	50.53	51.51	49.29
		Highest completed education		
Compulsory schooling	26.84	30.32	28.68	28.72
Upper secondary/vocational	42.49	40.80	41.04	39.11
University	25.08	20.82	23.53	20.92
		Region of Denmark		
Capital	33.38	31.27	34.24	33.37
Central	22.82	22.53	22.94	22.78
North	9.53	10.34	9.53	9.90
South	20.53	21.28	20.06	20.42
Zealand	13.74	14.58	13.23	13.54
		Region of origin		
Danish	89.04	87.53	87.30	83.63
Non-Western	7.46	7.98	8.90	10.92
Westen (non-Danish)	3.50	4.50	3.80	5.45
		Wealth and income		
Net wealth:				
Danish kroner (DKK)	211,692.82	350,293.31	60,964.38	70,977.51
Percentile rank	50.59	50.50	50.14	50.50
Total income:				
DKK	320,861.01	291,867.76	299,905.79	283,100.52
Percentile rank	54.94	50.50	54.04	50.50
		Number of individuals		
	3,081,344	4,994,502	2,424,798	3,099,442

Notes: This table compares key features of the app user population with the full Danish population. Individuals aged 13+ with official residence in Denmark - and hence a Danish social security (CPR) number - are eligible to use the app. The data includes all users up until end of year 2016 and the full population at that point in time. I only include individuals in each group for whom I observe all variables in the left-most column. Numbers in cells are the percentage of each group in the relevant population, except for *Wealth and income* where the numbers correspond to average Danish krone or percentile rank. In *Column 1* I include all app users and compare values to the full poplation of individuals in Denmark aged 13+ (*Column 2*). In *Column 3* I limit the sample of app users to children, ages 13-17, and working age adults, ages 18-54, and compare to the same sample in the full population (*Column 4*). *Highest completed education* divides each population into three groups corresponding (approximately) to length of completed education: compulsory schooling from ages 6-15; 2-3 years of voluntary upper secondary or vocational schooling, and university-level education (from bachelors level up to doctorate, hence an additional 3+ years of education). *Region of Denmark* splits each population based on address into the the five geographical administrative regions of the country (Danish (English): Hovedstaden (Capital), Midtjylland (Central), Nordjylland (North), Syddanmark (South), Sjælland (Zealand)). *Region of origin* splits each population into individuals of Danish origin and immigrants and descendents from the Western and non-Western world. The definition of *Western* comes from Statistics Denmark and includes all EU-28 countries, Andorra, Iceland, Liectenstein, Monaco, Norway, San Marino, Switzerland, Vatican City, Canada, USA, Australia and New Zealand (and *Non-Western* includes all other countries). *Net wealth* is the 2015 end of year value of wealth net of liabilities. *Total income* includes labor market, transfer and capital income. *Percentile ranks* are calculated within age of birth cohorts.

Table 3: Summary statistics on relations in the network.

	All users		**Ages 13-54**	
	(1)		(2)	
	Mean	SD	Mean	SD
Degree	40.81	45.32	46.69	47.37
DKK transferred	40,897.55	43,101.34	45,562.10	44,815.44
Number of transfers	148.75	164.89	172.20	173.87
Reciprocal links	8.76	8.76	10.10	9.20
Percent reciprocal of:				
Degree	23.96	16.05	23.57	14.37
DKK transferred	57.87	29.54	59.81	28.14
Number of transfers	56.42	26.92	57.86	25.80

Notes: This table presents summary statistics on relations in the network for the full population of users (*Column 1*) and users aged 13-54 (*Column 2*) as of December 2016. I calculate the statistics on an undirected network derived from the raw transfer data. That is, I take all transactions between every pair of individuals and record the total value of money exchanged and the total number of transactions for each and every pair up until end of year 2016. *Degree* refers to the number of distinct individuals a person transfers money to or from. Similarly, *DKK transferred* and *Number of transfers* measures the monetary value and number of transfers an individual is a counterparty to, whether on the recieving or sending end of the transaction. *Reciprocal relationships* refers to relationships where money is transferred in both directions.

Table 4: Overlap of links in the network derived from mobile money transfers with real-world family and institutional ties.

	Unweighted		**Weight: DKK transferred**		**Weight: Number of transfers**		**Reciprocal links (unweighted)**	
	(1)		(2)		(3)		(4)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Panel A: Percentage with at least one link by group								
Family	95.36	21.04						
Education	88.35	32.09						
Employer	75.92	42.76						
Address	13.38	34.04						
Any real-world	100.00	0.00						
Panel B: Percentage of individuals' social circle by group								
Family	11.32	13.40	42.62	27.35	29.92	22.86	25.23	23.70
Education	28.17	24.20	19.33	20.45	27.99	25.48	33.57	29.32
Employer	16.35	17.01	9.53	13.84	13.84	16.88	15.68	20.32
Address	0.60	2.81	1.44	7.02	1.19	5.66	1.11	5.38
Any real-world	56.44	22.21	72.92	22.84	72.94	21.36	75.60	23.15
Number of individuals:			1483810				1426763	

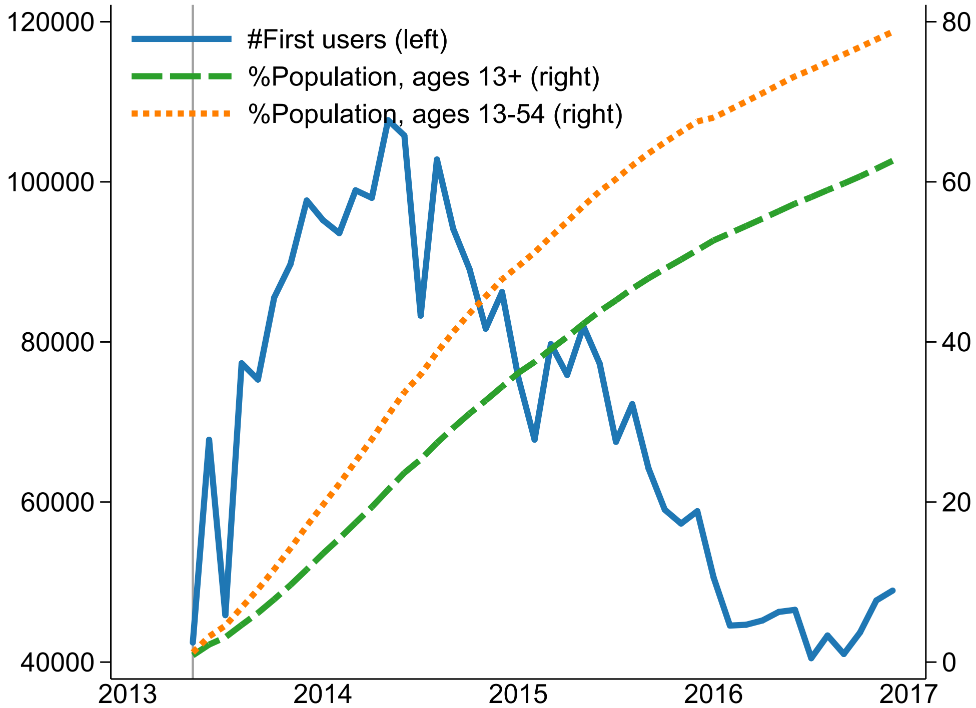
Notes: This table shows the overlap of the network formed from mobile money transfers with family and institutional networks and explores how this overlap varies with different measures of relationship intensity. *Column 1* reports results based on the unweighted (undirected) degree of each individual. *Columns 2-3* weight links by the amount transferred and the number of transfers. *Column 4* focusses on reciprocal links and includes only individuals with at least 1 reciprocal link (excluding 57,047 individuals). I limit the network to links between individuals born in Denmark after 1975. For this sample of users I observe full education, workplace, address and family histories. *Panel A* reports the percentage of individuals with at least one connection to someone from each, or any, of the family and insititional networks in their social circle. *Panel B* reports the average percentage (and SD of this value) of individuals' social circle accounted for by each, and any, of the family and institutional ties.

Table 5: Homophily of immigrants and their descendants.

		Percent in network population	Homophily: in-group share of social circle			Inbreeding homophily		
			Unweighted	Weight: DKK	Weight: Number	Unweighted	Weight: DKK	Weight: Number
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
(A) Immigrants and descendents of immigrants, excluding links between family members								
Origin:	Non-natives	10.42	0.47	0.54	0.52	0.40	0.49	0.46
	Global subregion		0.31	0.39	0.36	0.30	0.38	0.35
	Country		0.24	0.32	0.29	0.24	0.32	0.29
(B) All individuals, excluding links between family members								
Sex:	Men	47.60	0.62	0.67	0.66	0.28	0.37	0.35
Education:	Compulsory schooling	27.12	0.54	0.55	0.59	0.37	0.38	0.44
	Upper secondary/vocational	45.46	0.59	0.60	0.61	0.24	0.27	0.28
	University and above	27.42	0.52	0.51	0.54	0.34	0.33	0.36
Province (decreasing in population density):	Copenhagen City	14.64	0.58	0.61	0.62	0.51	0.54	0.56
	Greater Copenhagen	9.53	0.43	0.46	0.47	0.38	0.40	0.41
	North Zealand	8.21	0.54	0.56	0.57	0.50	0.52	0.53
	East Zealand	4.52	0.45	0.47	0.48	0.42	0.45	0.46
	East Jutland	15.90	0.75	0.75	0.78	0.70	0.71	0.74
	Funen	8.44	0.75	0.75	0.78	0.72	0.73	0.76
	West and South Zealand	9.34	0.66	0.68	0.69	0.63	0.65	0.66
	South Jutland	12.18	0.75	0.75	0.77	0.71	0.72	0.74
	North Jutland	9.57	0.77	0.78	0.80	0.75	0.75	0.78
	Bornholm	0.57	0.67	0.67	0.70	0.67	0.67	0.70
	West Jutland	7.09	0.69	0.70	0.72	0.67	0.67	0.70

Notes: This table reports statistics on homophily among non-natives (Panel A) and along dimensions commonly studied in the networks literature: sex, education and geography (Panel B). The left-most column splits the network population into groups. *Column (1)* reports the percentage of the network population in each group, which I present as the network population share of non-natives in *Panel A* since I average measured homophily across this group as a whole. *Columns (2)-(4)* report the average share of individuals' social circle who are in-group ties, by the groups defined on the left. *Column (2)* is unweighted and *Columns (3)-(4)* weight by the amount DKK transferred and the number of transfers. *Columns (5)-(7)* report average inbreeding homophily indices for each group, again in unweighted and weighted forms. The *Inbreeding homophily* index relates the observed excess propensity - the bias - of group members in linking to the the same group to the maximum possible bias they could exhibit. In *Panel A*, I present baseline and inbreeding homophily indices assuming all non-natives are "in-group" ties and then - in order to provide a summary - I report weighted averages of baseline and inbreeding homophily indices where I define the groups based on *Subregions* and *Countries*. A list of the subregions and top 100 countries among non-natives is provided in Table A1. In *Panel B:* For Sex, I only report statistics for men, since the values for women - the majority group - are co-determined. The Education grouping remains the same as reported in Table 1. For geography, I split the population by Province: Denmark is a country consisting of a connected land mass and several islands and this grouping reflects this diverse geography.

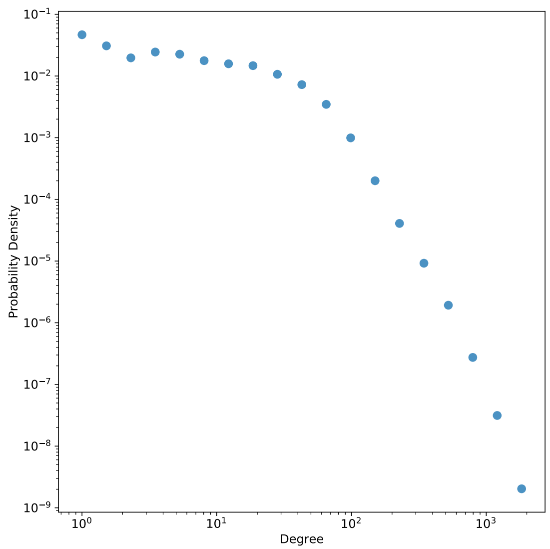
Figure 1: Adoption of the app since introduction.



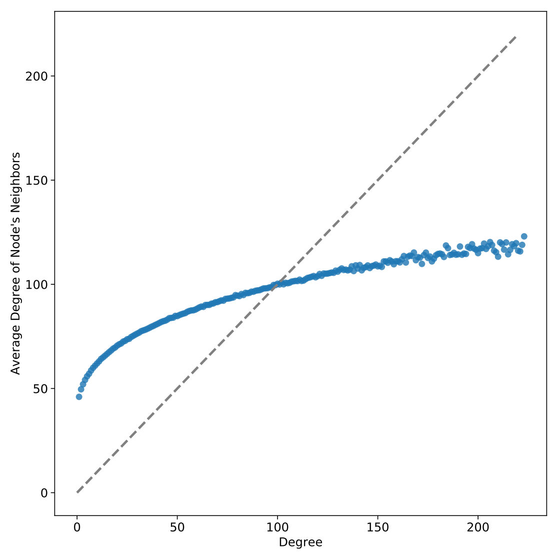
Notes: This figure shows the patterns of adoption of the money transfer app and the percentage of the population using the app at each point in time since introduction in May 2013 (marked by the grey vertical line) until December 2016. #First users (solid blue line, left vertical axis) records the number of individuals adopting the app in each month. %Population, ages 13+ (dashed green line, right vertical axis) records the percentage of the population aged 13 and above who are app users. %Population, ages 13-54 (dashed orange line, right vertical axis) records the percentage of the populations aged 13-54 years old who are app users. The data is based on the full sample of users, excluding merchants.

Figure 2: Structural properties of the social network.

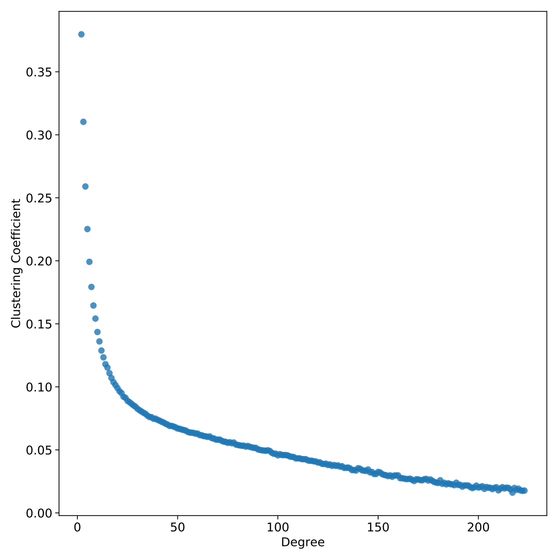
(a) Degree Distribution



(b) Degree Correlation



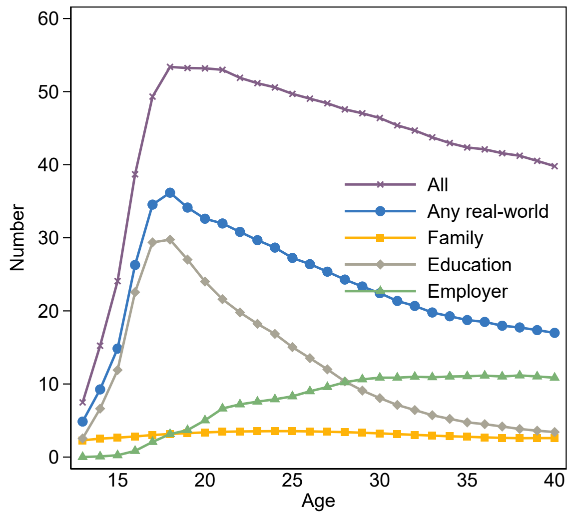
(c) Clustering



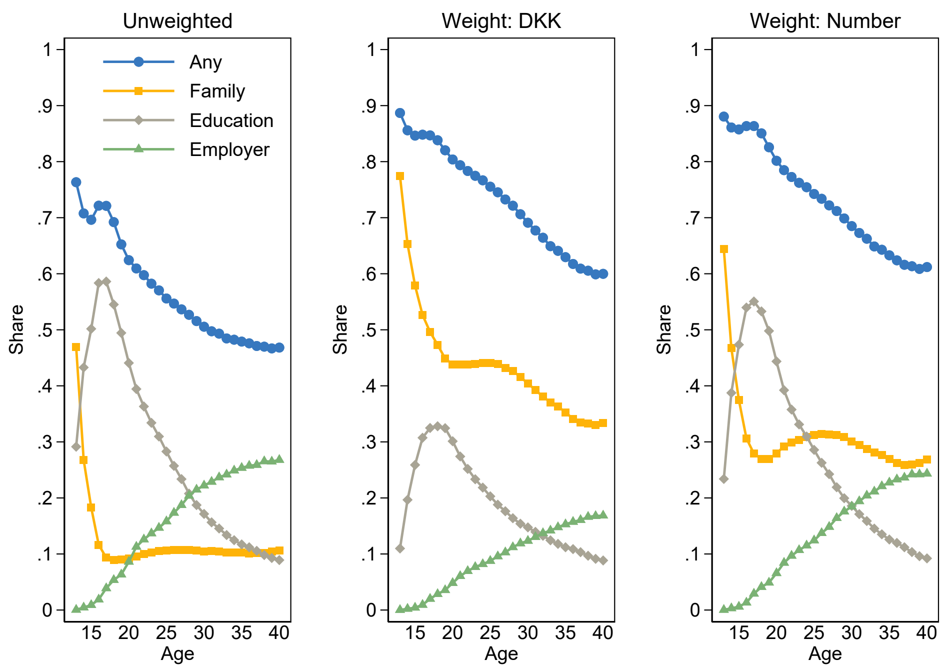
Notes: This figure shows the structure of the social network constructed from mobile money transfers. I convert the records on all transfers into an undirected network. That is, I take all transactions between every pair of individuals and record the total value of money exchanged and the total number of transactions for each and every pair up until end of year 2016. In Panel A, I plot the distribution of degrees across all individuals in the network. An individual's degree is the number of people that the individual transfers money with, whether receiving or sending. The degree distribution on a linear-scale - plotting the relative frequency of individuals at each degree - is extremely skewed: very many individuals have only a few links and a few hubs exist with a large number of links. In order to more easily assess the relative number of individuals of each degree - all the way up to very extreme values - I put the data into bins that increase in size with degree and then I plot the density of individuals within each bin, scaled by the width of the bin, on a logarithmic scale. In Panel B, I show the correlation in degree between individuals and the individuals they link to in the network, their social circle or neighbors. I calculate the average degree of an individual's neighbors and plot the average of this value for nodes of each degree. As with Panel C, I limit the plot to show statistics up to a degree of 5.5 times the average degree in the network. The grey dashed line is the 45-degree line; the points cross this line at a degree equivalent to 2.5 times the average degree. In Panel C, I plot the average clustering coefficient by degree. For a given individual the clustering coefficient is the proportion of all possible pairs of that individual's neighbors that are connected to each other, that is, the proportion that are also neighbors.

Figure 3: The life cycle of family and institutional ties.

(a) Number of Ties



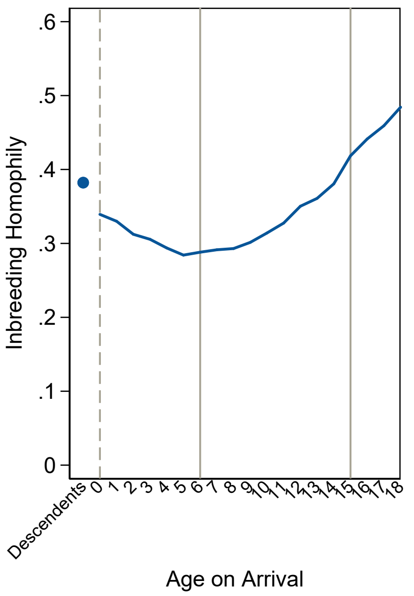
(b) Importance of Ties



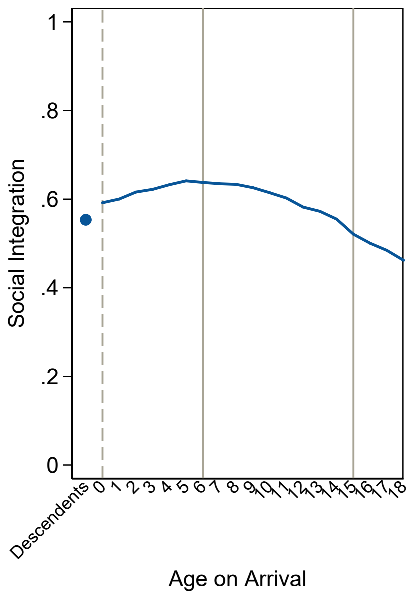
Notes: This figure shows the average number (Panel A) and share (Panel B) of different family and institutional ties in an individual's social circle by age. I focus on four types of real-world network: family, education, workplace and shared address. I limit the network to links between individuals born in Denmark after 1975. For this sample I observe full education, workplace, address and family histories. I don't include results for the address network since the average number of links to this network is low. In Panel A, the purple line with circle markers records the average degree for individuals of each age and the other lines report the average degrees attributed to each real-world network and any of the real-world networks combined. In Panel B, the blue line with circle markers reports the average share of links that can be attributed to any of the four real-world networks. The leftmost plot is unweighted, the middle plot is weighted by the DKK value transferred and the right plot is weighted by the number of transactions. The other lines in Panel B report the share of total links attributed to each real-world network.

Figure 4: Variation in homophily and integration by age on arrival and region of origin.

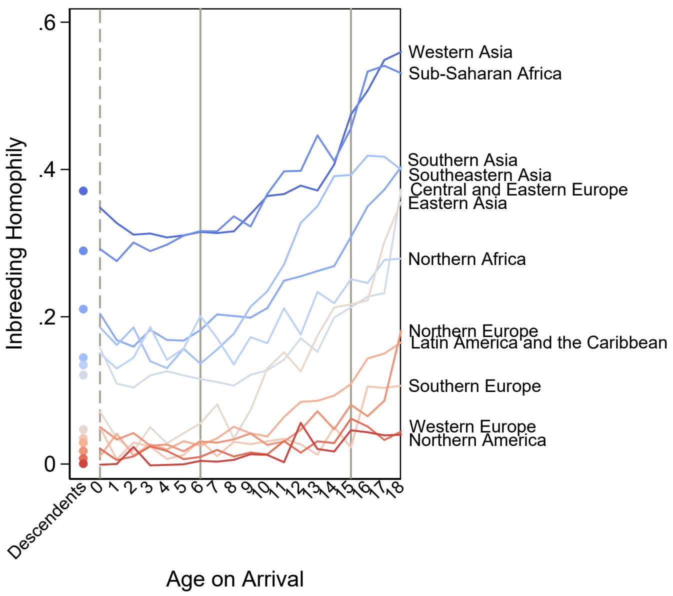
(a) Inbreeding Homophily (In-Group: All Non-Natives)



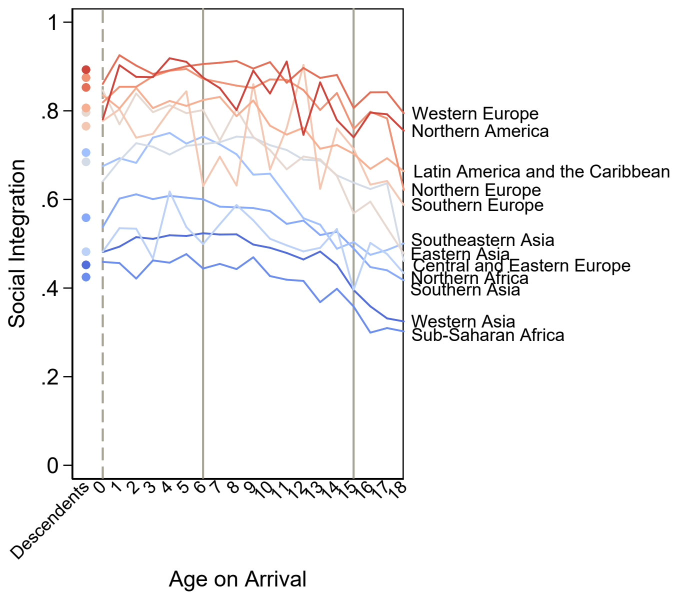
(b) Social Integration (Share Natives in Social Circle, All Non-Natives)



(c) Inbreeding Homophily (In-Group: All Non-Natives)

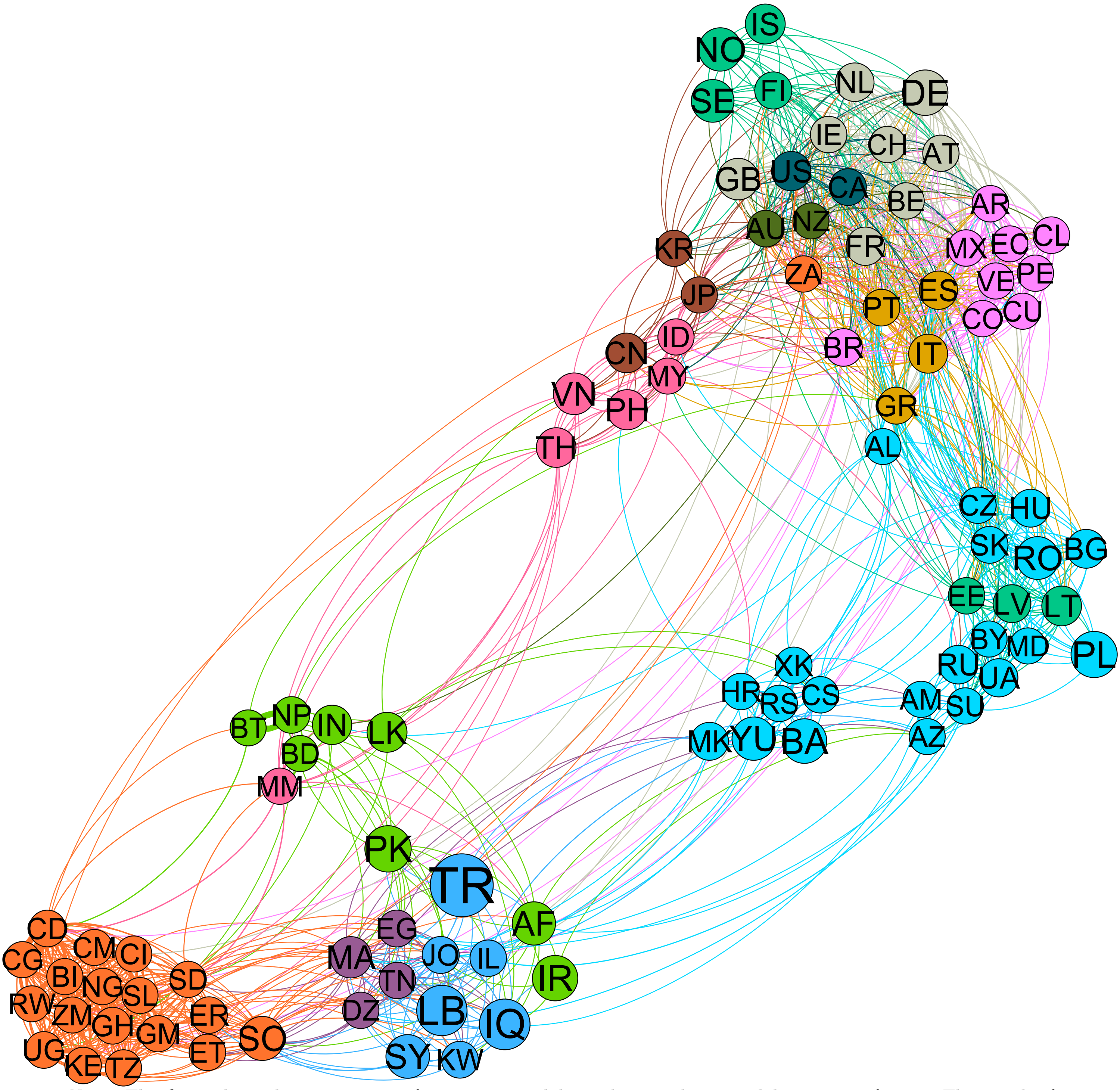


(d) Social Integration (Share Natives in Social Circle, By Subregion)



Notes: This figure shows variation in the extent of homophily and integration across immigrants and descendents by age on arrival and by region of origin. The sample includes all descendents of immigrants and all immigrants who arrived in Denmark up until the age of 18 who were aged 25-55 as of December 2016 (removing the age restriction or narrowing it to 25-35 makes no difference to the results). I exclude links between family members. Panel A reports the average inbreeding homophily index (vertical axis) for descendents (the marker point) and immigrants of different arrival ages (the connected line). In-group ties are defined as links to other immigrants or descendents (non-natives). That is, the index measures the amount of bias that descendents and immigrants of different arrival ages have towards forming links with any other immigrants or descendents relative to the maximum possible bias they could have based on the share of all immigrants and descendents in the network population. In Panel B, I report average social integration - measured as the share of native Danes in individuals' social circles - for descendants and within immigrants by age on arrival. In Panel C, I report average inbreeding homophily where I split immigrants and descendents by their global region of origin and change the definition of in-group ties to links to other immigrants or descendents from the same global region. I apply a diverging color scheme, based on the level of homophily among descendants. In Panel D, I split by region of origin and plot average integration for each, maintaining the color scheme. The solid vertical grey lines mark school starting age - 6 years old - and the final year of compulsory schooling at age 15.

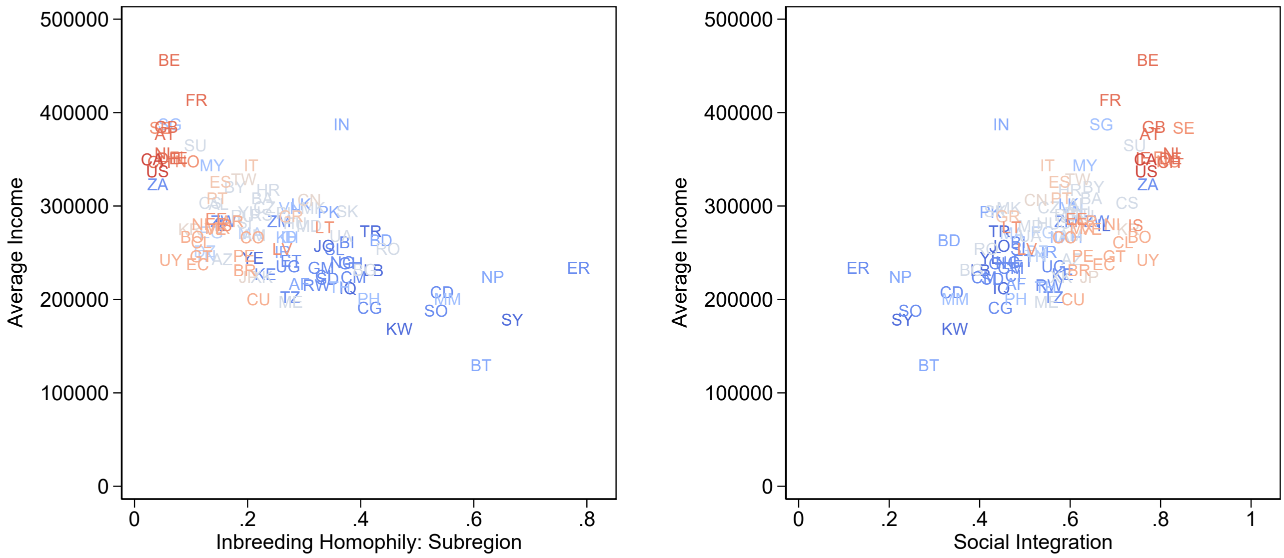
Figure 5: Immigrant communities by country of origin.



Notes: This figure shows the communities of immigrants and descendents in the network by country of origin. The sample of the network used to construct the figure includes links between immigrants from, or descendents of immigrants from, the 100 largest countries of origin in the network. I use the force atlas algorithim in Gephi to automatically determine the position of the countries relative to each other. Simply put, the algorithm pulls high interaction countries together and repels low interaction countries. The exact details of the alhorithm can be found in the paper by Jacomy et al. (2014). The colors of each country denote the global subregion the country belongs to. The links between each country are weighted. Taking individuals from Turkey (TR) as an example, the weight of the link between TR and Afghanistan (AF) is the fraction of links from individuals of Turkish origin to the 99 other countries of origin that are with individuals from AF, scaled by the fraction of individuals from the 100 largest countries of origin in the network who are from AF. To make the graph readable, I only include links with a weight of more than 1.05. The color of the link is determined by the subregion that the link originates from and the direction of the link is clockwise.

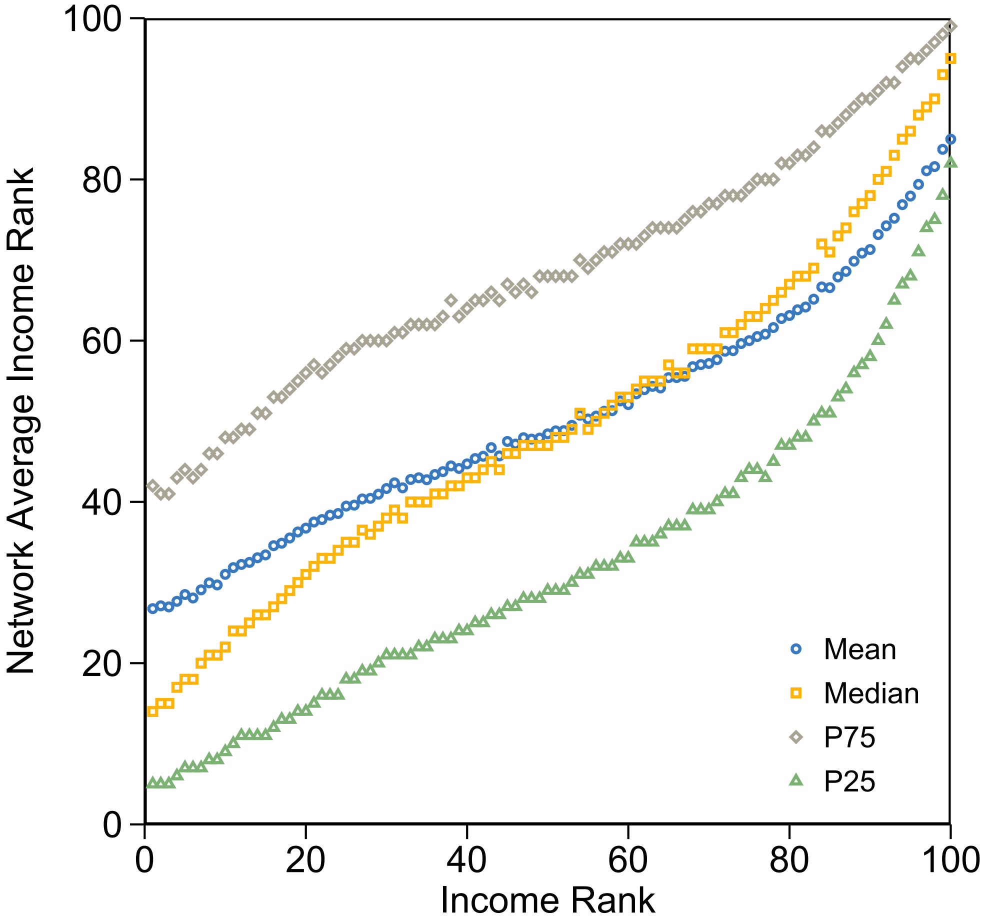
Figure 6: Homophily, social integration and economic integration, by country of origin.

(a) Inbreeding Homophily (In-Group: Subregion) (b) Social Integration (Share Natives in Social Circle)



Notes: This figure correlates average annual incomes for prime-working age adults among immigrants to Denmark and their descendents (non-natives) with measures of inbreeding homophily (Panel A) and social integration (Panel B). The measure of social integration is the average share of total links that are with Danes among individuals with each country of origin. I focus on individuals from the top 100 largest countries of origin among Danish immigrants and descendents. The income measure is constructed for prime-working age individuals but the social network statistics are measured as averages by country of origin among all individuals in the network. I color each point by the relevant global subregion. The coefficient on a regression of average income on the average share of Danish links - weighted by group size - suggests that a 10% higher share of links to Danes is correlated with average incomes that are 27,430 DKK higher.

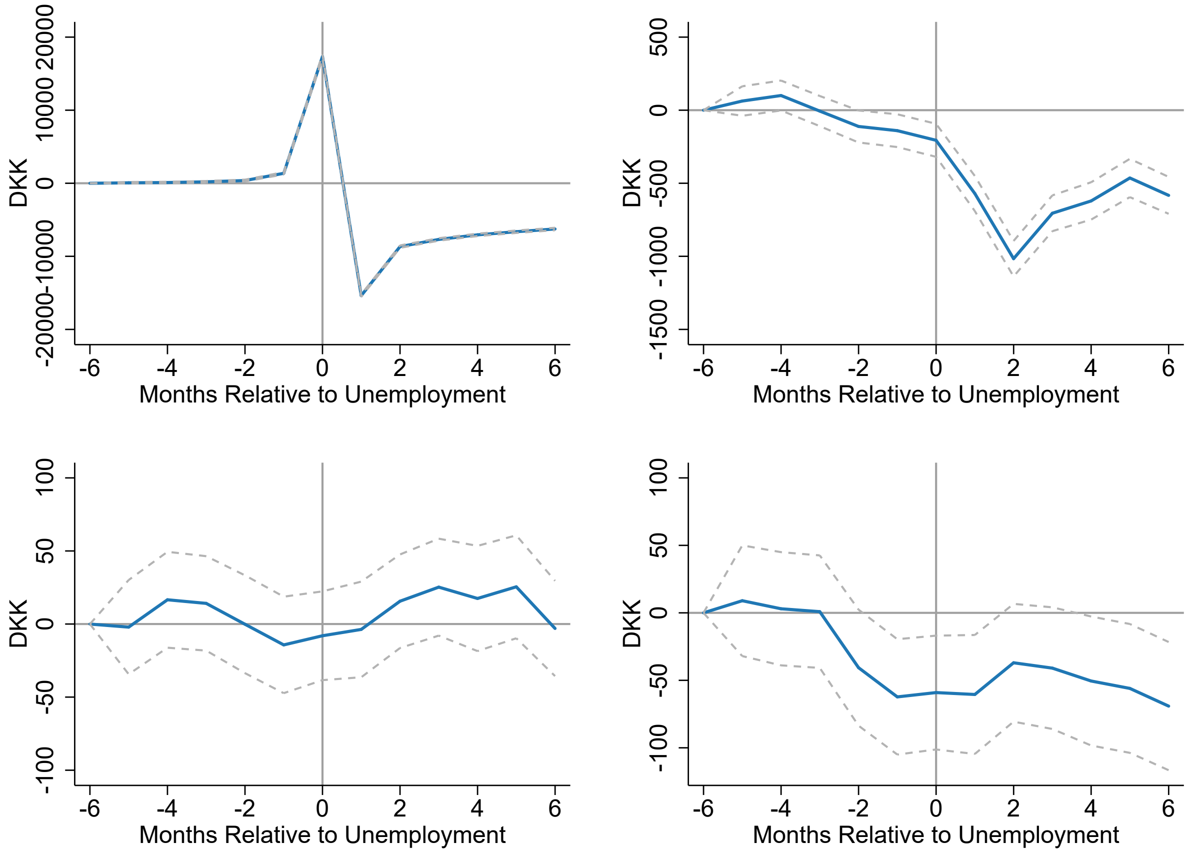
Figure 7: Social network stratification by economic status.



Slopes:
Mean: 0.50, Median: 0.69, P75: 0.51, P25: 0.63

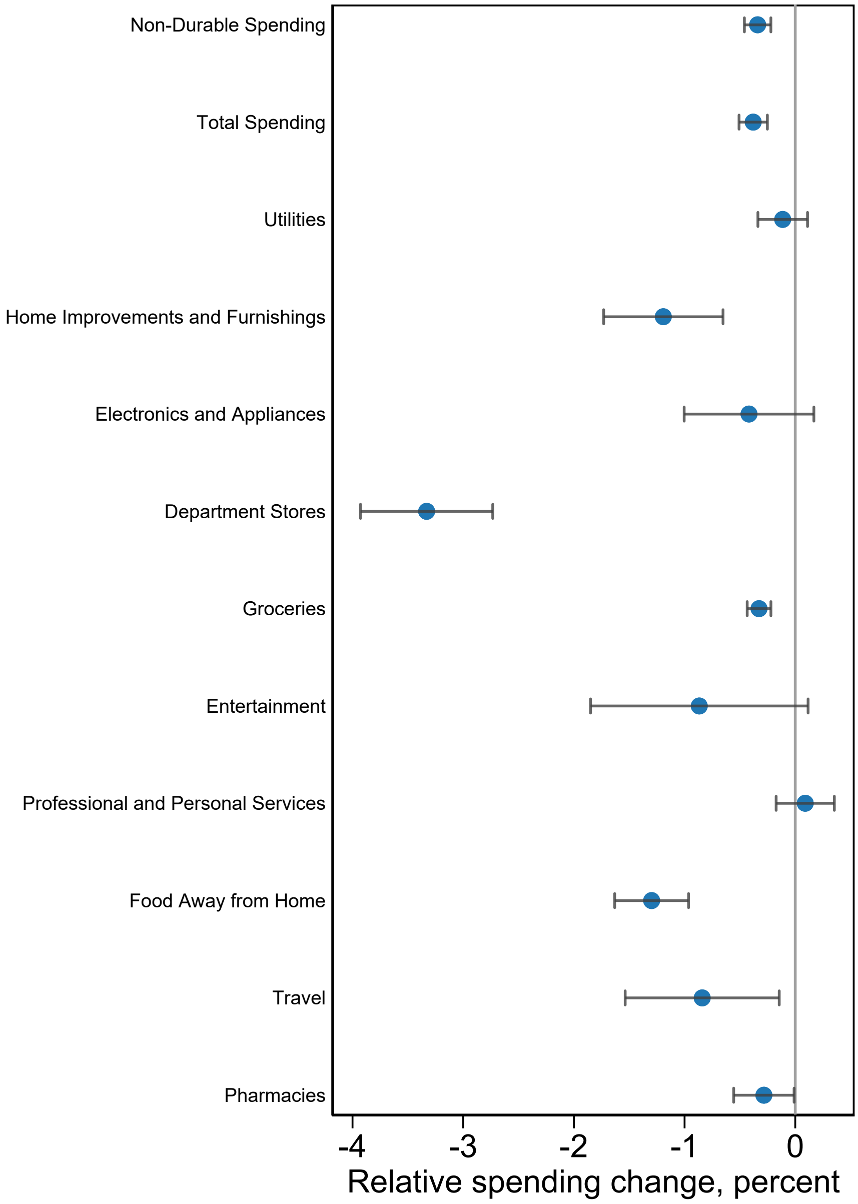
Notes: This figure shows the correlation between individual percentile income rank and the percentile rank of average income among individuals' social circles (network average income), controlling for sex and birth cohort effects. I focus on the subgraph of the network containing links between individuals aged 30 to 40 years old in December 2016 who are not family members. This subgraph of the full mobile money transfer network contains 578,226 individuals and 6.54 million links. For these individuals, I calculate their average annual total income over three years from 2013 to 2015, where total income includes labor market income, transfer income, and asset income. I refer to this measure of income as y_i. I regress y_i on birth cohort indicators, a sex indicator, and interactions between birth cohort indicators and sex. For each individual, I use these estimates to calculate predicted income based on sex and birth cohort $\hat{y}_i$ and unexplained, or residual, income as $y_i^{res} = y_i - \hat{y}_i$, measuring the portion of individual income not explained by birth cohort and sex. I then calculate, for each individual i, the average of y_i^{res} among their social circle, ${y_i}^{social}$, which I refer to as network average income. I then calculate percentile ranks of y_i^{res} and ${y_i}^{social}$ as p_i^{res} and p_i^{social} respetively and present 4 properties of their joint distribution in the figure: the conditional expectation of p_i^{social} given p_i^{res}, the median p_i^{social} given p_i^{res}, and the 25^{th} and 75^{th} percentiles of p_i^{social} at each p_i^{res}. All reported percentiles contain more than 100 individuals in order to maintain anonymity. The reported slopes are based on OLS regressions.

Figure 8: Event-study estimates of social influence on spending.



Notes: This figure shows estimates from event-study regressions for individuals who lose their jobs (Top row) and individuals who experience a social connection losing their job (Bottom row). The First column shows results for income and the Second column shows results for household non-durable expenditure. The solid blue line in the bottom row correspond to the estimated coefficients from regressions of the form given by equation X in the main text. The estimates on the top row are from an analagous regression but where the unit of observation is the individual experiencing the shock. The dashed grey lines represent 95% confidence intervals. Income is defined as monthly pre-tax income, including transfers. Non-durable expenditure includes card and bill payments for non-durable goods and services and cash withdrawals. Both variables are measured in Danish kroner (DKK). I omit the indicator for 6 months prior to job loss.

Figure 9: Summarizing social influence across types of spending.



Notes: This figure shows event-study estimates for components of total spending for households exposed to a job loss shock in their network. I collapse the event-time indicators into a single indicator identifying time periods after the shock and I present the estimates - shown by blue markers - and confidence intervals - grey capped bars - measured relative to average spending in each category 6 months prior to the event. The grey vertical line marks zero relative spending change. I measure non-durable spending as described in the text. Total spending is the sum of non-durable spending and durable spending. The other spending components are based on card and bill payments, using Merchant Category Codes (MCC) and a proprietary categorization respectively in order to construct each measure.

Appendix

Comparing the Structural Properties of the Network With the Facebook Network

The origins of the Facebook and the mobile money transfer networks are fundamentally different. Despite this, these networks are strikingly similar with respect to several of the structural properties examined here. First, the degree correlations are almost identical: in both networks, progressing from a degree of 1 to the 95th percentile of the degree distribution correlates with an increase in the average degree of individual's neighbors from the average degree to the 95th percentile of the degree distribution (see Figure 1 Panel B in Bailey et al. (2018)). Second, and following from the first point, it is only past the 95th percentile of the degree distribution that individuals tend to have fewer neighbors than their neighbors do. In other words, the extent of the "friendship paradox" – the observation that, on average, people are less popular than their friends (Feld, 1991; Jackson, 2017) – is almost identical in both networks. Third, the average clustering coefficient in both networks is identical to the decimal point (see Table 1 in Bailey et al. (2018)).

Overall, these similarities hint at universal structural properties of population-scale networks and fundamental processes governing their formation.

There are, however, some structural differences between the two networks. The shape of the degree distribution in the Facebook network, documented in Figure 1 Panel A in Ugander et al. (2011) and the same in Bailey et al. (2018), exhibits more curvature up to high degrees than in the mobile money transfer network. Moreover, there is bunching in the Facebook degree distribution at 5,000, the limit for the number of friends. The mobile money transfer app does not limit connectivity.

Finally, degrees are considerably higher, on average, in the Facebook network than in the mobile money transfer network. Relevant to this last point, a link in the mobile money transfer network represents a real-world interaction and it is, perhaps, more appropriate to compare the degree distribution in the mobile money transfer network to the distribution of the number of individuals a Facebook user actually interacts with on the platform, for example, through messaging or commenting. I am not able to find this distribution for the full Facebook network, but the analysis by Wilson et al. (2009) suggests the median number of individuals interacted with on Facebook is less than 10. Relative to the median Facebook user's total number of connections, this value is much more similar to the median degree in the mobile money transfer network. Jones et al. (2013) find that interaction on Facebook is highly predictive of real-world friendship. Combined, these considerations suggest that data from mobile money transfer networks might be more reflective of real-world social network degree distributions than Facebook friendships are, and hence these are a valuable source of data for learning about the universal structural properties of population-scale networks.

Appendix Tables and Figures

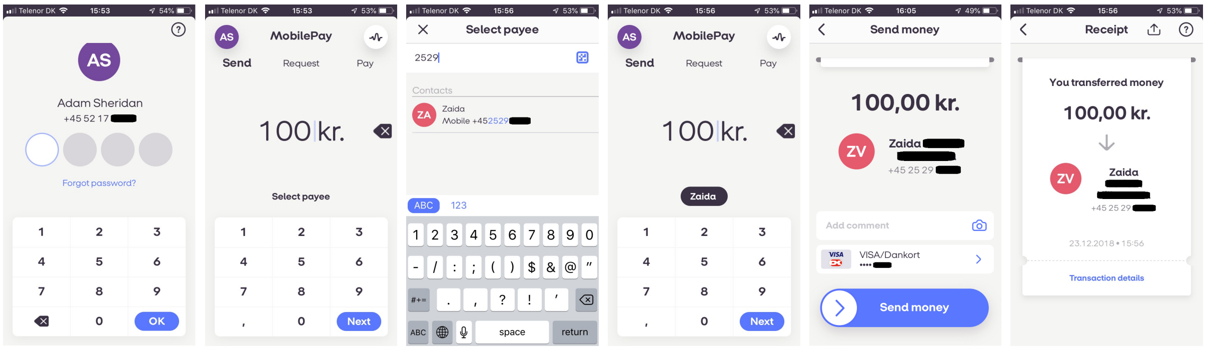
Table A1: Country-of-origin and region-of-origin definitions.

Country-of-origin	ISO alpha-2	Region-of-origin	Percent of network population
Denmark	DK	Denmark	89.58
New Zealand	NZ	Australia and New Zealand	0.05
Australia	AU		
Albania	AL	Central and Eastern Europe	2.05
Azerbaijan	AZ		
Moldova, Republic of	MD		
Belarus	BY		
Armenia	AM		
Soviet Union	SU		
Serbia	RS		
Czech Republic	CZ		
Czechoslovakia	CS		
Croatia	HR		
Kosovo	XK		
Slovakia	SK		
Macedonia, the Former Yugoslav Republic of	MK		
Hungary	HU		
Russian Federation	RU		
Ukraine	UA		
Bulgaria	BG		
Yugoslavia, Federal Republic	YU		
Romania	RO		
Bosnia and Herzegovina	BA		
Poland	PL		
Estonia	EE	Northern Europe	1.07
Finland	FI		
Latvia	LV		
Lithuania	LT		
Iceland	IS		
Sweden	SE		
Norway	NO		
Portugal	PT	Southern Europe	0.28
Greece	GR		
Spain	ES		
Italy	IT		
Belgium	BE	Western Europe	0.97
Austria	AT		
Switzerland	CH		
Ireland	IE		
France	FR		
Netherlands	NL		
United Kingdom	GB		
Germany	DE		
Korea, Republic of	KR	Eastern Asia	0.20
Japan	JP		
China	CN		
Malaysia	MY	South-Eastern Asia	0.67
Indonesia	ID		
Myanmar	MM		
Thailand	TH		
Philippines	PH		
Vietnam	VN		
Bhutan	BT	Southern Asia	1.33
Bangladesh	BD		
Nepal	NP		
India	IN		
Sri Lanka	LK		
Afghanistan	AF		
Iran, Islamic Republic of	IR		
Pakistan	PK		
Israel	IL	Western Asia	2.42
Kuwait	KW		
Jordan	JO		
Syrian Arab Republic	SY		
Lebanon	LB		
Iraq	IQ		
Turkey	TR		
Ecuador	EC	Latin America and the Caribbean	0.25
Cuba	CU		
Venezuela, Bolivarian Republic of	VE		
Peru	PE		
Colombia	CO		
Mexico	MX		
Argentina	AR		
Chile	CL		
Brazil	BR		
Algeria	DZ	Northern Africa	0.26
Tunisia	TN		
Egypt	EG		
Morocco	MA		
Sierra Leone	SL	Sub-Saharan Africa	0.67
Rwanda	RW		
Zambia	ZM		
Côte d'Ivoire	CI		
Cameroon	CM		
Congo	CG		
Sudan	SD		
Burundi	BI		
Tanzania, United Republic of	TZ		
Gambia	GM		
South Africa	ZA		
Nigeria	NG		
Ethiopia	ET		
Kenya	KE		
Uganda	UG		
Congo, the Democratic Republic of the	CD		
Ghana	GH		
Eritrea	ER		
Somalia	SO		
Canada	CA	Northern America	0.17
United States	US		

Notes: This table provides information on the largest 100 country-of-origin among immigrants and descendents in the network. I provide the ISO 3166-1 alpha-2 code - two-letter country codes - for each country as published by the International Organization for Standardization (ISO). A number of countries in this list no longer exist - SU, YU - and for these I use the last reported ISO alpha-2 code. I adopt XK for Kosovo, in accordance with the European Commission (https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Country_codes). I define global subregions in two ways. For countries on the European continent - exlcuding Denmark - I use the EuroVoc geographical classification (http://eurovoc.europa.eu/drupal/?q=request&uri=http://eurovoc.europa.eu/100277). EuroVoc is multilingual thesaurus maintained by the European Union. I group all other countries using subregions as defined by the UN Statistics Division "Standard country or area codes for statistics use" (https://unstats.un.org/unsd/methodology/m49/overview/).

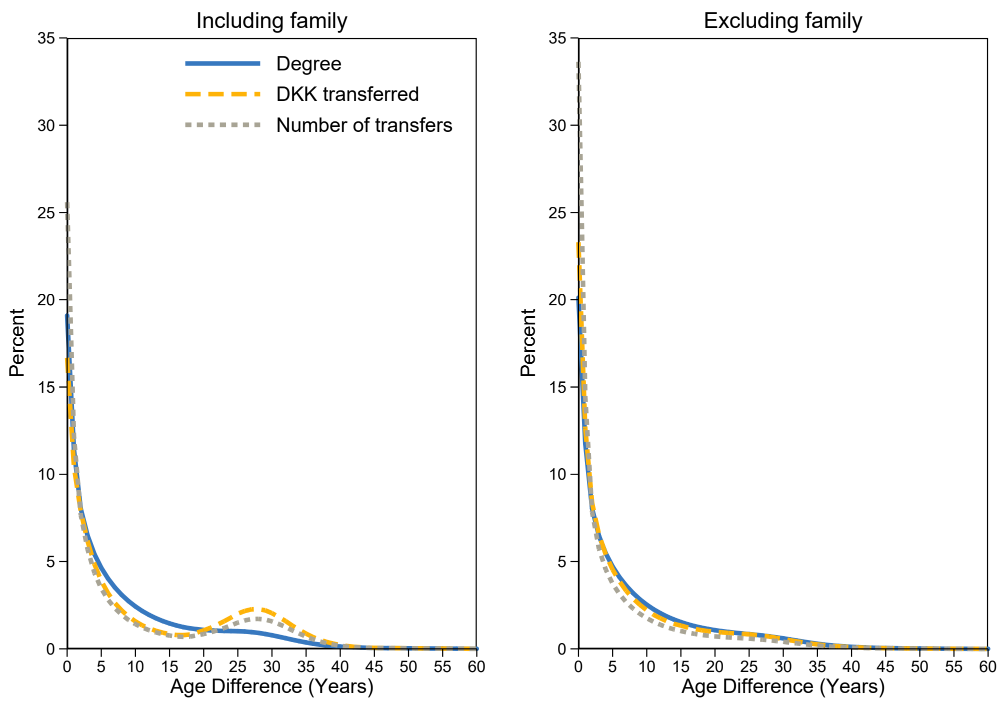
Figure A.1: The MobilePay transfer flow.

(a) Step 1 (b) Step 2 (c) Step 4 (d) Step 4 (e) Step 5 (f) Step 6



Notes: This figure presents screenshots of the user flow for making a transfer. In Step 1, the user logs in using a 4-digit code. In Step 2, the user enters they amount of money they would like to transfer. In Step 3, the user chooses who they would like to transfer the money to from their phone contacts (or through entering a phone number). In Step 4, the user can make final adjustments to the amount to send and presses 'Next' to proceed to payment. In Step 5, the use chooses their payment card, optinally adds a message and/or photo, and swipes right to transfer the amount. In Step 6, the user gets a receipt for the transfer.

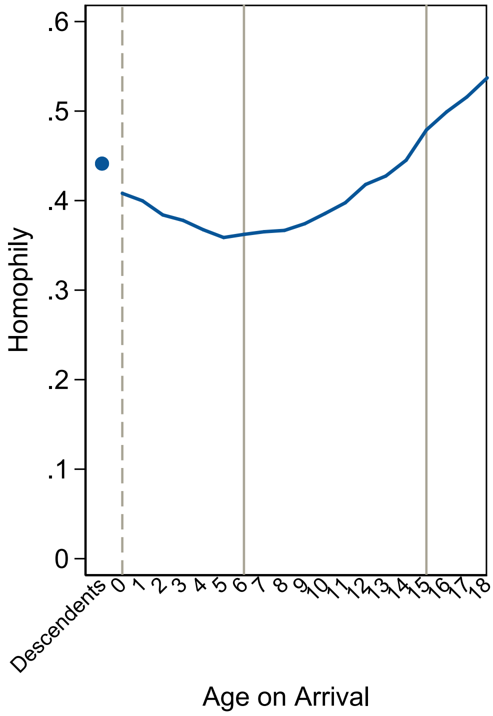
Figure A.2: Age homophily.



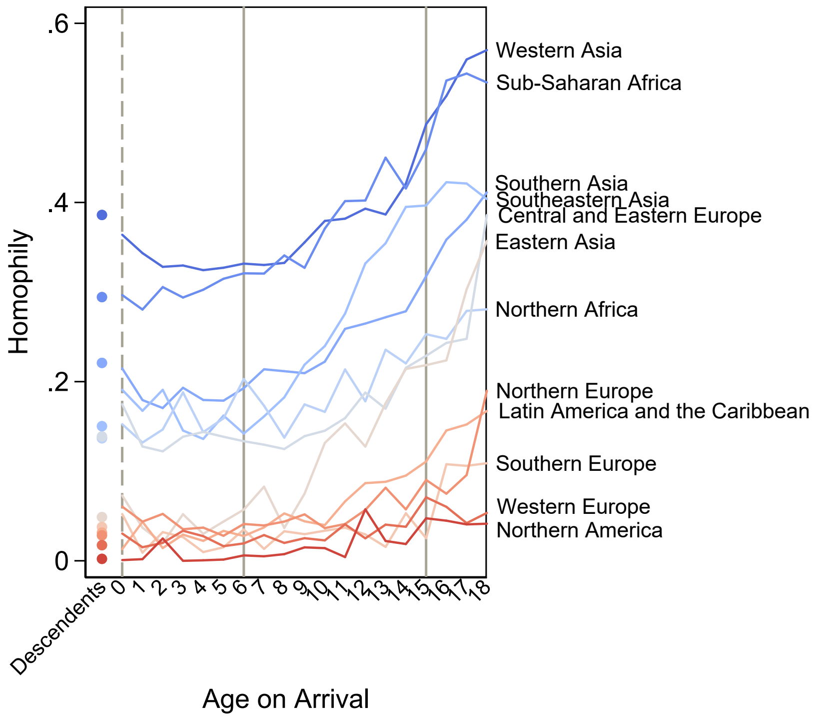
Notes: This figure shows the percentage of links (solid blue line, Links), DKK transferred (dashed yellow line, DKK transferred) and number of transfers (dotted grey line, Number of transfers) by the age difference between the counterparties in each link. Panel A includes links between family members and Panel B excludes family links, removing the bimodality.

Figure A.3: Variation in homophily using the basic homophily index.

(a) Homophily (In-Group: All Non-Natives)



(b) Homophily (In-Group: Region)



Notes: This figure repeats Panel A and Panel C from Figure 4 using the basic homophily index, rather than the inbreeding homophily index.