Abstract

Title: Application of the network scale-up method to estimate hard-to-count populations: the case of Brazilian immigrants in Florida, USA

Dimitri Fazito

Christopher McCarty

According to the US Census Bureau there were 212,428 Brazilian immigrants in the country in 2000 but concurrently the Brazilian government (extra-officially), the media and varied researchers argued against underestimate and claimed different numbers ranging from 400,000 to 1,500,000. In this study we propose to estimate the subpopulation of Brazilian immigrant residents in Florida, USA (including estimates about Mexican immigrants) based on the network scale-up (NSU) method. This method aims to estimate the size of (hard to count) unknown subpopulations assuming that the fraction of these subpopulations presented in the averaged active social networks of the general population members could be scaled up to their real sizes. In this paper we present the NSU rationale, explain its general application to a case study and conclude with remarks about the use of NSU in demography.

Extended Abstract

In this study we propose to estimate the subpopulation of Brazilian immigrant residents in Florida, USA (including estimates about Mexican immigrants) based on the network scale-up (NSU) method.

This method aims to estimate the size of (hard to count) unknown subpopulations assuming that the fraction of these subpopulations presented in the averaged active social networks of the general population members could be scaled up to their real sizes.

This is a rather new method to estimate population size in demography but does not use the traditional demographic techniques. It is entirely based on the concept intuition that "social interaction patterns" accounts for demographic behaviors and might be modeled to explain subtle associations between social factors and vital events (Boogarts and Watkins, 1996). The core of NSU rationale is that people's active (local) social networks should represent at some extent the global social environment – like a social structure – in which population takes form (i.e., the formation of population categories and segments and the categorization of demographic behaviors and events). Thus, the socio-demographic composition (attributes) of personal networks should reflect the general distribution of such attributes across the overall population. Consequently, modeling the association between personal contacts and individuals' attributes in one's social network should provide some information about the global social and demographic patterns (Kadushin et al., 2006; Moody, 2006; McCormmick and Zheng, 2007).

Network Scale-Up Rationale

According to McCarty et al. (2001:28) a major component of the network scale-up method "involves estimating the average size of personal networks for a large sample of people" and then identify the average number of people the respondent knows in a particular subpopulation. Thus, the assumption is that the number of people known by the respondent in a particular subpopulation is a function of the number of all people known in her personal network (that is, the size of personal networks).

Killworth et al. (1998a, 1998) tested thoroughly the assumption that, other things being equal, "the probability that any member of the respondent's network is in a subpopulation is the fraction of the larger population occupied by the subpopulation (in other words, if $1/100^{th}$ of the US population have some characteristic, then on average one would assume $1/100^{th}$ of any network to possess that characteristic also)" (McCarty et al. 2001:28).

The simple fractional model is $m/c = e/t$, where "m" is the reported number of people known in a subpopulation (network alters in the specific subpopulation), "c" is the size of the respondents' personal network, "e" is the size of the subpopulation and "t" is the size of the larger population.

In addition, the model above must satisfy 3 assumptions (Killworth et al. 1998):

- Everyone has the same probability of knowing someone in a given population;
- The size of the personal network (parameter "c") must be constant;
- Everyone is supposed to have "perfect" information about the alters in the network;

These assumptions are not evenly met across different populations. The violation of the first assumption is caused by "barrier effects" (i.e., there are spatial or sociodemographic characteristics that prevent people to know about each other) and "transmission effects" (i.e., for different reasons information across the personal networks are not transmitted evenly – people avoid sensitive questions, like HIV+ people who avoid to talk about their health conditions even to closer friends). The problem due to "transmission effects" is the most difficult to manage but efforts have been tackled it (Killworth et al. 2006; McCornnick and Zheng, 2007).

The second assumption depends on dynamics in social life, regarding individual's levels of social activity. However, as suggested by Dunbar (1998), usually people organize their social contacts (network of alters) according to hierarchical "layers of social inclusion", ranging from a very intimate and small core of active contacts to a wide, heterogeneous and weak-ties social environment. Hill and Dunbar (2003) supports that people's active social networks range between 150-200 contacts since human neocortex is not able to process (i.e., identify, classify and manipulate) detailed information about social contacts beyond that threshold which is coincident with the periphery of the wider social environment. Moreover, studies have also

shown that regular people in US present an average personal network size around 290 contacts (Killworth et al. 1998; McCarty et al., 2001; Kadushin et al. 2006).

The third assumption could be violated variously but is mainly a consequence of the "estimate effect" (people might be unable to define accurately the boundary of their networks and the "label" of their alters). Anyway this is a very common problem in the social surveys and can be prevented using traditional survey techniques.

Therefore, in order to estimate the subpopulation size (hard-to-count population "e") it is necessary to have the parameters "t", "m" and "c" in hand. Indeed, the total population "t" is not a problem at all as long as one can trust on census information. In addition, probabilistic surveys can estimate accurately the average "m" across the overall population "t" (for example, the average number of diabetics, homeless or Brazilian immigrants the people know). However, the problem for social network theory in general is to find out a reliable estimate for parameter "c", the personal network size (Killworth and Bernard, 1997). Although, barrier, transmission and estimate effects may produce "noise" around "c", scientists developed two methods to estimate reliably the parameter "c" in populations: estimate from scale-up known subpopulations and estimate from "summation" of type of relations (Killworth et al. 1998a, 1998b; Johnsen et al. 1995; McCarty et al., 2001).

1. Estimating "c" from scale-up method: knowing "m", "t" and known "e" (i.e., known subpopulations like married people, population of doctors, homeowners, etc) will allow us to use the simple model equation above to provide maximum likelihood estimates on different known subpopulations which maximizes the probability of finding an accurate averaged "c" (see, Killworth et al, 1998:292-293).
2. Estimating "c" from summation of types of relations: direct estimate of "c" can also be acquired using probabilistic surveys to ask respondents to "count and estimate" the number of people they know for each type of relation (like, intimate family, extended family, friendship, neighborhood, co-workers, etc) and then summing up – McCarty et al. 2001.

Across 7 independent applied researches in the US (using both methods) the parameter "c" was systematically estimated in 291 – the authors tested for reliability and showed that this number is not a statistical artifact. That is, on average, the size of the personal network of the Americans is 291 people (McCarty et al. 2001). It is important to note that this personal network means the "active network" that was gathered asking the respondents about the "people you know and they know you by sight or by name. You have had some form of contact with them in the past two years and you could contact them if you had to".

Case Study: the Brazilian Immigrant Subpopulation Resident in Florida, USA

Studies on Brazilian immigration in US have argued that the official estimate from US Census data (sample) might underestimate the number of Brazilians due to the irregular status of many immigrants (Margolis, 1994, 2003; Martes, 2000; Soares, 2002). Because Brazil has

experienced a sudden and steady increase of international emigration since 1980, today many public authorities, the mass media and the organized civil society claim that Brazil is experiencing a "diaspora", especially the Brazilian "transnational communities" in US (Fazito & Soares, 2008). Despite the lack of accurate estimates of Brazilian international emigrants and immigrants those main actors claim for high numbers: currently, many authorities believe that there are 4 million Brazilians living abroad and around 1,300,000 only in US – some have already claimed 1,500,000 (CPMI, 2006).

Here we provide an estimate for Brazilian immigrants resident in Florida, in the beginning of 2009, based on the "summation" approach of NSU. We ran a probabilistic survey (using digit dialing) and interviewed 250 residents in Florida asking only about their personal network contacts living in Florida in the last 2 years. In order to estimate the average "c" we asked respondents to count their contacts in each of 14 categories of relations – like "intimate family living in the same household", "intimate family living in different household", "close friends", "regular friends you meet at school/office", "unusual people you meet in your neighborhood" and so on – and then summed them up.

We found an average parameter "c" (social network size) about 157 contacts and an average "m" (proportion of Brazilian residents in the respondents' Floridian social networks) about 0.68. Considering the expected resident population of Florida in 2009 around 18,870,610 we estimated the total subpopulation of Brazilian immigrants in Florida by 81,264.

According to the Census Bureau statistics there were about 212,000 Brazilian immigrants living in the US in 2000 - and 44,500 immigrants only in Florida, the most expressive American state for Brazilian immigration. If NSU is credible, then it seems that the Brazilian migration flows to Florida almost doubled in less than ten years or the Census statistics did not get the real number, indeed.

It is important to make sure that this NSU application is still tentative and should not discredit census findings. But the new technique should be taken seriously as an alternative approach to rigorous estimation of population size.

Bibliography

Bongaarts, J. and S. C. Watkins (1996). Social interactions and contemporary fertility transitions. Population and Development Review, 22(4):639–682.

Dunbar, R. (1998) The Social Brain Hypothesis. *Evolutionary Anthropology* 6:178–190.

Hill, R. & Dunbar, R. (2003) Social Network Size in Humans. *Human Nature* 14(1):53-72.

Kadushin, C., Killwork, P., Bernard, R., Beveridge, A. (2006) Scale-Up Methods as Applied to Estimates of Heroin Use. *Journal of Drug Issues*, 22:417-440.

Killworth, P. D., McCarty, C., Bernard, H. R., Shelly, G. A., and Johnsen, E. C. (1998). Estimation of seroprevalence, rape, and homelessness in the U.S. using a social network approach. *Evaluation Review*, 22:289-308.

Killworth, P. D., McCarty, C., Johnsen, E. C., Bernard, H. R., and Shelley, G. A. (2006). Investigating the variation of personal network size under unknown error conditions. *Sociological Methods & Research*, 35(1):84-112.

McCarty, C, Killworth, PD, Bernard, HR, Johnsen, E and Shelley, GA. (2001). Comparing Two Methods for Estimating Network Size, *Human Organization,* 60:28-39.

McCormick, T. H. and Zheng, T. (2007). Adjusting for recall bias in \how many X's do you know?" surveys. Joint Statistical Meetings: Salt Lake City, Utah.

Moody, J. (2006) Fighting a Hydra: a Note on the Network Embeddedness of the War on Terror. *Structure and Dynamics: e-Journal of Anthropological and Related Sciences*, vol. 1(2) article 9.