



# HHS Public Access

Author manuscript

*Migr Stud.* Author manuscript; available in PMC 2016 October 12.

Published in final edited form as:

*Migr Stud.* 2016 July ; 4(2): 182–214. doi:10.1093/migration/mnw004.

## Sampling Migrants from their Social Networks: The Demography and Social Organization of Chinese Migrants in Dar es Salaam, Tanzania

**M. Giovanna Merli,**

Duke Population Research Institute, Sanford School of Public Policy, Department of Sociology and Duke Global Health Institute, Duke University, Box 90312, Durham, NC 27708

**Ashton Verdery,**

Department of Sociology & Criminology, Population Research Institute and Institute for CyberScience, The Pennsylvania State University, University Park PA, USA

**Ted Mouw,** and

Department of Sociology & Carolina Population Center, University of North Carolina, Chapel Hill NC, USA

**Jing Li**

National Center for STD Control, Nanjing, China

M. Giovanna Merli: giovanna.merli@duke.edu

### Abstract

The streams of Chinese migration to Africa are growing in tandem with rising Chinese investments and trade flows in and to the African continent. In spite of the high profile of this phenomenon in the media, there are few rich and broad descriptions of Chinese communities in Africa. Reasons for this include the rarity of official statistics on foreign-born populations in African censuses, the absence of predefined sampling frames required to draw representative samples with conventional survey methods and difficulties to reach certain segments of this population. Here, we use a novel network-based approach, Network Sampling with Memory, which overcomes the challenges of sampling ‘hidden’ populations in the absence of a sampling frame, to recruit a sample of recent Chinese immigrants in Dar es Salaam, Tanzania and collect information on the demographic characteristics, migration histories and social ties of members of this sample. These data reveal a heterogeneous Chinese community composed of “state-led” migrants who come to Africa to work on projects undertaken by large Chinese state-owned enterprises and “independent” migrants who come on their own accord to engage in various types of business ventures. They offer a rich description of the demographic profile and social organization of this community, highlight key differences between the two categories of migrants and map the structure of the social ties linking them. We highlight needs for future research on inter-group differences in individual motivations for migration, economic activities, migration outcomes, expectations about future residence in Africa, social integration and relations with local communities.

## Keywords

Africa; China; International Migration; Social networks; Survey Sampling

---

## 1. INTRODUCTION

Since the end of the 1990s, Chinese migration to Africa has increased rapidly. Statistics on the volume of this migration and on the number of Chinese currently living in various African countries remain elusive. Most of what is known relates to countries with a long history of immigration from China, e.g., South Africa, Madagascar and Mauritius. Estimates of the total number of Chinese migrants in Africa range widely, from around 580,000 to over one million (Park 2009; Li 2010; Sun 2014; Mohan et al. 2014). This uncertainty is proportionately larger for individual countries. Some of this uncertainty owes to endemic problems in African censuses and statistical systems; some is particular to the case of migration in Africa – lax immigration policies, poor tracking mechanisms, and corruption that facilitates illegal migration (Park 2009).

Knowledge on the characteristics of Chinese migrants in Africa is mostly based on the findings of in-depth, ethnographic studies of Chinese communities in selected African countries. From these, we know that some Chinese migrants arrive in Africa under formal labor contracts linked to publicly-funded Chinese investment projects, with fixed expectations regarding departure and length of stay, but the situation is more nebulous for others who go to Africa on their own accord to undertake various types of business ventures. Scholarly and journalistic accounts of China's human presence in Africa generally describe these two groups in isolation from each other. They either focus on Chinese workers recruited for large-scale investments and financial projects led by China's state-owned enterprises (SOEs; e.g. Lee 2009) or they describe the activities of the countless smaller actors operating independently from the state (French 2014; Haugen and Carling 2005; Lin 2014).

But the proliferation of China's economic cooperation, business and commercial activities in Africa has broadened the scope of China's presence beyond the inter-governmental realm. New detailed descriptions and comparisons between groups of migrants and across study sites are needed to characterize the heterogeneity of China's human presence in Africa. The few studies (e.g. Mohan et al. 2014) that have identified different migrant typologies among recent cohorts of migrants from mainland China to Africa, distinguishing state-led from independent migrants and describing social relations in Chinese communities across multiple African contexts, have been hampered by difficulties to gain access to formal state-led projects and have relied on the use of ad hoc sampling approaches that impede population inference and prevent meaningful comparisons between migrant types and across study sites.

Previous efforts to provide quantitative descriptions of Chinese migrants in Africa reveal the challenges of recruiting useful analytical samples of this population. First, the Chinese government is very sensitive to representations by scholars and the media alike of the role it plays in African development, especially in light of international criticism of the labor

relations between Chinese management and African workers on sites managed by Chinese companies in Zambia (Human Rights Watch 2011) and the specter of Chinese imperialism raised in the U.S. and Western Europe (Lee 2009). Based on our own experience, Chinese embassies in Africa are reluctant to share lists of Chinese companies and their employees, complicating access to formal state-run projects in Africa.

Second, Chinese migrant populations in Africa are hidden and rare. Hidden populations (Heckathorn 1997) are those that lack a sampling frame which would allow one to draw conventional probability samples that yield accurate descriptions of and comparisons between the population groups under study. Lists of Chinese in Africa from which to draw a probability sample are not available due to incomplete national statistical systems, absent official registries of foreign born persons and the undocumented status of many Chinese visitors who overstay their tourist visa. If these lists exist, they are very difficult to access. Chinese migrants in Africa are also likely to represent a small proportion of the general population. This rarity implies that the sample size required to ensure inclusion of a minimum number of migrants for analytical samples for inference would have to be extremely large with obvious cost considerations. Even when they are found, migrants may be reluctant to speak with interviewers without receiving prior information about the study from people they trust.

Here we use a novel network-based sampling approach, Network Sampling with Memory (NSM, Mouw and Verdery 2012), designed to circumvent the challenges to sample rare and hidden populations. Between June and July 2013, we used NSM to recruit a sample of 147 Chinese respondents living in Dar es Salaam, Tanzania for the Chinese in Africa Health Study (CAHS).<sup>1</sup> Respondents provided their own demographic, social and health information as well as basic demographic information for a total of 853 unique members of their Chinese social networks. Altogether, the data collected provide a rich and broad description of one Chinese community in Africa. They allow a characterization of the heterogeneity of China's human presence in an African country; they identify the presence of migrants sent by the state and those who have moved of their own accord including variation in demographic characteristics, employment activities and migration histories and differences in these characteristics between categories of migrants. A byproduct of NSM is the collection of extensive network data which allow one to map the social organization of this Chinese community, especially the relational and structural features of their co-ethnic social networks as they are patterned by their economic activities and province of origin in China. In this paper, we first review the literature on China's presence in Africa. We then introduce the new sampling approach for hidden and rare populations of migrants, and describe its implementation in a Tanzanian setting and the methods used to analyze the data. This is followed by a description of the characteristics of distinct categories of migrants, their migration histories and their patterns of social relations as they are structured by

---

<sup>1</sup>The CAHS survey was undertaken to explore linkages between migrant status, material circumstances, awareness and knowledge of major infections in the new disease environment, health-related behaviors (risky sexual behaviors, protective behaviors), utilization of health care, health insurance, and health outcomes (sexual health, mental health, chronic and infectious disease conditions) of Chinese migrants in Africa.

employment sector and Chinese province of origin. We conclude with a discussion of avenues for future research implied by our findings.

## 2. CHINA'S PRESENCE IN AFRICA AND TANZANIA

A review of the extant scholarly literature, mostly grounded in data collected through ethnographic research or ad hoc, convenience samples (see *inter alia* Haugen and Carling 2005; Lee 2009; Lin 2014; Mohan et al. 2014) and of journalistic reports (e.g. French 2014) on Chinese migration to Africa suggests two broad categories of migrants. The first group consists of “state-led” migrants, who arrive under formal one-to-three year-long labor contracts linked to projects undertaken by large Chinese SOEs, with operations in mining, infrastructure and telecommunications and by enterprises managed by provincial governments engaged in service contracts, water system construction contracts and natural resources (Sun 2014). Some are employees of these enterprises temporarily assigned to an overseas project in Africa; others are recruited as overseas contract workers by government-licensed private employment agencies that operate out of Beijing, Shandong, Jiangsu, Liaoning and Shanghai (Politzer 2008). These municipalities and provinces are among the top-ten sending areas of contract workers to Chinese overseas development investment (ODI) projects (CASS 2012:109).

Little is known about the characteristics and numbers of Chinese contract workers attached to these projects in Africa, whether Chinese companies are sending managers and highly skilled workers on temporary work contracts – much like Honda has sent Japanese workers to the U.S., or Ford has sent U.S. workers to Mexico – or if they are also shipping production workers. Popular perceptions that Chinese companies provide the entire workforce for an overseas project in lieu of hiring locally (e.g. Gong 2007; French 2014) remain unverified. In fact, the ratio of Chinese to local workers is likely to vary by type of project, labor requirements and costs, local policies regulating foreign investments and the position of the Chinese firm employing workers vis-à-vis the local government. The China-Africa Business Council, using data from a 2012 survey of 198 member companies (80 percent private and 20 percent public) with established presences in 32 African countries, estimated that these companies employ an average ratio of 5 African workers to 1 Chinese worker (Zhong 2013).

The second major group of Chinese in Africa is composed of “independent” migrants who are entrepreneurs and merchants who come to Africa on their own accord to undertake various types of business ventures: e.g. as owners and employees of small businesses, often recruited through business owners’ family, friendship or acquaintance networks (Lin 2014; Haugen and Carling 2005; Mohan et al. 2014), or as employees of small and medium size enterprises with no state backing but with investment interests in Africa including mining, manufacture, construction and agriculture (Pieke and Speelman 2013). Chinese traders sell retail or wholesale Chinese manufactured goods across African urban markets and remote rural outposts (Esoh 2005; Bredeloup and Bertoncetto 2006; Ma Mung 2008; Haugen and Carling 2005; Lin 2014; Mohan et al. 2014; French 2014). Owners of small Chinese businesses service both the growing Chinese population in Africa as well as the local communities. These include professionals running traditional Chinese medicine clinics (Hsu

2007, 2008). Reports abound of independent migrants moving to Africa pushed by a desire to venture out of China, start a business there and become their own boss (Haugen and Carling 2005) away from the monitoring and mediation of the Chinese state, empowered by the relatively higher status that comes with moving to a place where they can be a “big fish in a small pond” (Lin 2014). For some Chinese women in Africa, this move also provides a space to be free from many of the constraints and social pressures imposed by traditional gender roles in China (Mohan et al. 2014). These migrants forego harder-to-reach destinations in Europe, the U.S., Australia and Canada in favor of African countries where entry and sojourning rules are less strict (Ma Mung 2008). Many are from Guangdong, Fujian and Zhejiang, three Southern coastal provinces which have historically dominated world-wide Chinese emigration (Liang and Morooka 2004; Ma Mung 2008; Lin 2014), yet others come from central China (Henan, Hunan, Hubei, Sichuan) or the industrial Northeastern China (Liaoning, Jilin and Heilongjiang). After suffering from the dismantling of China’s work unit system and the associated massive numbers of layoffs, China’s Northeastern provinces are emerging as important sending areas of international emigrants (Paul 2002). There are also reports of people transitioning from state-led to independent migration, formerly state-led migrants with expired labor contracts staying behind or return to Africa as traders (Park 2009; French 2014), young university graduates who take jobs with large Chinese companies operating in Africa and who aspire to set up their own businesses (Mohan et al. 2014), and friends and relatives of members of Chinese medical teams to Africa, or former members of these teams, who set up small traditional Chinese medicine clinics (Hsu 2007), though it is not clear how common this transition is.

The relative size of the two main groups of Chinese in Africa is not known. One can only speculate that, in countries where Chinese SOEs export highly skilled workers and managers but no production workers, the size of the state-led group will be smaller than the size of the group of independent migrants who populate the markets and Chinatowns of African cities but are also found managing shops in remote rural outposts.

While detailed descriptions of Chinese migrants in Africa are limited, the macro forces underlying this migration are easier to enumerate. On the China side, there is a strong national interest in gaining access to Africa’s natural resources, especially since China has transitioned from an oil exporter to an oil importer in 1993. Since then, China has sought to become Africa’s major development partner with resource-backed development loans to countries that are resource-rich and cash-poor (Brautigam 2009; 2010). Second, the industrial restructuring of the 1990s has encouraged Chinese small- and medium-size SOEs to search for commercial opportunities overseas (Brautigam 2009: Chapter 3). At the same time, the Chinese government has launched concessional aid loans to finance joint venture investments in manufacturing and agriculture in Africa and to create demand for Chinese machinery and labor services. Since 2006, the Chinese government has furthered investment in Africa by lowering the cost of loans to SOEs through its “policy banks” (e.g. China Eximbank and China Development Bank; Brautigam 2009), encouraging SOEs to go global under the auspices of the Going Out Policy (*zou chu qu*) (Biggeri and Sanfilippo 2009; Brautigam 2009; Brautigam and Tang 2011; Pairault 2013). This approach has turned many SOEs into global multinational corporations in search of overseas investments and lower-cost overseas production sites, especially for manufacturing. It has provided support for

China-Africa cooperation zones where Chinese companies export production to many parts of the African continent (Brautigam and Tang 2011; Brautigam 2009; Brautigam 2010). Indeed, Africa offers new grounds to expand business for Chinese firms that face increasing competition and declining revenues at home (Sun 2014). The sustained growth experienced by many African economies over the past decade has implied a potential for growing demand and markets for Chinese goods. African markets' allure to China has grown in the wake of the global financial crisis and high income countries' declining demand for Chinese exports (Battat 2006; Zweig 2009; Biggeri and Sanfilippo 2009).<sup>2</sup>

On the Africa side, the governments of African countries have seen an advantage in promoting cooperation with China. While most Western competitors are reluctant to invest in Africa, China is regarded as a key participant in the African commodity trade and a keen investor.

These factors have led to an upsurge in market relations between China and Africa. Over the last decade, the net trade volume between China and Africa increased from less than US \$100 million in 2000 to US\$198.49 billion in 2012 (State Council 2013). The bulk of these flows consists of the extraction of natural resources to China and of machinery, textiles, plastic products, transport equipment, chemicals and electronic goods to Africa (Ross 2013:135; Global Trade Atlas 2013). Following the growth in trade relations, China's investment stock in the continent has increased from US\$ 1.6 billion in 2005 to US\$21.23 billion by the end of 2012. As of March 2013, 2,372 Chinese investment projects in Africa have gone through the required approvals (MOFCOM statistics cited in Brautigam 2014).<sup>3</sup>

Although the phenomenon of Chinese migration to Africa pre-dates in some instances the recent inter-governmental economic cooperation between China and Africa, Chinese migration flows to Africa — and, while not the aim of this study, from Africa to China (Brown et al. 2010) — have grown since 2000, a time which coincides with the Chinese government's lifting of international travel restrictions in preparation for China's entry into the World Trade Organization (Xiang 2003).

Tanzania is one of many Sub-Saharan African countries that has recently experienced a surge of Chinese migration. A widely circulating though unverifiable estimate puts the total number of Chinese migrants in Dar es Salaam at 20,000 and the total number in Tanzania at 30,000. Even though this stock comprises mostly recent migrants, Chinese migration to Tanzania dates back to the 1960s–1970s, when Mao's China, for ideological and economic reasons, engaged in economic cooperation and development assistance to multiple African countries. Tanzania, under the presidency of Julius Nyerere, was a major recipient of this

<sup>2</sup>Surveys of Chinese firms have shown that dominant reasons for going overseas are the search for markets, new sales and distribution networks, and the search for manufacturing opportunities in countries where there is growing demand for their products in order to avoid import duties (Zweig 2009; Battat 2006).

<sup>3</sup>The bulk of the stock of Chinese ODI investment in Africa (excluding the financial sector) in 2009 was in mining (38.8%), manufacturing (24.9%) and construction (16.2%). SOEs under the direct supervision of the central government or provincial governments and their subsidiaries are believed to be responsible for about 80% of China's non-financial overseas direct investment (ODI) worldwide (Pairault 2013: Table 3). In Africa, SOEs under the direct supervision of the central government are managing mining and construction while manufacturing investment projects are left to the management of SOEs under the supervision of provinces, a distinction that originates from China's Soviet-style system of industrial organization during the Mao era (Pairault 2013:276–78).



development assistance. Since 1968, China has dispatched a total of 1,000 medical workers from China's Shandong province to Tanzania as part of the early medical assistance program which twinned Chinese provinces with African nations, while medical teams from Jiangsu province were sent to Zanzibar, a large semi-autonomous island archipelago about 40 miles off the coast from Dar es Salaam (Hsu 2008). Of the agricultural and industrial development projects in Africa financed by the Chinese government during that period, two Chinese projects in Tanzania stand out: the *Urafiki* (Friendship) Textile Mill (Monson 2009; Lee 2009) which started operation in 1968 in Dar es Salaam and the construction of the Tazara railway linking landlocked Zambia to the coast through Tanzania. This massive Chinese project was built in 1970–1975 and brought to Tanzania between 30,000 and 50,000 temporary Chinese workers who were repatriated at project completion (Monson 2009). More recently, Tanzania has become a hotspot of natural resource extractions with finds of oil, gas, iron ore and coal with new Chinese investments in infrastructure and resource extraction (Ross 2013) located in geostrategic proximity to the Indian Ocean coast. Tanzania is also one of Africa's emerging economies (Radelet 2010) characterized by robust economic growth, poverty reduction and political accountability. These factors have been shown to significantly influence Chinese foreign direct investment (FDI) and trade towards Africa (Biggeri and Sanfilippo 2009).

### 3. DATA AND METHODS

#### 3.1. Network Sampling with Memory

Recent innovations in link-tracing designs (i.e., designs where referrals from a set of respondents are used to recruit other respondents) seek to provide a probability-based inferential structure for the representation of populations that are difficult to sample because they are rare or hidden, in contrast to approaches like snowball sampling that do not offer probability-based inference (Volz and Heckathorn 2008). To recruit a sample of Chinese migrants for CAHS we used a variant of one such approach, Network Sampling with Memory (NSM, described in detail in Mouw and Verdery 2012). NSM capitalizes on the network structure of the target population to identify and interview respondents who are successively referred into the sampling frame by prior respondents over “waves” of data collection and are targeted by researchers who use an algorithm to identify which individuals to sample next (discussed below). NSM sample recruitment starts with the identification of a limited number of initial respondents (“seeds”) who are asked to provide minimally identifying information (e.g. last name and last four digit of cell number) on a given number of friends and acquaintances who are members of their social network, called network alters. Typically respondents are given space to name their alters with the following question wording: “a person whose name you know and he or she know yours and you might stop and talk at least for a moment with if you ran into them on the street”. This type of question is referred to in the literature as a “name generator”, which is a common and well-studied method of eliciting socially relevant peers with reasonably high levels of validity despite some biases toward nominating peers who respondents interact with more frequently (Campbell and Lee 1991; Marsden 1993; Straights 2000; Marin 2004). Minimally identifying information collected in this roster allows one to combine all nominated individuals in a single network and to identify people who were nominated by more than one

respondent. As the NSM sample progresses, this reconstructed network increasingly resembles the true social network linking members of the target population. To accelerate this process, NSM uses the currently revealed network at each step and a sampling algorithm to “direct” the recruitment process to spread rapidly through the underlying population’s social network by placing higher sampling probabilities on people who have been nominated less frequently by previously sampled respondents, increasing the chance of discovering unknown sections of the network. By testing the sampling process using simulated sampling on actual school and university networks, Mouw and Verdery (2012) show that NSM dramatically lowers the design effects (a measure of sampling efficiency calculated as the ratio of sampling variance to the sampling variance of simple random sampling) compared to more popular network sampling alternatives such as Respondent Driven Sampling (RDS, Heckathorn 1997, 2002) which suffers from very large design effects (Goel and Salganik 2010) and biased variance estimators (Verdery et al. 2015). Smaller design effects imply that researchers can achieve the same or better statistical precision (and, consequently, narrower confidence intervals) using NSM on the basis of smaller samples than would be required for RDS, i.e., with NSM they need only collect samples in the hundreds rather than in the thousands to obtain precise and accurate estimates (Mouw and Verdery 2012).

In the CAHS survey, we employed a variant of NSM, forward network sampling with memory (fNSM; Mouw et al. 2014), that improves the feasibility of implementing it in the field. In the original version of NSM, respondents might be contacted multiple times after the interview to provide full contact information on their nominated network alters who were randomly selected to be interviewed, increasing the burden on respondents in the field. With fNSM, respondents are asked during the interview to provide full contact information on only up to three randomly selected peers among the list of nominated peers for whom contact information has not previously been obtained. These individuals then form a queue from which researchers draw potential respondents. Sampling proceeds in this fashion from an ever expanding sampling queue of potential respondents with full contact information until the target sample size is reached. Mouw et al. (2014) showed with simulations that this variant only slightly reduces sampling efficiency with an average design effect in samples of size 500 of 2.2 compared to design effects of 1.16 in conventional NSM samples of similar size.

Eligibility for participation in CAHS included residence in Tanzania for one month or more and being at least 18 years old. Data collection started in early June 2013 and was preceded by two field visits of several months each to Dar es Salaam by the first and fourth author in late spring 2012 and winter-spring 2013 to conduct formative research in preparation for the survey. This included familiarizing ourselves with the Chinese community in Dar, assessing whether Chinese in Dar had extensive social connections to each other, identifying suitable seed respondents and gaining the trust of the Chinese community. Study approval was obtained from the ethical review board of Duke University and the National Institute of Medical Research (NIMRI) in Tanzania.



### 3.2. Questionnaires

The CAHS survey began with the selection of seven seeds known to the researchers, stratified by migrant group according to the ownership sector of their employment and by province of origin. Each seed was invited to the CAHS interview office located in the same compound where the largest Chinese grocery store in Dar es Salaam was located. Using the fNSM approach, a total of 147 respondents were ultimately selected for participation and administered two questionnaires: the main questionnaire and a network roster questionnaire. The individual questionnaire permits the characterization of differences between migrant groups, their migration experience, and economic activities. Respondents' migration history was capped at four migration spells. Respondents were asked about the start date and duration of their current migration spell in Tanzania, of their first ever migration to Tanzania if different from the current one, of the most recent migration to a country other than Tanzania, along with retrospective information about the employment activities associated with each of these migration spells. Respondents were also asked about their residence and employment activities in China before emigrating overseas and the duration in that place of residence. To identify internal migration in China, information on their most recent place of residence and the place of their household registration (*hukou*) in China was also asked.

The network roster questionnaire was administered to all respondents prior to the main interview. Respondents were asked to nominate up to 10 alters defined as other Chinese migrants they knew and who resided in Dar es Salaam using the name generator discussed in the prior section. After nominations were elicited, respondents were asked to provide the last name and last four digits of the cell phone number as well as basic demographic information (gender, age and province of usual residence in mainland China prior to migration to Tanzania) of each of their nominated alters. As explained in section 3.1, this minimally identifying information on respondents' nominated network alters allowed us to combine all nominated individuals into a single network and to identify individuals nominated by more than one respondent.

### 3.3. Creating the network

To identify unique individuals in the network (that is individuals nominated by at least one respondent), we wrote a matching program in Stata to match individuals on their gender, last name, and last four digits of their cell phone numbers. The combined network rosters resulted in a total of 1,000 uniquely identified individuals who had been nominated a total of 1,282 times, the difference between 1,000 and 1,282 are individuals nominated by multiple respondents. During each interview, after asking questions about the respondent's network roster (which provides information on up to 10 network alters), the fNSM sampling algorithm randomly selects up to three nominated alters for whom full contact information is asked. These randomly selected alters exclude any network member for whom contact information has already been collected on previous interviews. The goal of this step is to speed up the process of eliciting referrals by reducing the burden on respondents while preventing the collection of redundant contact information on previously nominated contacts. According to this protocol, the CAHS study team asked for full contact information for 296 randomly selected members of respondents' networks, and obtained a first-stage response rate of 87.5% on these requests for contact information. Among the 259

nominations with full contact information in the queue to be contacted, 12 remained in queue and were not contacted before the end of the study, and 247 were invited for an interview. Of these, 147 accepted the invitation and became CAHS respondents, 6 accepted but did not show up for the interview, and 94 rejected the invitation resulting in a second-stage response rate of  $147/247 = 59.5$  percent<sup>4</sup>. Because basic demographic characteristics were collected from respondents about their nominated alters, we were able to compare the demographic characteristics of nominated alters who agreed to become respondents and non-respondents. This comparison revealed no significant difference between the two groups.

A visual inspection of Figure 1 illustrates the key features of NSM in sampling from a hidden population. The figure shows the final accumulated roster of 1,000 uniquely nominated individuals of the target population sampled by CAHS. These individuals are represented as nodes of the social network (black, dark gray and light gray points on the graph). Black and dark gray nodes include seeds and sampled and interviewed nodes from among the nominations, while light gray nodes were nominated by respondents but not selected for an interview by the sampling algorithm. Interviewed nodes are color coded by ownership sector. State-led migrants (dark gray nodes) are respondents who reported employment with a state-owned enterprise (including those under central or provincial control and their subsidiaries), while independent migrants (black nodes) are respondents who reported employment with a private enterprise, the self-employed or those who worked for a business owned by a family. The seven original seeds – two are state-led and five are independent migrants, are denoted by a square symbol. The seed nodes are dispersed throughout the network, with some being peripherally located (far from the densest points of interconnection in the center of the figure) and others central. The figure shows that some state-led migrants nominated independent migrants and vice-versa (i.e., cross-recruitment nominations are not uncommon). Tracing through the connections generated by these nomination patterns, we can see links between seeds who are member of one group leading to sampled nodes of the other group. The size of each nominated but unsampled (light gray) node in the graph is proportional to the probability of that node's being sampled had the survey continued, while sampled nodes are portrayed with a fixed size. Because the NSM sample recruitment process places higher sampling probability on the nodes likely to lead to undiscovered portions of the networks, un-sampled nodes which are closer to the center of the graph have smaller sampling probabilities than the nodes in the peripheral areas of the network.

### 3.4. Nomination patterns

An important finding regarding the organization of migrant communities in places of destination is that migrants from the same place of origin have a tendency to form community (e.g. Jampaklay, Korinek and Entwisle 2007; Entwisle et al. 2007). In the Chinese migration context in Tanzania, the compartmentalization of economic and social activities of different groups of migrants may impede the formation of social ties across

---

<sup>4</sup>The overall response rate is the combination of the first- and second-stage response rates, excluding the 12 referrals that remained in the queue, and is calculated as  $0.875 \times 0.595 = 0.520$ , or 52%.

ownership sectors. To examine how social ties of Chinese migrants in Tanzania are patterned by ownership sector of employment and province of origin, we focus on nomination patterns of respondents by each of these two characteristics and use two methods to characterize the patterning of these ties: mixing matrices and homophily models for dyadic data as described below.

**3.4.1. Mixing matrices**—Using the example of provinces of origin, we are interested in whether respondents tend to nominate alters from their own province of origin (i.e., intra-province nomination) and, to the extent there is cross-nomination by province, which provinces are most prominently featured in these patterns (inter-province nominations). Reports by respondents about their alters allow us to directly measure nomination patterns within and across attributes by forming a “mixing matrix” (Morris et al. 2009). A mixing matrix is defined by cross-classifying all nominations according to the attribute of the respondent and the reported attribute of his nominated alters. Thus, for instance, with two categories of respondents (e.g., migrants from Shandong and migrants from all other provinces) a naïve version of the mixing matrix would be formed using a 2 by 2 table with each cell entry recording how many nominations went from, e.g., Shandong migrants to Shandong migrants, Shandong to other, other to Shandong, and other to other. However, when constructing attribute mixing matrices based on naïve frequencies or percentages, there are two potential threats to validity. The first is that the amount of observed inter-provincial nominations is conditioned by the relative numbers of people in each province. To cope with this issue, we use the logic behind the familiar  $\chi^2$  test, except that we compute the ratio of the observed-to-expected values in each cell. In an observed/expected ratio, we draw on the naïve frequency of nominations within each cell of the mixing matrix (observed nominations; i.e., how many nominations from migrants of one type to migrants of another type), and the expected number of nominations under conditions of statistical independence (defined mathematically below). We construct the observed/expected ratio (O/E Ratio) for the pair of province  $i$  and  $j$  as follows:

$$O/E \text{ Ratio}(i, j) = \frac{k_{ij}}{\left( \sum_i k_i \sum_j k_j \right) / \sum_i \sum_j k_{ij}}$$

where  $k_{ij}$  represents the number of nominations observed between people from province  $i$  and people from province  $j$ ,  $\sum_i k_i$  is the number of nominations sent by people from province  $i$ ,  $\sum_j k_j$  is the number of nominations received by people from province  $j$ , and  $\sum_i \sum_j k_{ij}$  is the total number of nominations in the table. We interpret the O/E Ratio as a selection coefficient which measures proportionate deviation from what would be expected given population composition and nomination rates by province. This is not an uncommon practice in the networks literature (cf. Morris et al. 2009).

The second potential threat to validity in naïve mixing tables is represented by differential numbers of nominations. Because respondents may nominate different numbers of alters, if those who nominate more alters are differentially likely to nominate intra-province ties, this can bias estimates of the mixing across provinces (cf. Young et al. 2013). To cope with this problem, we apply a solution similar to that described by Young et al.: 1) randomly sample one alter per respondent and calculate the observed/expected ratio across this sample of respondents; 2) repeat step 1 many times, 200 in our case; 3) average the observed/expected ratios across the 200 samples. This bootstrapping approach, named within-cluster resampling approach (WCR) by its proponents, has the advantage of including only one observation per respondent, thus reducing bias owing to conditional dependencies between cases.

**3.4.2. Homophily models**—To examine the relational structure of Chinese in Dar es Salaam, we examine patterns of homophily, which is defined as the tendency of respondents to nominate alters with attributes similar to themselves (e.g. McPherson et al. 2001). We look specifically at homophily over nominations by ownership sector and province of origin. We use recently developed models that offer a parsimonious framework in which to examine homophily in networks, parsing out the extent to which respondents nominate within vs. between ownership sectors or provinces owing to the demography of group sizes or nomination rates (Daw, Margolis, Verdery 2015; Cameron et al. 2011; DeFries-Fulker 1985). We first construct a data set comprising all pairs of individuals in the network, whether or not they share a tie and estimate the following equation (Model 1):

$$\ln(\text{pr}(Y_i=1)/\text{pr}(Y_i=0))=\alpha+\beta_1Y_j+\beta_2T_{ij}+\beta_3Y_jT_{ij}+u_i+u_j+e_{ij},$$

where  $Y_i$  indicates the ownership sector (SOE=1, independent migrant=0) of respondent  $i$ ,  $Y_j$  indicates the ownership sector of respondent  $j$ ,  $T_{ij}$  indicates whether or not respondent  $j$  nominated respondent  $i$  (or vice versa), and  $u_i$ ,  $u_j$ , and  $e_{ij}$  are individual  $i$ , individual  $j$ , and dyad specific error terms. The coefficient  $\beta_1$  measures what is called “baseline homophily” (McPherson et al. 2001), i.e., “homophily created by the demography of the potential tie pool”.  $\beta_1$  can be understood as a measure of the extent to which homophily in the network owes to group sizes (an issue discussed above with respect to mixing matrices). In this model,  $\beta_2$  measures the homophily created by nomination rates (also discussed above), while the interaction  $\beta_3$  measures inbreeding homophily, that is homophily which is “explicitly over and above the opportunity set” (McPherson et al. 2001:419). Inbreeding homophily, in other words, measures how unexpected the levels of within and cross group nominations are, after controlling for group sizes and nomination rates. Inbreeding homophily is equivalent to previously described observed-expected ratios, but formalizing it in a regression model allows us to conduct statistical hypothesis testing and compare its performance across models. So the next model (Model 2) explores deviations from inbreeding homophily in cases where both individuals are from the same province. This is represented by the  $\beta_7$  coefficient in the following equation:

$$Y_i=\alpha+\beta_1Y_j+\beta_2T_{ij}+\beta_3Y_jT_{ij}+\beta_4P_{ij}+\beta_5Y_jP_{ij}+\beta_6T_{ij}P_{ij}+\beta_7Y_jT_{ij}P_{ij}+u_i+u_j+e_{ij},$$

where  $P_{ij}$  measures whether  $i$  and  $j$  are from the same province of origin in China. A positive coefficient would indicate that the tendency to nominate within ownership sector for dyads from the same province is greater than would be expected on the basis of group sizes and nomination rates.

### 3.5. Bootstrap sampling weights

Mouw and Verdery (2012) showed that the average of the mean estimates from NSM samples across repeated simulated samples converge to the population mean (i.e., it is unbiased) when the revealed network size converges with the size of the population social network estimated with a capture-recapture approach, a method for estimating population size based on network data (e.g. Dombrowski et al. 2012). However, there is no guarantee that these results hold before this stopping point is reached. The CAHS survey was terminated after recruitment of a sample of 147 respondents and before the revealed and estimated network sizes converged. When the total sample size is small (i.e., when few waves have been collected), network sampling methods, like NSM, will be biased if the seed respondents are not randomly selected, though NSM is less susceptible to this tendency than other network sampling approaches (e.g., RDS) because of its targeted algorithmic searching of the network. Nonetheless, NSM samples, particularly small ones, may still have biases, but these can be mitigated using advanced methodologies. For example, to reduce potential biases due to seed dependence and other biases arising from an incompletely revealed network and to correct biases towards sampling popular individuals, we leverage the unique data collected in NSM to construct bootstrap sampling weights that we use to reweight cases. Weights for each person are constructed based on the number of times he or she was sampled in simulated NSM sampling chains running over synthetic population social networks using a procedure similar to Gile and Handcock (2011). The full procedures used to generate sampling weights and to adjust means and proportions are described in the Appendix. These sampling weights have been shown to reduce potential biases due to small sample sizes (Mouw et al. 2014) but to have little impact on the design effect, which in samples of the size of the CAHS sample tend to range between 1.1 and 1.8 (Mouw and Verdery 2012), meaning that an NSM sample of 200 cases would have the same statistical precision of simple random samples between 111 to 182.

## 4. RESULTS

### 4.1. Means and proportions

Table 1 reports adjusted and unadjusted percentage distributions and means for the main socio-demographic and migration characteristics of CAHS respondents. Overall, Chinese in Dar es Salaam are overwhelmingly male, with a mean age of 33 years. More than half of the sample is married and almost half reports having attained a college degree. Less than half have experienced an internal migration spell in China (i.e. the place of their most recent residence in China was not the place of their household registration, or *hukou*) and for 41 percent their place of most recent residence in China is different from the place where they grew up (i.e. the place they used to live when they were 14 years old). The mean duration of their current residence in Tanzania is 3 years. Chinese migrants in Dar es Salaam typically live with colleagues in collective housing provided by their employers on construction sites

and factory dorms or in private houses. The average respondent lives with other Chinese, that is Chinese represent 94.6 percent of the cohabiters of the average CAHS respondent. The table also describes the characteristics of the sample by whether a respondent was a state-led or an independent migrant.

State-led migrants account for 27 percent of the sample. There are marked socio-demographic differences between state-led and independent migrants. Compared to independent migrants, state-led migrants are younger, less likely to be married, with more years of schooling, more likely to have a college degree and more urbanized, with 86 percent having an urban household registration in China compared with 45 percent of independent migrants. The majority of state led migrants are from Beijing, Shanghai and Tianjin where the headquarters of large SOEs are located or from Shandong, Jiangsu, Hunan, Henan and Sichuan where state-sponsored migration agencies recruit contract laborers for job assignments overseas. About half of independent migrants originate from Zhejiang, Fujian and Guangdong, the three historical sending provinces of emigration to Europe, the U.S. and other parts of China or from China's northeastern provinces, an area which has only recently gained its place on the emigration map. Half of independent migrants, many of whom are of rural origin, have had a previous experience of internal migration in China, while this fraction is much smaller among state-led migrants. The mean duration in Tanzania of independent migrants is one year longer than that of state-led migrants. Figure 2 shows the frequencies of CAHS respondents in each ownership sector by year they moved to Tanzania. Two factors are likely to explain the shorter and more recent residence spells of state-led migrants. First, state led migration is a phenomenon which took off especially after 2006 upon the Chinese government encouraging SOEs to "go out". Second, labor contracts typically expire after three years, although renewal of contract is not uncommon (personal communication during formative research; Park 2009). Regardless, only a few members of the sample, all of them independent migrants, had arrived to Tanzania before 2005 and very few before 2000, the year that marks the onset of Chinese migration flows to Africa.

Figure 3 characterizes the economic activities of the sample of independent and state-led migrants. Most SOE respondents are employed in construction and mining, while retail, construction, and manufacturing, stand out among the main economic sectors of independent migrants. This is consistent with the economic specialization of state and privately owned enterprises in Africa. In contrast to SOEs which operate in key resource sectors, private Chinese companies are more sophisticated in processing local products like cotton and leather into manufactured goods such as garments and shoes, while retail represents the main economic activities of self-employed traders and small-time entrepreneurs who populate African cities' markets and Chinatowns and who rely upon a supply chain of goods, family members and social contacts stretching back to the mainland.

The CAHS survey identified 36 owners of small to medium private and family-run businesses who were asked the number of Chinese and of Tanzanian workers they employed. Based on the reported counts of their employees by ethnicity, we computed the ratio of Chinese to Tanzanian workers. Figure 4 shows that, with the exception of mining, this ratio is close to 4:1, or 80 percent of all workers are hired locally, regardless of the economic sector of the business. This is highly consistent with Tanzania's labor policies which



mandate foreign companies to reserve 80 percent of their job openings for local employees. The 4:1 ratio also compares well with the ratio of 5:1 drawn from the 2012 China-Africa Business Council survey of 198 member private and public companies with established presence on the African continent (Zhong 2013) and is consistent with the observation by Mohan et al. (2014) that, across the 85 Chinese enterprises in Nigeria and Ghana they studied, the ratio of African to Chinese workers was highly skewed towards the former. It is, however, remarkably inconsistent with the widespread perception that China is filling its projects with its own workers. Although it is hard to draw conclusive statements from very few observations, mining represents an exception both due to the higher skills required and/or the fact that respondents interviewed in Dar working in mining were managers who spent part of the year in the Dar es Salaam office, not workers in the mining sites outside of Dar es Salaam. This evidence, corroborated by our own observations of Chinese family-run businesses in Dar-es-Salaam, cautiously suggests that at least independently-owned Chinese businesses in Tanzania appear to comply with local employment policies.

#### 4.2. Migration histories

An important question surrounding Chinese migration is whether, following China's lifting of emigration restrictions, the restructuring of the state sector in urban China and the proactive role taken by some local governments in encouraging emigration (Lu et al. 2013; Pieke and Speelman 2013), the profile of Chinese international migrants is changing and whether the lure of and opportunities for international migration are expanding beyond geographic areas with well-defined streams between origin (e.g. Fujian, Zhejiang, Guangdong) and destination (e.g. Europe and the U.S.). A description of the migration histories of CAHS respondents represented in Figure 5 sheds light on the experience of Chinese emigrants to new frontiers. This figure tracks the duration and type of migration spells of CAHS respondents by Chinese provinces of origin. Provinces are grouped according to whether they share similar histories and emigration environments and respondents are assigned to province based on their place of permanent household registration (*hukou*). In the CAHS study, the migration experience of respondents was truncated, by design, to one event per migration type: current residence in Tanzania, most recent internal migration spell in China, first ever residence in Tanzania and the most recent migration spell in another foreign country. Date of birth is included to highlight the relationship between the number and length of migration events with age.

Respondents with household registration in provinces that historically send emigrants to Europe, the U.S. and other parts of China (Zhejiang, Fujian, Guangdong) report multiple types of migration experiences including internal migration and migration to another foreign country prior to their current residence in Tanzania. A similar experience of migration to multiple destinations pertains to respondents from central Chinese provinces (Hunan, Hubei, Henan, and Sichuan). These are areas which have long specialized in sending unskilled construction workers to Chinese cities. On the other hand, the experiences of respondents from large urban centers, the Northeastern provinces (Jilin, Liaoning, Heilongjiang), Shandong and Jiangsu are, for the most part, limited to current or previous migration spells to Tanzania. This is because these provinces have only recently gained a place in the Chinese emigration map. Beijing, Tianjin and Shanghai are home to the largest Chinese SOEs, while

the residents of China's Northeastern provinces, who have borne the brunt of massive layoffs following the restructuring of the state sector, are now looking at Africa as one place of opportunity. In Shandong, state-sponsored migration agencies recruit contract workers for overseas projects. Shandong is also a province where the commercialization of emigration has become a phenomenon involving independent agencies and local governments (Paul 2002; Xiang 2005). The province also boasts historical ties with Tanzania dating back to an earlier period of Chinese health development assistance to Africa. Indeed, respondents from Shandong report the most previous trips to Tanzania. Finally, the number and total duration of the migration experience based on the four most recent migration events seem to be correlated with age with respondents born in earlier years reporting in general a longer cumulative experience of migration. However, because the length of the migration experience of migrants with more than one migration spell of each type is truncated, this evidence is not conclusive.

Figure 6 highlights the different migration experiences of independent (left panel) and state-led migrants (right panel). Compared with state-led migrants, independent migrants report longer and multiple, international as well as internal migration events. For independent migrants, opportunities in Africa have replaced internal migration opportunities, especially among older respondents with a longer exposure to migration experiences. For most state-led migrants, on the other hand, assignments to Tanzania or another foreign country represent the first instance of migration away from their usual place of residence.

### 4.3. Nominations patterns

**4.3.1. Mixing matrices**—To explore whether it is province of origin or ownership sector of employment to pattern social ties among Chinese migrants in Tanzania, Table 2 and Table 3 report O/E ratios among nominations dyads of sampled respondents by ownership sector<sup>5</sup> and by province respectively. Nomination patterns of SOE employees are characterized by strong homophily, with a tendency to nominate and refer each other at a rate that is 2.3 times higher than expected given population composition and nomination rates, compared with a rate only 12 percent higher than expected among independent migrants. The question is whether being from the same province of origin increases the propensity to nominate across sector given the observed tendency of migrants from the same origin area in other settings to form social ties. To address this question, we first use information on province of origin of respondents and their nominated alters to investigate with mixing matrices nomination patterns by migrants' province of origin. The network roster questionnaire asked respondents about province of residence of their alters, a question which participants found easy to answer. Of the 1,282 nominations, only 113 had missing information on province.

A feature of Figure 1 shown above is that while nodes representing independent and state-led migrants form single-colored clusters in certain areas of the network and especially in the periphery, there is notable overlap of black and dark gray nodes in the central area of the network. Figure 7 has the same configuration as Figure 1 but, instead of the ownership

<sup>5</sup>The network roster questionnaire did not collect information on the ownership sector of CAHS respondents' alters because, during the formative stage of the study, we found that informants were more likely to know the economic sector than the ownership sector of their alters. For this reason, this table is generated from a count of nomination dyads of sampled respondents only.

sector of 147 respondents, it shows province of origin of the 1,000 uniquely identified nodes in the network. In general, nodes cluster by province, with the clustering being most evident for migrants from historical sending areas of emigration, from provinces new to emigration and from Jiangsu and Henan province. The center of the graph is the place where respondents from different ownership sectors overlap. This place is populated by nodes from multiple provinces, with a small cluster of nodes from Shandong, suggesting the possibility of nominations across province and ownership sector.

Table 3 presents adjusted O/E ratios computed from nomination patterns by provinces. The largest coefficients are along the diagonal, suggesting a general tendency among Chinese migrants in Dar es Salaam to nominate alters from the same province of origin. The coefficients are all greater than 3 times what is expected for migrants from historical emigration provinces (Zhejiang, Fujian, Guangdong), from provinces new to emigration (Northeast provinces), and from Shandong, Jiangsu and Henan provinces. Migrants from three historical emigration provinces and from the three Northeastern provinces new to emigration make up about half of the independent migrants group. Migrants from these groups of provinces are the most isolated in the network. They nominate alters from other provinces at well below the expected rate, and they are always nominated at half the expected rate by migrants from other provinces. These nominations patterns are consistent with the provincial clusters of nodes observed in Figure 7. Nodes representing migrants from provinces new to emigration are clustered in a peripheral part of the network, while the position of nodes denoting historical areas of emigration is mostly peripheral to the network except for a few nodes interspersed throughout the network. Of particular note is the social distance between these groups of independent migrants and migrants from the big Chinese cities, some of the main places sending state-led migrants to Africa. Migrants from the Northeast and those from Zhejiang, Fujian and Guangdong nominate and are nominated at less than half the expected rate by migrants in Beijing, Tianjin and Shanghai.

Nomination patterns between provinces which are sending state-led migrants to Tanzania are quite different. Migrants from the big cities nominate others from their same province group and from residual provinces at a rate higher than expected and are nominated by those from Shandong, Henan, Jiangsu and residual provinces at a higher than expected rate. Business ties and alliances among state-led migrants are coordinated by sector-specific business associations under the oversight of the Chinese embassy which coordinates the bidding process for local large-scale projects (e.g. construction projects), so ties among state-led migrants can cut across provinces. This is consistent with strong patterns of intra-sector nominations for state-led migrants but with some nominations across provinces of origin of SOE managers and contract workers.

Lastly, migrants from Shandong tend to nominate and be nominated across provinces at a higher rate. Coefficients reflecting people from Shandong nominating those in other provinces are rarely below half the expected rate and Shandong people are never nominated at less than half the expected rate by those from other provinces. This is consistent with Shandong nodes' central position in the network graph. We can only speculate that the multiple earlier trips to Tanzania reported by CAHS respondents from Shandong may

explain their central position in the network as their longer experience in Tanzania may have exposed them to a variety of other Chinese.

**4.3.2. Homophily Models**—We are interested in whether respondents from the same province of origin are more likely than would be expected based on group sizes and nomination rates (i.e., inbreeding homophily) to nominate within vs. between ownership sectors. To examine this, Table 4 shows the coefficients of the two homophily models. In both models, there is substantial evidence of inbreeding homophily on ownership sector, meaning that those from the same ownership sector are more likely to nominate each other than would be expected based on the size of each group and the average number of people they nominate. When province of origin is accounted for in Model 2, being from the same province substantially increases the chance of homophily on ownership sector. However, while sharing a province with an alter increases the probability of intra-sectoral ties, it does not increase the probability of ties between state-led and independent migrants. Table 5 shows predicted probabilities computed from Model 2. While people who are nominated by a state-led migrant from the same province have a 86.4 percent probability of being in the same ownership sector (93.2 percent probability of being in same sector for independent migrants from the same province), people who are nominated by a state-led migrant from the same province have only a 13.6 percent probability of being an independent migrant compared with a 56 percent probability of being an independent migrant for people who are nominated by a state led migrant from a different province. While same-province ties are less likely to cross ownership sector than cross-province ties, this difference is not statistically distinguishable. Sharing the same province of origin increases the probability of intra-sectoral but not inter-sectoral ties.

## 5. DISCUSSION AND CONCLUSIONS

To provide a novel description of the demography and social organization of a Chinese community in Dar es Salaam, Tanzania, we used a variant of Network Sampling with Memory. This approach has numerous advantages for recruiting useful analytical samples of hidden and rare populations such as Chinese migrants in Africa. It overcomes the challenges of sampling populations that lack a sampling frame and yields accurate and reasonably precise estimates of their characteristics with smaller samples than required by other popular network-based sampling approaches. It uses a probability based inferential framework, albeit one still under development, that facilitates group comparisons. It collects network data from sample respondents which enables the mapping of social ties among members of the hidden population.

The NSM sample of Chinese migrants we recruited consists of 147 Chinese migrants who reported on themselves and on 853 unique Chinese peers also living in Dar es Salaam. These data shed light on the contours of a heterogeneous Chinese community and on the structure of the social ties linking its members. This community is composed of two distinct categories of migrants: workers who are sent on temporary labor contracts to work on large scale projects run by China's SOEs and various types of employees of private medium and small size enterprises, small-time entrepreneurs and traders – and those working for them –

whose commercial activities are unfolding independently from the coordination of the Chinese state.

The data corroborate much of what is known from a growing body of, mostly qualitative, literature set across numerous African countries, such as Angola, Cape Verde, Ghana, Nigeria, South Africa and Zambia, on the individual-level characteristics of recent cohorts of migrants from mainland China (e.g. Haugen and Carling 2005; Lee 2009; Park 2009; Lin 2014; Mohan et al. 2014). Importantly, they add socio-demographic details and highlight meaningful differences between the two groups of migrants. Compared with the highly educated and urbanized state-led migrants, independent migrants are older, with fewer years of education, more likely to be married, to originate from rural areas in China, to have experienced at least one prior internal migration spell in China and a longer migration spell in Tanzania. The province of origin composition of Chinese migrants in Dar es Salaam also suggests that Chinese emigration is expanding beyond the traditional sending provinces such as those that have long sent Chinese emigrants to the U.S. and Europe and is diversifying in terms of destinations and array of opportunities sought in these destinations. These findings lend credence to the rise of new types of migratory flows emanating from China, which are shaping the new Chinese migration order (Pieke 2007) characterized by the globalization of emigration, a new geography of sending areas from China, and a diversification of socio-economic backgrounds and migration decisions of emigrants.

The social organization of the Chinese community in Dar es Salaam as gauged by nomination patterns in the CAHS survey is still mostly structured along ownership sector with significant social distance observed between the two groups of migrants. Province of origin is an attribute that defines the intra-sectoral but not the inter-sectoral ties of the Chinese community in Dar es Salaam. The finding of intra-sectoral ties characterized by a shared province of origin is not inconsistent with findings from studies which have shown that sharing family ties and same place of origin is a crucial feature of co-ethnic social organization of the Chinatown communities in the U.S. (Zhou 1992; Lin 1998) characterized by a significant overlap between economic activities and place of origin. Yet, the absence of significant inter-sectoral ties also suggests that, in this community of Chinese in Africa at least in the current, early phase of emigration from mainland China to Africa, social relations are more likely to be based on ties defined by a shared economic imperative than by a shared place of origin. The segregated lifestyles and long working hours of members of the two groups of Chinese migrants in Dar es Salaam limit opportunities for social contacts and nominations across ownership sectors. Employees SOEs in Dar es Salaam are no different from those working on state-led projects in countries such as Zambia or Angola, who reside in “enclaves” (Lee 2009; Mohan et al. 2014) and secluded residential quarters, gated compounds, separated from the local workforce, “complete with [their] own security guards, cooks, kitchen, satellite dishes, television and karaoke rooms, video and DVDs from China, ping-pong tables and basketball courts” (Lee 2009).

Similarly, the opportunities of independent migrants to form ties across ownership sector are rare. First, the organization of Chinese business communities in Africa where Chinese sector-specific business associations and chambers of commerce only represent the interest of well-established businesses or of SOEs (Mohan et al. 2014) prevent the formation of

business alliances between the two groups of migrants. Second, the geographic clustering of Chinese traders, retailers and wholesalers in the boisterous markets of Dar es Salaam limit the formation of ties to migrants from Chinese provinces specializing in sending small-time entrepreneurs overseas, and often only among those people working for non-competing businesses. These migrants work long hours, they typically live in quarters located in proximity to their shop and are embedded in their own communities of similar others with little free time and few, tenuous social bonds with other Chinese living and working outside of their neighborhoods. Low levels of community cohesion were also observed among Chinese communities in Ghana and Nigeria by Mohan et al. (2014) who attributed the absence of strong social bonds to high competition among Chinese migrants working in similar lines of retail and wholesale business and to the social isolation of those who traveled to Africa alone and found employment in small Chinese-run businesses.

Furthermore, the social barriers entailed by different sociodemographic positions of members of the two groups before leaving China may replicate the social distance between the two groups in countries of destination. If China's socially stratified rural-urban divide reinforced by the *hukou* system (Wu and Treiman 2004, 2007) is any telling, this divide may propagate to this Chinese community in Tanzania, where the two groups of migrants characterized by different socio-demographic profiles live segregated social lives and have no incentive to seek social contacts outside of the ownership sector of their current employment.

The differences between the two categories of Chinese migrants we have identified have numerous implications for future research on Chinese migration to Africa. First, although the recent rise of international migration from China, after a standstill during the first four decades of the People's Republic of China, has resulted in a growing number of empirical studies on Chinese international migration (e.g. Liang et al. 2008; Liang and Morooka 2004; Lu, Liang and Miao 2013), the primary emphasis in this literature is on emigration flows to developed countries from geographic areas in China with historically well-defined streams between origin and destination. Yet the diversity in provincial origin of Chinese in Dar es Salaam suggests that Chinese emigration is expanding beyond these areas. New analyses are needed to explore the changing profile and growing multiplicity and forms of China's present international migration flows to include those towards African and other less developed countries.

Second, a deeper understanding of the role of the Chinese state in shaping international migration and its implication for migration outcomes in the place of destination is needed. For example, there may be meaningful differences in the well-being of people who are sent by their employer to work on SOE overseas projects with fixed labor contracts, prior knowledge of expected earnings and residential arrangements at destination organized by their employers and migrants who move independently with less clear job prospects and without guaranteed earnings before migration. These features suggest different costs, barriers and benefits of migration faced by independent migrants compared with state-led migrants. Differences between the two groups may extend over and above the differences in human capital required by migrants' occupations and ownership sectors. Further, if the nature of migration selection on a set of characteristics, such as age, education and risk



preferences depends on the correlation between these characteristics and the costs and benefits of migration, we would expect the nature of the different types of migration to Africa to be associated with differences in migrants' individual outcomes, even if the selection mechanism operating within each group cannot be observed due to the absence, as in our study, of a comparison sample of non-migrants and of full information on status characteristics and risk preferences prior to migration. In addition to the information analyzed here, the CAHS survey also collected a risk preference module and numerous health behaviors modules which, albeit outside the scope of the current study, will allow one to address questions related to migrant health outcomes at destination.

Third, the different types of Chinese migrants in Africa may have different impacts on local African communities and may contribute different images of China in Africa. The high profile and the large size of projects which bring state-led migrants to Africa imply close coordination and monitoring of Chinese workers by their superiors representing the Chinese state and limited interactions with local communities beyond the work site. Such supervision hardly exists for smaller, private Chinese companies and small-time entrepreneurs, traders, retailers and wholesalers, who, because of the nature of their economic activities, have frequent interactions with locals, hence greater potential for being at the center of tensions with local communities (e.g. BBC 2011). Moreover, if Chinese migrants are working in Tanzanian-dominated employment settings, it will be crucial to examine patterns of ethnic segregation within workplaces and the nature of contacts with co-workers in these firms.

With the rapid rise of China's global engagement and the relaxation of emigration restrictions from China, any discussion surrounding Chinese migration to Africa must consider these different types of migrant flows, their differences in individual motivations for migration, economic activities, migration outcomes, social networks at destination, expectations about their future residence on the African continent and their relationships with and impact on local communities.

## Acknowledgments

The data collection and analyses described here were made possible by various grants to Giovanna Merli, including a pilot grant from the Duke Global Health Institute (DGHI), a PFirst award from the Office of the Provost, Duke University and an NICHD grant 1R01HD068523.

## APPENDIX: CONSTRUCTION OF NSM SAMPLING WEIGHTS

We employ bootstrap sampling weights to improve sample proportion estimates from the fNSM sample. Similar approaches have been attempted in the respondent driven sampling (RDS) literature with different types of data input. Bootstrap sampling approaches have long been used in an effort to approximate RDS's sampling variance under a first order Markov model on the nodal attribute of interest (cf. Salganik 2006; Yamanis et al. 2013, Appendix). Other work has focused on using bootstrap sampling to obtain mean estimators that are more robust to violations of RDS assumptions. For instance, Gile (2011) used bootstrap sampling with probability proportional to degree without replacement on an estimated population degree distribution in order to approximate the constraints imposed on RDS by it being conducted without replacement (cf. Gile and Handcock 2010). Another approach uses

Exponential Random Graph Modeling (ERGM) on RDS recruitment data and attribute homophily to estimate network parameters and simulate synthetic population social networks, which are then repeatedly resampled with simulated RDS chains. This approach was used to construct an RDS estimator which weights nodes by the number of simulated chains in which they were sampled (Gile and Handcock 2011) and was shown to outperform standard RDS estimators. In a related vein, Merli et al. (2015) used ERGM to simulate the network combined with repeated RDS resampling over this network to assess the crucial RDS assumption that nodes are sampled with probability proportional to degree.

However, these approaches remain fundamentally limited by the type of data collected by RDS, a myopic trace of a potentially small portion of the network (Verdery et al. 2015). They also restrict the type of post-estimation strategies that can be used to correct biases induced by the idiosyncrasies of the RDS sample recruitment process over each given population network.

NSM has two main advantages over RDS or other naïve random walk approaches: It collects complete network data and enables one to direct the sample to explore the network more fully, thereby ensuring a more complete picture of the network of interest (Mouw and Verdery 2012).

We take advantage of these unique features of NSM data combined with an ERGM framework to construct the bootstrap sampling weights and proceed as follows:

1. Run an ERGM model on the empirical network data provided by CAHS respondents. This included indicator variables for nodes which were sampled vs. nodes which were nominated but unsampled.
2. Predict ties between nominated but unsampled nodes using this model and the simulation applications available from ERGM (cf. Handcock et al., 2008) and overlay this atop the actual network observed between nominated nodes (i.e., only ties between nominated but unsampled nodes are predicted).
3. Use these procedures to construct 100 synthetic networks.
4. Simulate the NSM sampling algorithm five times each upon these predicted networks, tracking the number of times each node was sampled ( $U_i$ ) and the other types of data collected by NSM.
5. Use the number of times each node is sampled in a generalized Horvitz Thompson estimator of the form:

$$\hat{\mu}_{NSM} = \frac{\sum_i y_i / \bar{\pi}_i}{\sum_i 1 / \bar{\pi}_i}$$

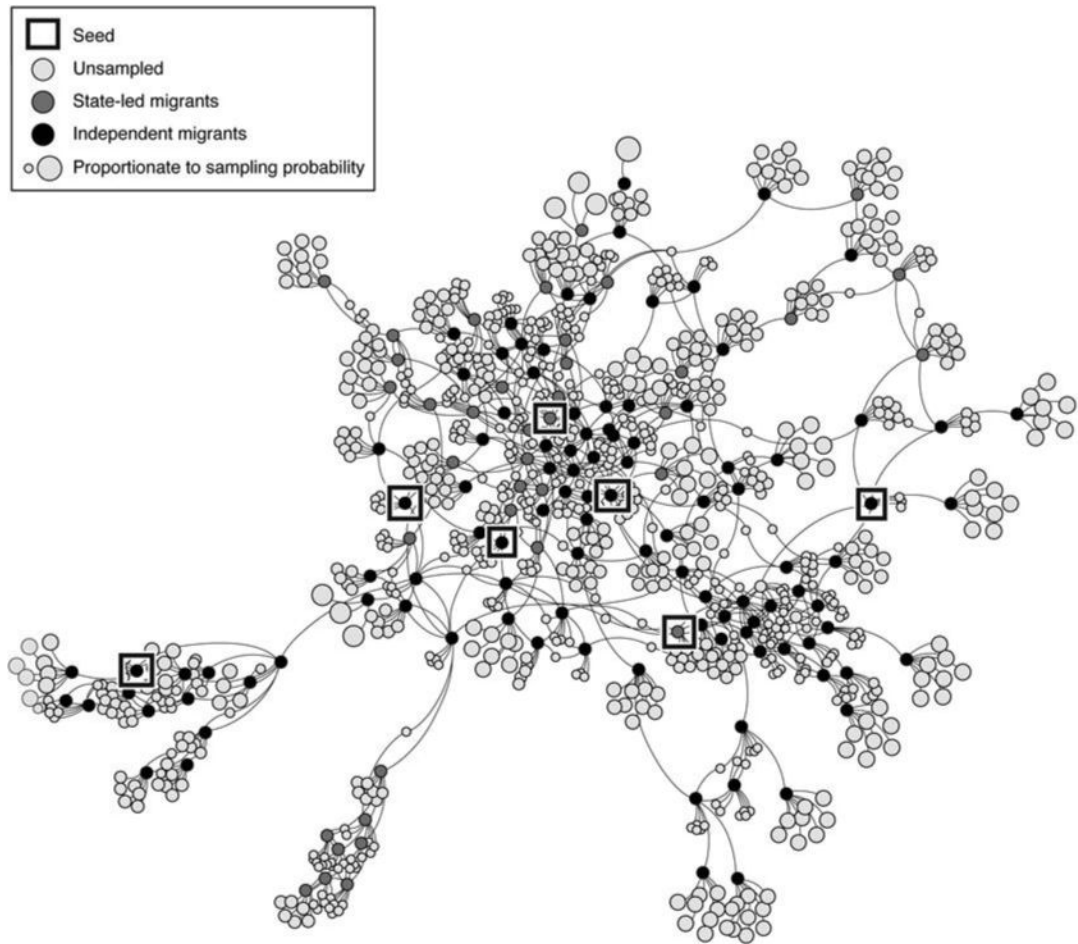
where  $\bar{\pi}_i$  is the proportion of times node  $i$  was sampled in the NSM bootstraps (i.e.,  $U_i/500$ ).

## References

- BBC. Tanzania orders Chinese out of Dar es Salaam market. 2011. <http://www.bbc.co.uk/news/world-africa-12138811>. Accessed on April 1, 2014
- Biggeri, Mario; Sanfilippo, Marco. Understanding China's move into Africa: An empirical analysis. *Journal of Chinese Economic and Business Studies*. 2009; 7(1):31–54.
- Battat, J. China's Outward Foreign Direct Investment Foreign Investment Advisory Services. The World Bank; 2006. 2006
- Brautigam, DA. *Dragon's Gift: The Real Story of China in Africa*. Oxford University Press; USA: 2009.
- Brautigam DA. Africa's Eastern Promise. *Foreign Affairs*. Jan 5.2010 (2010).
- Brautigam DA, Tang XY. China's Engagement in African Agriculture: "Down to the Countryside". *The China Quarterly*. 2007; 199:686–706.
- Brautigam DA, Tang XY. African Shenzhen: China's Special Economic Zones in Africa. *Journal of Modern African Studies*. 2011; 49(1):27–54.
- Brautigam, DA. Web log comment. China in Africa: The Real Story. 2014. N.p., n.d. Web. 19 Apr. 2014 <http://www.chinaafricarealstory.com/p/chinese-fdi.html>
- Bredeloup S, Bertoncello B. La Migration Chinoise en Afrique: Accélérateur du Développement ou 'Sanglot de l'Homme Noir'? *Afrique Contemporaine*. 2006; 218:199–224.
- Brown, Alison; Lyons, Michal; Dankoco, Ibrahima. Street Traders and the Emerging Spaces for Urban Voice and Citizenship in African cities. *Urban Studies*. 2010; 47(3):666–683.
- Cameron AC, Gelbach JB, Miller DL. Robust inference with multiway clustering. *Journal of Business & Economic Statistics*. 2011; 29(2)
- Campbell, Karen E.; Lee, Barret A. Name Generators in Surveys of Personal Networks. *Social Networks*. 1991; 13:203–221.
- CASS (Chinese Academy of Social Sciences). *Zhongguo guoji yimin baogao No. 1 [Annual Report on Chinese International Migration, 2012, No. 1]* [in Chinese]. Beijing: Social Sciences Academic Press; 2012.
- Daw J, Margolis R, Verdery AM. Siblings, friends, course-mates, club-mates: How adolescent health behavior homophily varies by race, class, gender, and health status. *Social Science & Medicine*. 2015; 125:32–39. [PubMed: 24673889]
- DeFries JC, Fulker DW. Multiple regression analysis of twin data. *Behavior genetics*. 1985; 15(5):467–473. [PubMed: 4074272]
- Dombrowski K, Khan B, Wendel T, McLean K, Misshula E, Curtis R. Estimating the Size of the Methamphetamine-Using Population in New York City Using Network Sampling Techniques. *Advances in Applied Sociology*. 2012; 2:245–252. [PubMed: 24672746]
- Entwisle, Barbara; Faust, Katherine; Rindfuss, Ronald; Kaneda, Toshiko. Networks and Contexts: Community Variation in the Structure of Social Ties. *American Journal of Sociology*. 2007; 112(5):1495–1533.
- Esoh, Elame. L'immigration en Afrique Noire dans le context de la mondialisation. 2005. [www.lestamp.com/publications\\_modialisation/publications.esoh.htm](http://www.lestamp.com/publications_modialisation/publications.esoh.htm)
- French, HW. *China's Second Continent How a Million Migrants Are Building a New Empire in Africa*. New York: Alfred A. Knopf; 2014.
- Gile KJ, Handcock MS. Respondent-Driven Sampling: An Assessment of Current Methodology. *Sociological Methodology*. 2010; 40:285–327. [PubMed: 22969167]
- Gile KJ, Handcock MS. Network Model-Assisted Inference from Respondent-Driven Sampling Data. 2011 Unpublished manuscript.
- Gile KJ. Improved Inference for Respondent-Driven Sampling Data With Application to HIV Prevalence Estimation. *Journal of the American Statistical Association*. 2011; 106(493):135–146.
- Goel S, Salganik MJ. Assessing Respondent-Driven Sampling. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(15):6743–6747. [PubMed: 20351258]
- Gong, Sasha. 2007. <http://128.97.165.17/media/files/80.pdf>. Accessed on April 15, 2014

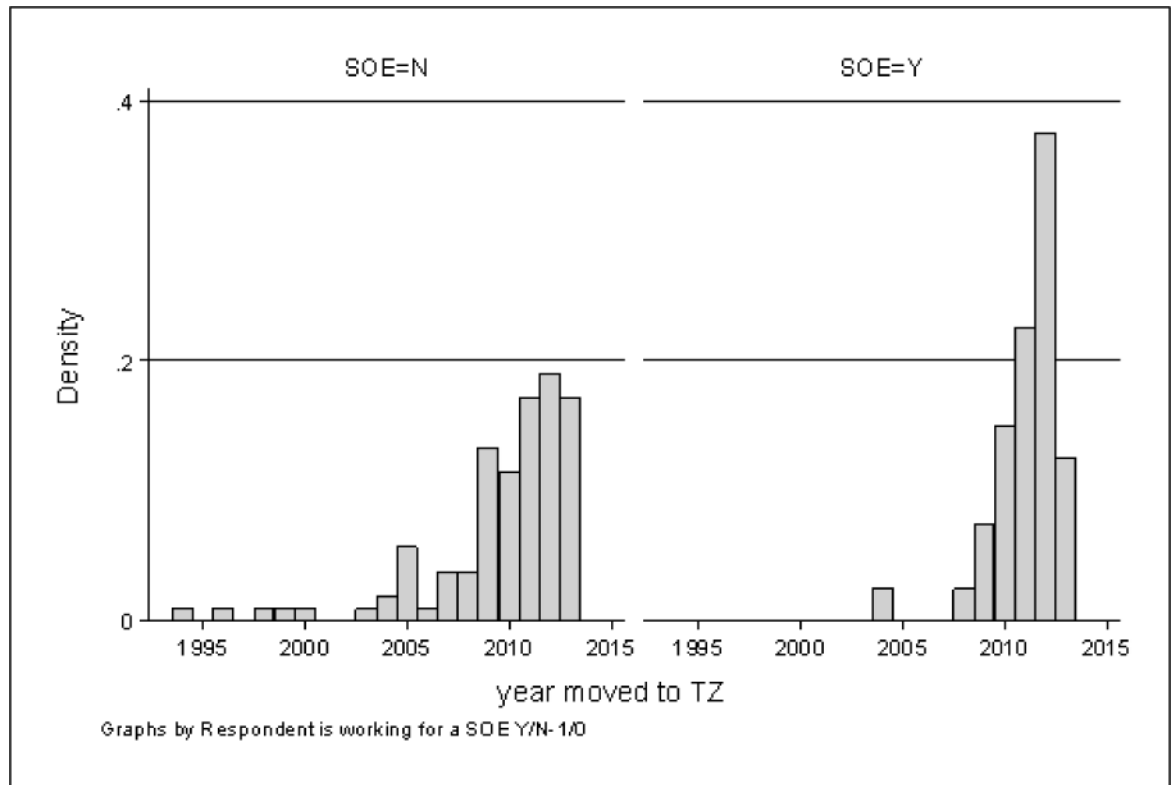
- Handcock MS, Goodreau SM, Hunter DR, Butts CT, Morris M. ERGM: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. *Journal of Statistical Software*. 2008; 24:1–29. [PubMed: 18612375]
- Haugen HO, Carling J. On the edge of the Chinese diaspora: The surge of baihuo business in an African city. *Ethnic and Racial Studies*. 2005; 28(4):639–662.
- Heckathorn DD. Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems*. 1997; 44:174–199.
- Heckathorn DD. Respondent-driven sampling II: Deriving valid population estimates from chain-referral samples of hidden populations. *Social Problems*. 2002; 49:11–34.
- Hsu E. Zanzibar and its Chinese Communities. *Population, Space and Place*. 2007; 13:113–124.
- Hsu, E. Medicine as business: Chinese medicine in Tanzania. In: Alden, C.; Large, D.; Soares de Oliveira, R., editors. *China Returns to Africa: A Superpower and a Continent Embrace*. C. Hurst and Company; London: 2008. p. 221–235.
- Human Rights Watch. “You’ll Be Fired if You Refuse” Labor Abuses in Zambia’s Chinese State-owned Copper Mines. New York: Nov. 2011 (2011)
- Jampaklay, Aree; Korinek, Kim; Entwisle, Barbara. Residential Clustering in Urban Settings of Thailand. *Asian and Pacific Migration Journal*. 2007; 16(4):485–510.
- Lee CK. Raw Encounters: Chinese Managers, African Workers and the Politics of Casualization in Africa’s Chinese Enclaves. *The China Quarterly*. 2009; 199:647–666.
- Li PT. The development of Sino-African relations and New Chinese Migrants in Africa. *Overseas Chinese History Studies*. 2010; 4:26–32. (In Chinese).
- Liang, Zai; Morooka, Hideki. Recent Trends of Emigration from China: 1982–2000. *International Migration*. 2004; 42(3):145–163.
- Liang, Zai; Chunyu Miao, David; Zhuang, Guotu; Ye, Wenzhen. Cumulative Causation, Market Transition, and Emigration from China. *American Journal of Sociology*. 2008; 114(3):706–737.
- Lin, Edwin. “Big Fish In a Small Pond”: Chinese migrant shopkeepers in South Africa. *International Migration Review*. 2014; 48(1):181–215.
- Lin, Jan. *Reconstructing Chinatown: Ethnic Enclave, Global Change*. University of Minnesota Press; Minneapolis: 1998.
- Lu, Yao; Liang, Zai; Chunyu Miao, David. Emigration from China in comparative perspective. *Social Forces*. 2013; 92(2):631–658.
- Ma Mung, Emmanuel. Chinese migration and China’s foreign policy in Africa. *Journal of Overseas Chinese*. 2008; 4(1):91–109.
- Marin, Alexandra. Are respondents more likely to list alters with certain characteristics?: Implications for name generator data. *Social Networks*. 2004; 26(4):289–307.
- Marsden, Peter V. The reliability of network density and composition measures. *Social Networks*. 1993; 15:399–421.
- McCreech N, Frost SD, Seeley J, Katongole J, Tarsh MN, Ndunguse R, Jichi F, Lunel NL, Maher D, Johnston LG, Sonnenberg P, Copas AJ, Hayes RJ, White RG. Evaluation of Respondent-Driven Sampling. *Epidemiology*. 2012; 23(1):138–147. [PubMed: 22157309]
- McPherson, Miller; Smith-Lovin, Lynn; Cook, James M. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*. 2001; 27:415–44.
- Merli MG, Moody J, Smith J, Li J, Weir S, Chen X. Challenges to recruit representative samples of female sex workers in China using Respondent Driven Sampling. *Social Science and Medicine*. 2015; 125:79–93. [PubMed: 24834869]
- Mohan, Giles; Lampert, Ben; Tan-Mullins, May; Chang, Daphne. *Chinese Migrants and Africa’s Development. New Imperialists or Agents of Change?*. London: Zed Books; 2014.
- Monson, Jamie. *Africa’s Freedom Railway*. Indiana University Press; Bloomington and Indianapolis: 2009.
- Morris M, Kurth AE, Hamilton DT, Moody J, Wakefield S. Concurrent partnerships and HIV prevalence disparities by race: linking science and public health practice. *American Journal of Public Health*. 2009; 99(6):1023–1031. [PubMed: 19372508]

- Mouw T, Verdery A. Network Sampling with Memory: A proposal for more efficient sampling from social networks. *Sociological Methodology*. 2012; 42(1):206–256. [PubMed: 24159246]
- Mouw, Ted; Verdery, Ashton M.; Giovanna Merli, M.; Li, Jing; Shen, Jennifer. Sampling a Hidden Population without a Sampling Frame: A Practical Application of Network Sampling with Memory. Paper Presented at the Population Association of America; Boston, MA. May 1–3 2014; 2014b.
- Pairault T. Chinese direct investment in Africa. A state strategy? *Région et Développement*. 2013; 37:259–284.
- Park, Yoon Jung. Chinese Migration in Africa. 2009. SAIIA Occasional Paper No. 24
- Paul, M. The Dongbei: The new Chinese immigration to Paris. In: Nyiri, P.; Savelieve, IR., editors. *Globalising Chinese Migration*. Ashgate; Surrey, UK: 2002. p. 120-128.
- Pieke, Frank N. Editorial Introduction: Community and identity in the new Chinese migration order. *Population, Space and Place*. 2007; 13:81–94.
- Pieke, Frank N.; Speelman, Tabitha. Report for the Migration Policy Center. European University Institute; Florence: Jan. 2013 Chinese investment strategies and migration: Does diaspora matter?. (2013)
- Politzer, Malia. China and Africa: Stronger Economic Ties Mean More Migration. Migration Policy Institute; 2008. 2008
- Radelet, Steven. *Emerging Africa*. Center for Global Development; Washington D.C: 2010.
- Ross, Anthony. Infrastructure and influence. China's presence on the coast of East Africa. *Journal of the Indian Ocean Region*. 2013; 9(20):134–149.
- Salganik MJ. Variance estimation, design effects, and sample size calculation for respondent-driven sampling. *Journal of Urban Health*. 2006; 83:98–111.
- State Council of China. *China Africa Economic and Trade Cooperation*. 2013
- Straights, Bruce. Ego's important discussants or significant people: an experiment in varying the wording of personal network name generators. *Social Networks*. 2000; 22(2):123–140.
- Sun, Yun. *Africa in China's Foreign Policy*. Brookings; Washington, D.C: 2014.
- Verdery, Ashton M.; Mouw, Ted; Bauldry, Shawn; Mucha, Peter J. Network Structure and Biased Variance Estimation in Respondent Driven Sampling. *PLoS ONE*. 2015:e0145296. [PubMed: 26679927]
- Volz E, Heckathorn DD. Probability based estimation theory for Respondent Driven Sampling. *The Journal of Official Statistics*. 2008; 24(1):79–97.
- Wu, Xiaogang; Treiman, Donald. The Household Registration System and Social Stratification in China, 1955–1996. *Demography*. 2004; 41:363–384. [PubMed: 15209045]
- Wu, Xiaogang; Treiman, Donald. Inequality and Equality under Chinese Socialism: The Hukou System and Intergenerational Occupational Mobility. *American Journal of Sociology*. 2007; 113(2):415–45.
- Xiang, Biao. Emigration from China: A sending country perspective. *International Migration*. 2003; 41(3):21–28.
- Yamanis N, Merli MG, Neely WW, Tian FF, Moody J, Tu X, Gao E. An empirical analysis of the impact of recruitment patterns on RDS estimates among a socially ordered population of female sex workers. *Sociological Methods and Research*. 2013; 42(3):392–425.
- Young SK, Lyles RH, Kupper LL, Keys JR, Martin SL, Costenbader EC. Assortativity coefficient-based estimation of population patterns of sexual mixing when cluster size is informative. *Sexually Transmitted Infections*. 2014; 90(4):332–336. [PubMed: 24482487]
- Zhong, Nan. Two new funds to boost trade with Africa. *China Daily*. Jun 13th.2013 (2013).
- Zhou, Min. *Chinatown: The Socioeconomic Potential of an Urban Enclave*. Temple University Press; Philadelphia: 1992.
- Zweig, D. New 'Trading State' Meets the Developing World. Hong Kong University of Science and Technology, Center on China's Transnational Relations; 2010. Working Paper No. 31



**Figure 1.**  
CAHS network of 1,000 uniquely nominated nodes, including 147 interviewed nodes colored by ownership sector of current employment





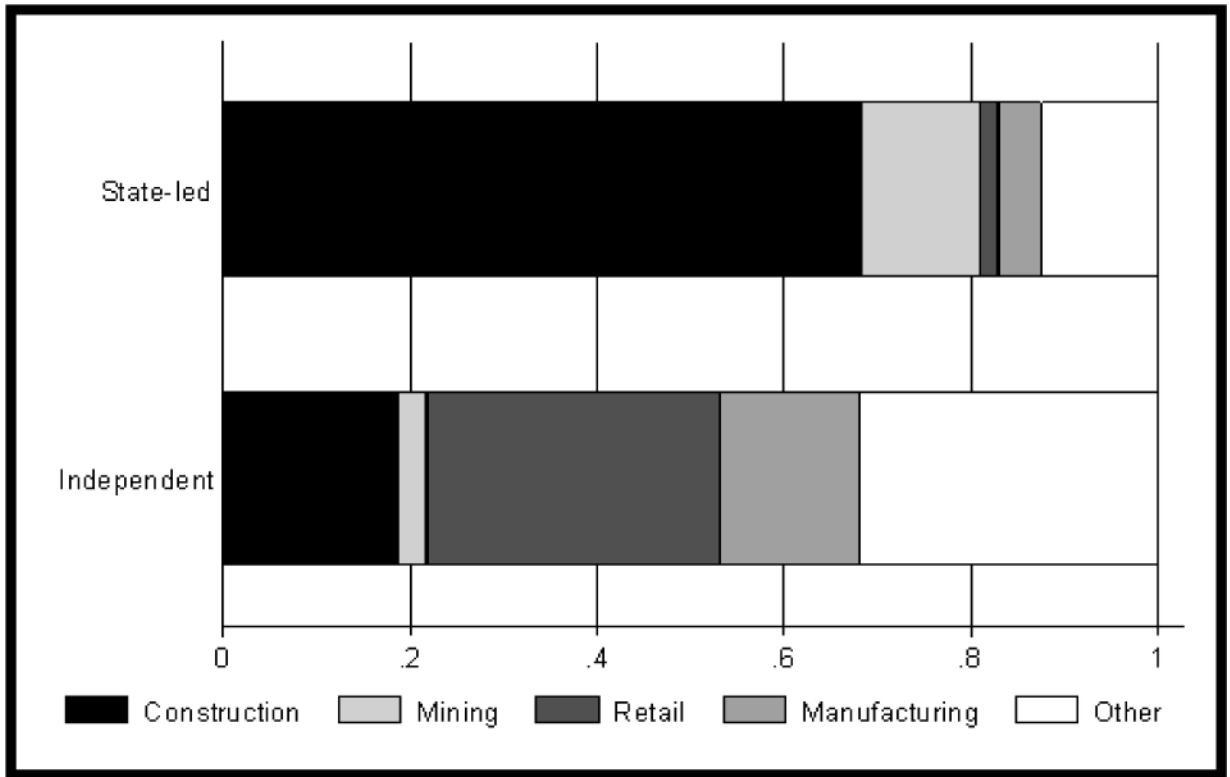
**Figure 2.** Year moved to Tanzania (this migration spell) by ownership sector of current employment, CAHS 2013 (N=145)

Author Manuscript

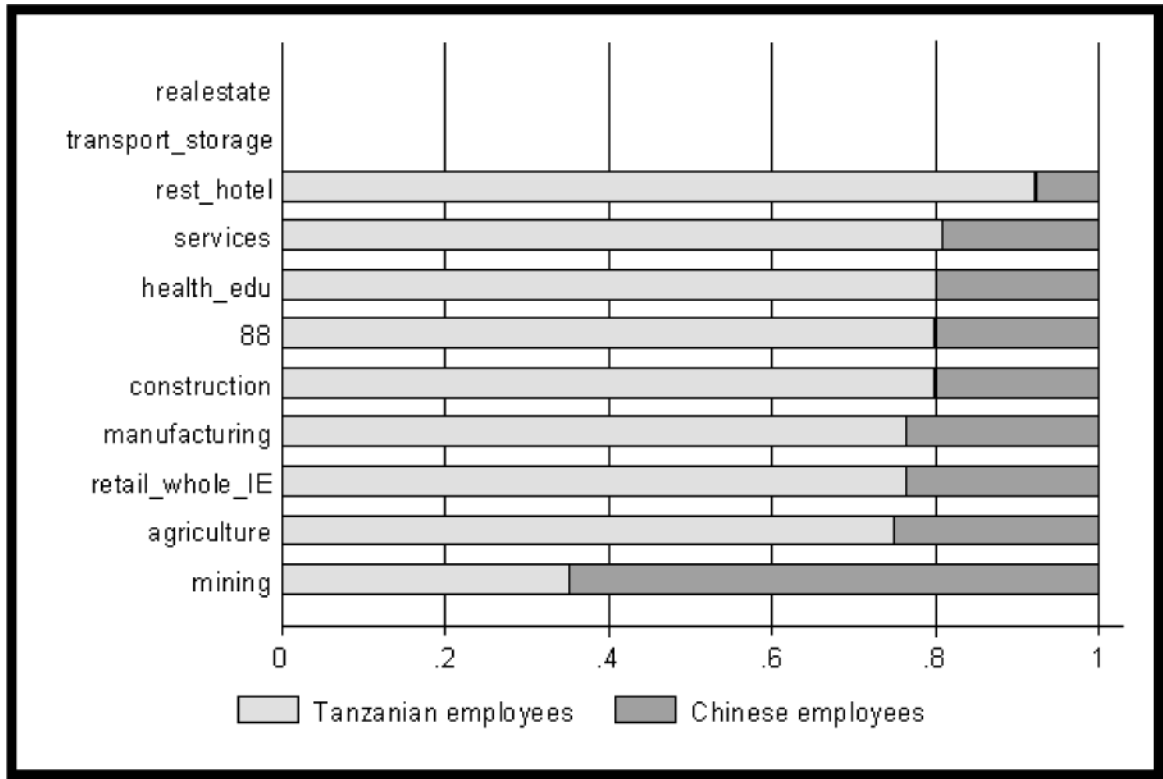
Author Manuscript

Author Manuscript

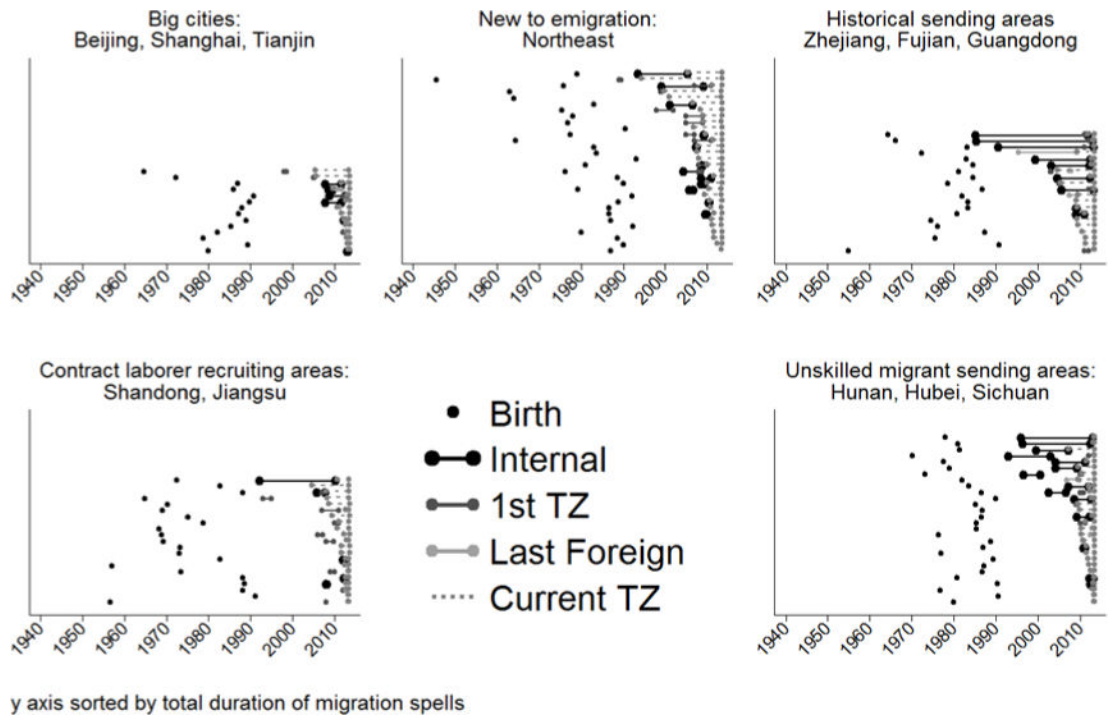
Author Manuscript



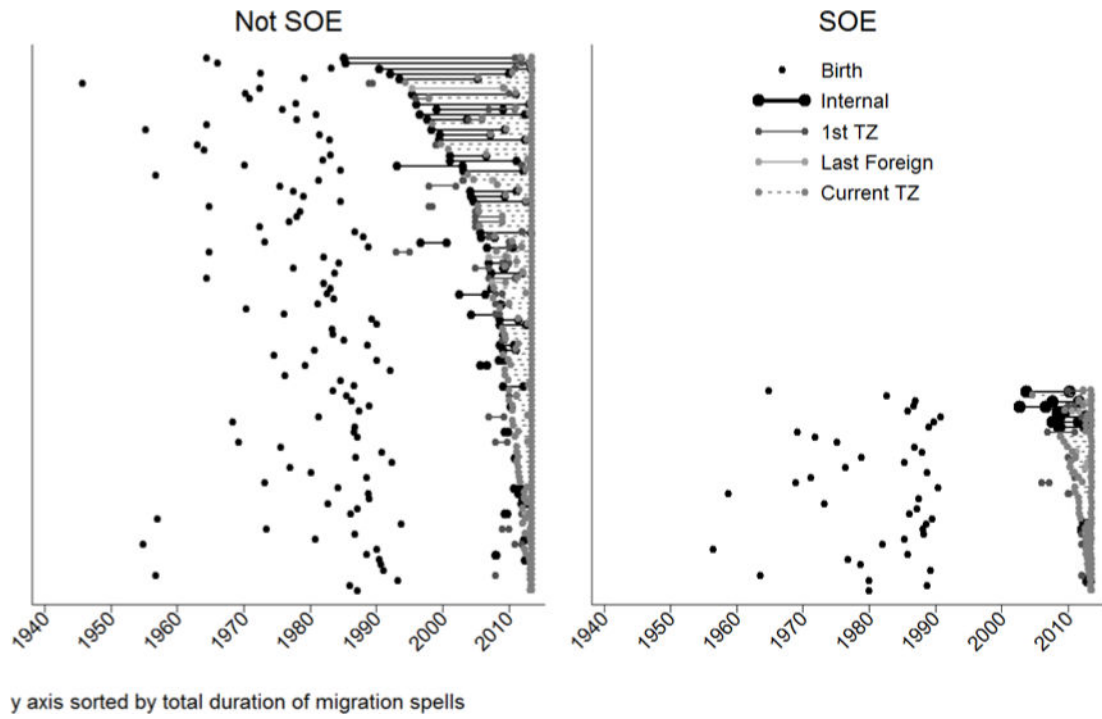
**Figure 3.** Proportionate distribution of Chinese state-led and independent migrants by economic sector, CAHS 2013 (N=145). NSM adjusted proportions



**Figure 4.** Proportionate distribution of Tanzanian and Chinese employees. Asked of 36 business owners in the sample, CAHS 2013



**Figure 5.** Migration spells for CAHS respondents by type of migration spell and Chinese province of origin (N=111)



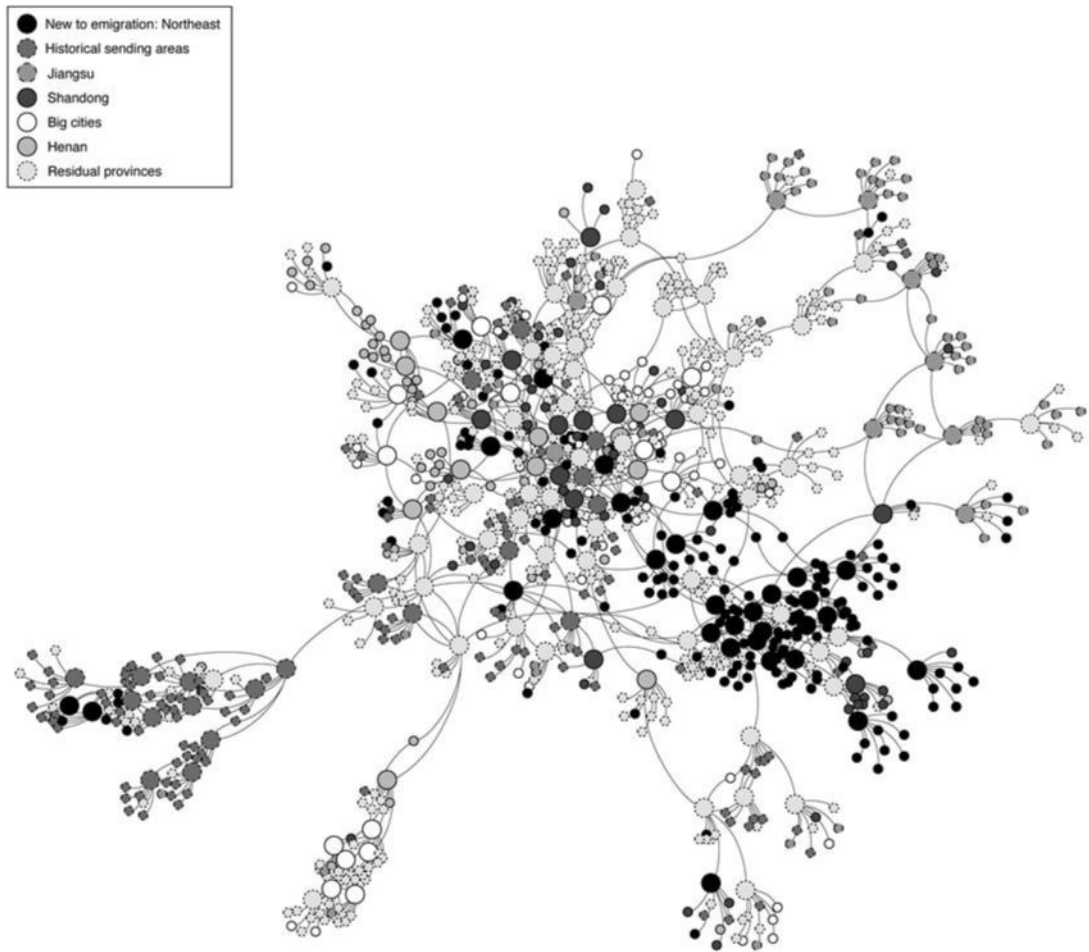
**Figure 6.** Migration spells for CAHS respondents by type of migration spell and ownership sector of current employment (N=145)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 7.**  
CAHS network of 1,000 uniquely nominated nodes colored by Chinese province of permanent residence



**Table 1**

Characteristics of CAHS sample by ownership sector of current employment, NSM adjusted percentages and means (unadjusted percentages and means)

	State-led	Independent	Total
<b>Count</b>	40	105	147
<b>% in group</b>	26.5 (27.2)	72.2 (71.4)	98.6 (98.6)
<b>DEMOGRAPHIC CHARACTERISTICS</b>			
<b>Age (mean)</b>	30.9 (32.0)	34.4 (33.9)	33.3 (33.2)
<b>% Male</b>	90.7 (87.5)	77.8 (82.9)	81.1 (83.7)
<b>% currently married</b>	43.8 (55.0)	68.1 (60.0)	60.7 (57.8)
<b>Years of schooling (mean)</b>	15.9 (15.6)	11.6 (12.3)	12.8 (13.2)
<b>% with a college degree</b>	85.5 (82.5)	30.2 (38.1)	44.9 (50.3)
<b>Duration in Tanzania this spell (mean years)</b>	1.7 (1.9)	3.7 (3.4)	3.3 (3.0)
<b>PERMANENT RESIDENCE IN CHINA</b>			
<b>% by hukou province</b>			
Northeast China	0.0 (0.0)	25.5 (26.7)	19.8 (20.4)
Zhejiang/Fujian/Guangdong	0.0 (0.0)	21.4 (17.1)	15.4 (12.2)
Shandong/Jiangsu	26.9 (20.0)	15.0 (12.4)	18.0 (14.3)
Hunan/Henan/Sichuan	16.1 (20.0)	18.4 (19.1)	17.5 (19.1)
Beijing/Shanghai/Tianjin	20.2 (27.5)	2.4 (2.9)	7.1 (9.5)
Other provinces	36.8 (32.5)	17.3 (21.9)	22.2 (24.5)
<b>% with urban hukou</b>	86.2 (82.5)	44.8 (44.8)	55.1 (54.4)
<b>EXPERIENCE OF MIGRATION IN CHINA</b>			
<b>% w/internal migration experience in China</b>	24.9 (32.5)	50.0 (49.5)	42.7 (44.2)
<b>% whose most recent residence in China is not in the place where they grew up</b>	42.4 (40.0)	39.5 (38.1)	41.1 (39.5)
<b>LIVING ARRANGEMENTS IN DAR ES SALAAM</b>			
<b>% living in collective housing provided by employer (e.g. construction site, dormitory)</b>	50.1 (50.0)	38.7 (41.9)	42.1 (44.2)
<b>% of Chinese cohabiters in living place (mean)</b>	94.1 (96.7)	94.7 (94.7)	94.6 (95.3)

Notes: Two respondents did not report ownership sector because they were not working at the time of the interview. Totals may not add up exactly owing to rounding error.

**Table 2**

Adjusted observed-expected ratios of nominations by ownership sector of current employment (unadjusted row percentages).

	<b>Independent migrant</b>	<b>State-led migrant</b>	<b>Raw counts (total)</b>
Independent	1.12 (87.3)	0.51 (12.7)	189 (100.0)
SOE	0.62 (46.8)	2.33 (53.2)	77 (100.0)
Raw counts (total)	201	65	266

N=total number of dyads between sampled nodes who reported ownership sector.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3**

Adjusted observed-expected ratios of nominations by province group (unadjusted row percentages). N= total number of dyads between nominated and sampled nodes

	New to emigration (Northeast provinces)	Historical sending areas (Zhejiang, Fujian, Guangdong)	Jiangsu	Shandong	Big Cities (Beijing, Tianjin, Shanghai)	Henan	Residual provinces	Raw Counts (Total)
New to emigration	3.42 (64.93)	0.40 (6.72)	0.07 (0.75)	0.54 (4.48)	0.20 (1.49)	0.36 (2.24)	0.54 (19.40)	268 (100.0)
Historical sending areas	0.27 (5.81)	3.21 (51.74)	0.61 (4.07)	0.89 (7.56)	0.32 (2.33)	0.41 (2.33)	0.68 (26.16)	172 (100.0)
Jiangsu	0.22 (3.85)	0.27 (3.85)	8.35 (65.38)	1.09 (7.69)	0.20 (1.28)	0.00 (0.00)	0.52 (17.95)	78 (100.0)
Shandong	0.36 (6.12)	0.77 (14.29)	0.72 (5.10)	3.05 (23.47)	1.93 (16.33)	0.85 (3.06)	0.87 (31.63)	98 (100.0)
Big Cities	0.31 (6.25)	0.34 (6.25)	0.32 (3.13)	0.52 (4.69)	2.56 (17.19)	0.43 (3.13)	1.72 (59.38)	128 (100.0)
Henan	0.28 (5.56)	0.41 (7.41)	0.10 (0.93)	1.52 (12.04)	1.53 (10.19)	5.89 (31.48)	0.87 (32.41)	108 (100.0)
Residual provinces	0.50 (10.47)	0.99 (17.91)	0.86 (4.88)	0.82 (6.98)	1.11 (6.05)	0.82 (5.81)	1.35 (47.91)	430 (100.0)
Raw counts (Total)	252	217	91	103	84	76	459	1,282

Note: The adjusted matrix was generated using a modification to the Within Cluster Resampling (WCR) approach developed by Young et al. (2014). 113 nominations with missing information on provinces were collapsed into the residual province category. For the residual category of provinces, the matching between dyads was done according to whether both nodes were in any of the residual provinces. N= total number of dyads between nominated and sampled nodes who reported province.

**Table 4**

Estimates from DeFries-Fulker style homophily regressions with two-way clustered standard errors

Coefficient	Variable	(1)	(2)
$\alpha$	<i>Constant</i>	-0.943 <sup>**</sup> [0.19]	-0.907 <sup>**</sup> [0.19]
$\beta_1$	$Y_j$	-0.084 <sup>*</sup> [0.04]	-0.163 <sup>**</sup> [0.04]
$\beta_2$	$T_{ij}$	-0.719 <sup>**</sup> [0.24]	-0.522 <sup>+</sup> [0.27]
$\beta_3$	$Y_j T_{ij}$	2.099 <sup>**</sup> [0.36]	1.347 <sup>**</sup> [0.45]
$\beta_4$	$P_{ij}$		-0.700 <sup>**</sup> [0.22]
$\beta_5$	$Y_j P_{ij}$		1.509 <sup>**</sup> [0.34]
$\beta_6$	$T_{ij} P_{ij}$		-0.486 [0.55]
$\beta_7$	$Y_j T_{ij} P_{ij}$		1.768 <sup>*</sup> [0.87]
	Observations	20880	20880
	Log-likelihood	-12256.5	-12188.6
	LR chi2		124.03 <sup>**</sup>
	BIC	-23.58	-32.38

Notes: Robust two-way clustered standard errors in brackets;

<sup>\*\*</sup>  
p<0.01,<sup>\*</sup>  
p<0.05,<sup>+</sup>  
p<0.1.

N=total number of dyads between sampled nodes who reported sector and province.

**Table 5**

Predicted probabilities of ties within and between ownership sectors and in same or different province of origin

	Predicted probability of being a state-led migrant	Predicted probability of being an independent migrant
Conditional on being tied to:		
State led migrant from same province	0.864	0.136
Independent migrant from same province	0.068	0.932
State-led migrant from different province	0.439	0.561
Independent migrant from different province	0.193	0.807

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript