

**Estimating religious populations with the network scale-up
method: a practical alternative to self-report**

Journal:	<i>Journal for the Scientific Study of Religion</i>
Manuscript ID	JSSR-OA-11-2016-210.R1
Manuscript Type:	Original Article
Keywords:	network scale-up method, religious identity, survey methodology, measurement, social desirability

Estimating religious populations with the network scale-up method: a practical alternative to self-report

Abstract

Getting accurate information on religious demographics from survey-based self-reports presents a difficult task, suffering from the biases of social desirability, personal safety concerns, and the ambiguous definitions of religious identity. An estimation strategy based on the enumeration of social network ties, the network scale-up method (NSUM), has recently been employed to estimate the sizes of hidden populations in criminology and public health, but has not been utilized in the study of religion. In this study, we argue for the advantages of NSUM in overcoming the biases associated with self-reports, and lay out a practical guide for the scholars of religion to the design and calculation of NSUM. We used a recent survey of Chinese international students to illustrate the use of NSUM and estimate the percentage of Buddhists (4.3%) and Christians (8.6%) in this population. We recommend interested scholars to adopt NSUM for its reliability, easy implementation, and more importantly--the affinity between the sociological perspective on religion and the socially-oriented assumption of NSUM.

Key words: network-scale up method, survey methodology, measurement, religious identity, social desirability

Introduction

Getting data on individual religious beliefs, behaviors, and identifications has been hindered from measurement and conceptual obstacles. In the United States, questions on religions are in stock in almost every social survey, official records, and polls initiated by different interested parties, all contain religious information. However, self-reported religious behaviors and beliefs are notoriously inflated by the over-reported desirable images such as being a frequent and sociable church-goer (Brenner, 2011, 2014; C Kirk Hadaway, Marler, & Chaves, 1993, 1998; Presser & Stinson, 1996). Self-reports on religious identity can be deflated in contexts where some minority groups suffer from social and political stigma, such as Christians in China, Muslims in Sri Lanka and Myanmar, and atheists in the United States. Commitment to a religion hardly comes as a factor of the individuals' religious behaviors, whose normative values vary significantly across religious traditions, and thus measuring religious identity faces the fundamental problem of finding a universal criterion (Mockabee, Monson, & Grant, 2001). Recall bias—inaccurately reporting one's religious behaviors in retrospection—constitutes another common cognitive bias in answering questionnaires, which could be both intentional or unintentional (Presser & Stinson, 1996).

The problem with self-reports was evident more than 20 years ago when a group of scholars challenged the high rates of self-reported affiliation and attendance in the United States. Robert Wuthnow's 2015 book *Inventing American Religion* has a good summary review of that debate (Wuthnow, 2015). However, the discipline has been trapped in a quandary: we know the existing method is not good or sufficient but nothing better can be done.

In this study, we introduce a new method of estimating the prevalence of religious populations not based on self-reports. This network-based method shifts the burden of information from individuals themselves to other people connected to them (e.g. friends and acquaintances), thus avoids unreliable report due to social desirability. Meanwhile, it also collects self-reported information from to adjust for other types of biases against transparent reporting: primarily the barrier bias and transmission bias. To help practitioners, we will show to readers a practical guide to the application of this method in their own survey designs. For the purpose of illustration, in this paper we use this method to estimate the prevalence of Christians and Buddhists in a population of Chinese international students in the United States. We also compare these estimates with that generated by self-report, and discuss another conceptual contribution of NSUM: religious identity should be understood as a feature perceptible in the social world rather than merely an attribute of individuals. As Durkheim says, "religion is something eminently social" (Durkheim [1912]2008: 10).

Literature review

Underreporting religious identity in self-reports

The first type of misreports on religious identity arises from social stigma and safety concerns. Belonging to certain religious organizations in many countries is a self-sabotaging act if detected by authorities. Aside from a few democratic secular regimes, the majority of the world's nations retain restrictions on religious activities and affairs to a varying degree (Grim & Finke, 2007; Pew Research Center, 2016). The existing Christians and Jews in a few Middle Eastern countries, Muslims and Christians in Myanmar and Sri Lanka, suffer from explicit legal

and governmental restrictions in daily life. In the post-communist countries, people tend to have lower level of trust toward strangers and the state institution in general, and the same conclusion may be drawn about people's attitudes toward surveyors in countries such as China (Howard, 2003).

Even without explicit dangers or stigma, people may still distort their responses to religious questions due to the existence of the interviewer effect. Scholars have shown that people are more reluctant to disclose important information to strangers, and the differences in demographic background between interviewers and interviewees make people less likely to disclose authentic information (Liu & Stainback, 2013; Tourangeau & Yan, 2007). Fewer studies directly discussed the interviewer effects on answering religious questions, but it is highly plausible that many would choose to distort or withhold their genuine identities and behaviors if the question is deemed sensitive. The Coptic Christians in Egypt, for example, tend to withhold their religious activity information when interviewers were visibly Muslim (veiled or wearing Takiya)(Blaydes & Gillum, 2013). In the U.S., atheists receive lowest trust by the public and are least likely to genuinely disclose (non)religious preferences to others (Edgell, Gerteis, & Hartmann, 2006). Therefore, many self-reported prevalence of religious groups may have severely undervalued the actual scale of existence of the target populations.

Overreporting religious identity in self-reports

Under other conditions, people may intentionally exaggerate their participation in religious activities, and falsely claim to be a member of a preferred group. It is a long-standing phenomenon known among sociologists of religion that Americans over-report their religious attendance frequency. Using alternative counting method, Hadaway, Marler and Chaves pointed out that the real attendance rate is only half as claimed by survey respondents (C Kirk Hadaway et al., 1993, 1998). They have also found that the size of religious gatherings is often smaller than registered or claimed by the organizations (C. Kirk Hadaway & Marler, 2005). With an overwhelmingly moralistic cultural atmosphere and the foundational role of religion in organizing communal life and shaping neighborhood structures in the U.S, asking whether one

attends a church or synagogue can be tantamount to questioning his/her personal integrity or loyalty to an ethnic group. Under the influence of social desirability, many questionnaire items on religious identity or religious participation are essentially invalid measures of the actual level of belonging (Brenner, 2011; Presser & Stinson, 1996).

Although religion does not always convey a positive social desirability to uplift the real estimates of religious groups in other countries compared to the U.S., the problem nevertheless exists in other sectors of the population. Finke, Bader and Polson (2010) have pointed out that only systematic error will cause essential bias in modeling. The systematic error based on gender, age, or any other given covariate, distort the general picture of religiosity. For example, Chinese men are more likely to claim belief in ancestor spirits than women (Chen, 2014; Li & Lavelly, 2003), but they may have over-reported this because the patrilineal nature of ancestor worship demands their narrated masculinity. Similarly, if people from coastal India perceive a more favorable image of Christianity (Frykenberg, 2008), miss-reporting religious identity will be subject to a geographical factor, which is further confounded by socioeconomic development. Thus, a varying degree of social desirability in a special niche will create problems for estimating religious populations based on self-reports.

The (un)reliability of official census

The interwoven complex between ethnic identity and religious identity constructed by political and nationalist interests poses a significant problem for obtaining accurate religious identity information from official records. Official records such as census or municipal documents are the primary means of gathering a population's religious information in many countries that are cautious about the religious demography. However, official records are often detached from the reality. For instances, Lebanon has not collected such information since the last century due to the concern about electoral instability; China has never clearly publicized its methodology of the current general information on religions. Often times, official records of religious information rely on inaccurate and unidentifiable sources. In many former USSR countries including Russia, Kazakhstan, and China, the sizes of religious groups in official publications are often some crude

aggregates of the ethnic populaces that traditionally follow that particular religion, so that the Huis in China are automatically classified as Muslims despite sizable atheists, Christians and Jews among them, Russians in Kazakhstan are automatically counted as Orthodox Christians even many are atheists or have converted to Islam (Gladney, 1998; Peyrouse, 2004). Official records also fail to take into account of the non-registered religions or sects outside the oligopoly state-church system. For example, a variety of folk religions in China including Shamanism and animism have no recorded information about their sizes, despite millions of Chinese routinely resort to folk religious practices for ritual and spiritual treatment (F. Yang, 2006; F. Yang & Hu, 2012).

An alternative: the network scale up method

For the above reasons and based on both theoretical and practical concerns, estimating religious demography through conventional surveys suffers from some challenging biases. To make the matter worse, estimating hidden and semi-underground religious groups in hostile contexts does not just suffer from greater level of measurement biases, but also risks the potential violation of the respondents' personal safety. To solve the bias issues and protect respondents, scientists have lately revitalized an interest in the network-based survey methods. One methodology that realizes this advantageous survey strategy is the network scale-up method (NSUM). By gauging information that does not pertain to the respondents *per se*, but to other people known to them, NSUM can eliminate biases related to social desirability and disclosure concerns, it also saves researchers from the challenge of identifying covert populations in a random sampling.

Conventional statistical inferences draw conclusions about the population from the sample, a scale-down process. Instead, the NSUM first uses the population's information to infer about the sample, and then draw some conclusions back to the population. Specifically, the NSUM uses the information known in the population to estimate the information unknown from the sample by, typically, collecting surnames known to each respondent (Killworth, Johnsen, McCarty, Shelley, & Bernard, 1998) {Killworth, 1998 #1175} {Killworth, 1998 #1175}. Because hidden

populations, such as HIV infected persons, ex-convicts, gays, etc., cannot be directly and reliably reached through interviewing the respondents themselves due to negative social desirability and stigma, the network scale-up method shifts the attention to the people who may know these targets.

Furthermore, conventional surveys and censuses have particular difficulty in reaching out to small subpopulations. Scholars have noticed that many denominations were not included in population censuses, and substantial variability may emerge in analyses based in small areas (Clogg, Massagli, & Eliason, 1989; Finke et al., 2010). Stratified oversampling, a common strategy to sample small subpopulations, typically demands greater financial and human resources in implementation. But with NSUM, exponentially fewer sampling units are required to identify a small subpopulation. Imagining a population consisting of only four persons Alex, Ben, Carol, and Dawood, where Alex is connected to Ben, Ben connected to Carol, Carol connected to Dawood, and Dawood to Alex. Now we want to find Dawood. The probability of reaching Dawood in one simple random selection equals $\frac{1}{4}$. However, if the surveyor can obtain Dawood's information from either Alex or Carol, who are simultaneously connected to Dawood, one random selection has $\frac{1}{2}$ chance of identifying him. The more densely a social network is composed of, the easier for a simple random selection to find a member from any subpopulation through the reports of other people connected to this subpopulation, a phenomenon known as the six degree world (Watts & Strogatz, 1998).

The NSUM has been used by public health scientists and criminologists (Bernard et al., 2010; Salganik et al., 2011), but no one has ever employed this method to estimate relevant information that may interest scholars of religion. The method's utility of assessing hidden populations shows great potential for the scientific study of religion, where scholars are increasingly paying attention to the issues of self-reports. Perhaps even more importantly, the very social nature of religious identity, dated back to a Durkheimian conceptualization of religions as distinct signals perceptible to others (Durkheim, 1912 (2008)), requires some

measurements that do not assume an individualistic assumption but instead engage religious identity in social networks.

Given its advantages, NSUM is not necessarily a more complicated method than what is already based upon self-reports. The method does require additional questions in a survey other than simply soliciting a self-report, but these additional questions can be used for a range of other research purposes not restricted to religion. In the following sections, we will show the logic and formula of applying NSUM, and how to properly design and collect the information to be used for NSUM.

Methodology

The classic network scale-up estimates

{Killworth, 1998 #1175}The basic approach of the network scale-up method proposed by Killworth and colleagues (Kadushin, Killworth, Bernard, & Beveridge, 2006; Killworth, Johnsen, et al., 1998) is first to estimate a respondent's network size:

$$D_i = N * \sum \frac{n_{ik}}{N_k}, \quad (\text{eq. 1})$$

where D_i is i th respondent's network size (i.e. degree centrality), N is the size of total population (e.g. all people in Nashville), N_k is the total population size of the k th known subgroup (e.g. all "Tysha" in Nashville, whose size should be known from the Census or other sources), and n_{ik} refers to the number of k th known subgroup **known to** i th respondent (e.g. "how many 'Tysha' do you personally know in Nashville"). Then, since we have obtained the size of personal network D_i for each respondent in equation 1, we can simply estimate the total size of a hidden subgroup h (e.g. Shia Muslims) as the weighted proportion of that subgroup in all respondents' networks:

$$\hat{N}_h = N * \sum \frac{n_{ih}}{D_i}, \quad (\text{eq. 2})$$

where n_{ih} is the size of h th hidden subgroup **known to** i th respondent.

Implementing the classic NSUM in a survey

Now we will demonstrate a gentle guide on how to conduct NSUM, from survey design to estimation with software. The first fundamental step, as shown in equation 1, is to calculate the network degree of each individual respondent—namely how many people do you know (the meaning of knowing should be defined by researchers). The utility of this information is extremely profound and certainly not restricted to the application of NSUM. A network's degree centrality is found to be associated with a wide range of social outcomes, such as health wellbeing, upward mobility, sexual risk behaviors, and socioeconomic advantages (Bian, 1997; Burt, 2000; Pollard, Tucker, Green, Kennedy, & Go, 2010; Wasserman & Faust, 1994; X. Y. Yang, Kelly, & Yang, 2015). Conventional survey items often directly ask the question of network centrality in such manner as "How many people do you consider as acquaintance/friends/confidants". Direct estimates of network centrality based on such questions are not inherently problematic when used as a covariate. However, they can be systematically biased when the value of such measures *per se* is of interest, and there is limited way of correcting such biases without other information (Marsden, 2002; Watanabe, Olson, & Falci, 2016). NSUM on the other hand, is technically capable of ensuring the estimated network degrees are true (see the discussion of the hold-out method in Results section).

In NSUM, according to equation 1, we first solicit respondents for the number of people they know (n_{ik}) in several known categories (N_k). To give an extreme example where the number of known categories is only 1 ($k=1$): how many Tysha do you know in Nashville. If a respondent knows exactly two Tysha, and we also know from the Census that there are currently 2000 Tysha in the Nashville, we can infer that this respondent's social network size is 0.1% of the total Nashville population. Multiply this fraction by the total Nashville population $N=1,830,000$, we conclude that this respondent's network degree D_i is 1,830.

The value of reported number of people known to i th respondent (n_{ik}) is sensitive to researcher's definition of "knowing". We suggest that researchers use the same criteria and very close wording when describing what counts as a valid relationship. For example, a time frame

and a defined action should be deployed in the survey question in order to reduce the recall bias and arbitrariness, such as “during the **past 30 days**, how many Tysha **have you talked to** in Nashville”.

Another caveat is that these known populations chosen as the denominators (N_k) would better be distributed randomly in the entire population. Although corrective methods exist when the known populations show non-random mixing in the population (McCormick, Salganik, & Zheng, 2010), such methods require treating the known populations as some over-dispersed parameters in a binomial model, which easily frustrates practitioners in the field. Avoid non-random mixing is thought-provoking but feasible, for the only requirement is that our known population of choice should, assumedly, randomly distribute in our sampling frame. Surnames may be a good example: although first names in many cultures demonstrate a generational divergence, surnames do not suffer from this generational influence. For example, there are more “Gloria” among the Boomers than Millennials, but the surname “O’Brien” should not significantly decrease over the years. The advantage of using surnames as known population characteristics is most pronounced in mono-ethnic societies such as China, Japan, and Vietnam, where both generational and cultural influences on choosing a name disappear (e.g., surname “Hu” shows a nearly uniform distribution across genders, generations, and regions in China). Researchers may also adventurously use other characteristics as known populations, such as people who are left-handed, as long as such population is randomly distributed and we do know their actual size in the population.

With the network degree for each respondent (D_i) calculated, the size of unknown population \hat{N}_n is simply a weighted maximum likelihood estimation of the fraction of each respondent’s enumeration of the target population (n_{in}) divided by his/her network degree, as shown in equation 2. To be consistent, when asking for the number of the target population (n_{in}), the same wording should be used as in previous questions soliciting the sizes of other populations known to the respondent. For example, the same time frame and defined action should be used

as previously: “during the **past 30 days**, how many Shia Muslims **have you talked** to in Nashville”.

At first glance, NSUM seems to require a number of additional questions to be incorporated in a survey, and this may pester survey designers who are often concerned about the space and time constraints in questionnaire. But this is not necessarily the case. As we have discussed, five questions about the known populations can be sufficient to estimate network degree. More importantly, these questions may serve many other purposes than NSUM estimation. For instance, scholars may use “how many of your friends have got flu in the last 30 days” as a known predictor, assuming the occurrence of flu is random in a population. With known information about the total flu infections in an area in a given month (N_k), this question does not only serve as the k th known population predictor (n_{ik}) for network degree (D_i) in equation 1, it can be by itself a variable of interest for scholars studying health perceptions and infection risk.

Correcting the classic NSUM estimates

Although a representative and random sampling would mitigate the issue, the classic form of NSUM suffers from three biases. First, there is a bias from the barrier effect, in which situation people with certain characteristics are less likely to know a specific population. For an extreme example, sick and bed-ridden people are systematically barred from knowing many young cheerleaders. The second bias comes from the transmission effect, in which situation one does not realize that s/he actually knows someone from a specific population. For example, an average American may not realize he has befriended a Shia Muslim. Third, recall bias may occur when respondents have access to a specific population, clearly know their identity, but fail to remember it when filling a questionnaire. This is more a measurement error than modeling issue, and pertinent to a variety of survey questions other than NSUM (Coughlin, 1990; Graham, Catania, Brand, Duong, & Canchola, 2003).

To reduce recall bias, the known population groups used as denominators (N_k) should constitute 0.2-0.3% of the whole population, and best not to exceed 5% of the population in order to avoid the cognitive burden of recalling blunt and ambiguous targets, such as people whose first name

is “John” (McCormick et al., 2010). There is no consensus on the best number of known population groups, but researchers found that five known populations ($k=5$) would suffice and sometimes less information produces surprisingly better accuracy (D. M. Feehan, Umubyeyi, Mahy, Hladik, & Salganik, 2016). Also, as mentioned earlier, the definition of the network alters should be explicitly stated with a specified time frame and a clear description. If the questions are not clearly defined, researchers may alternatively record the response latency of answering a specific question and adjust for the difficulty of recall and evaluation (Mulligan, Grant, Mockabee, & Monson, 2003). With these caveats paid attention to and relevant strategies deployed during survey design, recall bias should be significantly reduced.

The barrier bias arises from non-random mixing between the target population and other sampled respondents. Consequently, the barrier bias is less problematic when the sampling satisfies the conditions of randomness and representativeness. If bed-ridden people are systematically barred from knowing cheerleaders, athletic people on the other hand are systematically overexposed to them. In a random and representative sampling, both bedridden and athletic people are thus self-weighted. However, when the target population is extremely small and secluded, random mixing between the target and the whole population is less likely to be observed.

Some scholars, particularly McCormick and others (Maltiel, Raftery, McCormick, & Baraff, 2015; McCormick et al., 2010), used a model-based approach to treat the biases as arising from an underlying binomial distribution, such that the barrier bias is subject to a probability process $q_{ik} \sim \text{beta}(m_k, \rho_k)$ and that the transmission bias is subject to a probability process $\tau_k \sim \text{beta}(\eta_k, \nu_k)$. Zheng, Salganik, and Gelman (2006) modeled the biases by overdispersed Poisson model with the similar model-based ideas. Behind these model-based approaches is the assumption that the two biases come from some underlying distributions of different characteristics in the population, and such distributions can be modeled.

Feehan and Salganik recently have developed another method for correcting the barrier and transmission biases, and this method may be more attractive to social scientists because it

emphasizes survey design rather than parameter modeling (D. M. Feehan & Salganik, 2016). They have topologically proven that the true estimate of an unknown population can be achieved via weighting the classic NSUM by three correction factors obtained from a probabilistic sampling among the unknown population. With the correctors, the generalized NSUM estimator equals:

$$\widehat{N}_h = \text{classic } \widehat{N}_h \times \frac{1}{\phi_f} \times \frac{1}{\delta_f} \times \frac{1}{\tau_f} \quad (\text{eq. 3})$$

$$\phi_f = \frac{\text{average number of connections from a member in sample } F \text{ to the rest of sample } F}{\text{average number of connections from a member in the population to the sample } F} \quad (\text{eq.3.1})$$

$$\delta_f = \frac{\text{average number of connections from a member in the unknown group to sample } F}{\text{average number of connections from a member in sample } F \text{ to the rest of sample } F} \quad (\text{eq.3.2})$$

$$\tau_f = \frac{\text{times that members in the unknown group being reported by the sample } F}{\text{total ties connecting members in the unknown group and the sample } F} \quad (\text{eq.3.3})$$

It is further demonstrated that when the sample frame F is a random and representative selection of the total population, ϕ_f can be eliminated. Thus, researchers only need to correct for: δ_f , the different ratio of network degrees between the unknown group and the whole sample, and this refers to the barrier bias; τ_f , the true positive rate of reports, and this refers to the transmission bias. Using a concrete example where our target unknown population is again Shia Muslim in Nashville. If on average Shia Muslims have 50 non-Shia friends, non-Shia people have 30 non-Shia friends, δ_f equals 1.67 and Shia Muslims are considered proactive and well-exposed to the public. If on average Shia Muslims' religious identity are known to 40 out of their 50 friends, τ_f equals 0.8. The classic NSUM estimate of Shia Muslims should be weighted by a combined factor of 0.75 ($=1/1.67 \times 1/0.8$). We do not need additional information to obtain δ_f , as long as the target unknown population can be included in our sample with a relative probability. To obtain the true positive rate τ_f , we need to evaluate among the Shiites what fraction of their non-Shia connections actually know their identity (i.e. times being reported by sample F divided by their total connections in sample F). This requires researchers to add a question such as "of all my coworkers/classmates/neighbors who are not Muslim, how many percentage really know my

religious identity?” This measurement itself, may be of great interest for the study of religious salience.

Example and application

Background

We will now demonstrate how NSUM can be applied in the field of religion with an example from Chinese international students, in which we are interested in the sizes of two religious groups—Buddhists and Christians. There is currently no information about the religious composition of this population, whose immigration experience and the affiliation with an officially atheist party-state poses an interesting enough case for scholars of religion. A survey of Chinese international students at a large public university in the Midwest of the United States was collected in 2016. The target institute is among the top three American universities with the largest number of Chinese students, who have enrolled in all major disciplines including engineering, sciences, agriculture, and social sciences. The sample ($n=767$) is closely representative to the population, and the small remaining discrepancy is adjusted by post-stratification weights based on college, graduate status, and gender (Lohr, 2010).

Known populations were designated as Chinese surnames: Zhu, Wu, Liu, Sun, Yang, Zhou, and Zheng. These surnames were very carefully chosen to be representative in a wide range of areas of China in order to avoid the concentration or underrepresentation of certain regional-specific surnames (Zheng et al., 2006). According to the Sixth National Census of China (Wu & Yang, 2014), the number of people carrying these surnames exists on a similar degree across northern, southern, coastal, and western China. Population sizes of these surnames were obtained from the Registrar at the surveyed university.

We asked the respondents to enumerate the number of Chinese people they “have talked to face-to-face in the past 2-3 months at this University”—the criteria to define acquaintance—from the following categories: people with surnames as Zhu, Wu, Liu, Sun, Yang,

Zhou, Zheng; Chinese who are Christians; Chinese who are Buddhists; Chinese who cheated in an exam; Chinese who have dated or married to a non-Chinese; Chinese who own a luxury car. All other groups than the surnames can be considered as unknown populations whose real prevalence is to be estimated. The enumerations of the seven surnames were capped at a maximum of 20 people in order to avoid careless responses and exaggeration.

Adjustment factor for transmission bias (the true positive rate τ_f) is the average report of the question “what percentage of your good friends/people in your major clearly know your religious identity” among a religion's believers. We used good friends and people in your own major as categories of reference here. Another category of reference that may be used is the exact same one used to enumerate network connections, in this study, the Chinese people you “have talked to face-to-face in the past 2-3 months at this University”. A game of contact method is also suggested by scholars, which asks respondents from the target group to assess their visibility to several social groups whom they would have equal chance to contact as everyone else, such as mailpersons (Bernard et al., 2010; Salganik et al., 2011).

As for how did we define who are a religion's believers, the survey asked for self-reported religious beliefs in five major religions (Buddhism, Daoism, Protestantism, Catholicism, Chinese folk religions) on a four-point scale ranging from “don't believe at all”, “don't believe much”, to “somewhat believe” and “totally believe”. A believer of a religion is someone who chose “totally believe” for that religion.

Although manual calculation is feasible, we utilized an R package “networkreporting” for convenient implementation of NSUM (D. Feehan & Salganik, 2014). Operating R codes for all procedures will be available on the first author's website upon formal publication.

Results

By enlisting the seven surnames as their acquaintances, respondents' network degrees can be estimated as the fraction of the number of surname carriers known to them over the surname carriers out in the population (see equation 1). To be cautious, how do we know the estimated

network degrees are accurate? If the calculated network degrees are reliable, the ratio of known surnames in the **sample** to the **sample's** average network size should equal the ratio of known surnames in the **population** to the **entire** population size. To test this, a “hold-out” method was proposed (Killworth, McCarty, Bernard, Shelley, & Johnsen, 1998) to examine whether a **known** population's size can be correctly estimated by other known populations. This method first takes out the k th known population, and uses the rest of the $k-1$ populations to derive D_i . It then estimates the k th known population as if it is an unknown \hat{N}_k , and compare the estimated \hat{N}_k with the real N_k . Repeat this process through all known populations, we will be able to find out which k th population has not been estimated correctly.

Table 1 shows the estimated sizes of all seven surnames in comparison to the real sizes of the surnames. Each surname is estimated as unknown at a time. It turns out that the estimated “Liu” negatively deviates from its real size. People systematically underreport “Liu” in their networks, very likely due to the bias of recalling a larger target population. “Liu” constitutes 5.6% of the total population, which clearly exceeds the standard recommended by earlier studies (McCormick et al., 2010; Zheng et al., 2006). As a consequence, it will drag down network degree estimates D_i , and inflate the relative proportion of other surnames in respondents' social networks. As shown in the right panel of table 1 and visualized in figure 1, the estimated sizes of surnames are much closer to their real sizes after removing “Liu”.

Finally, figure 2 presents the distribution of network degrees D_i as estimated by the remaining six surnames. The average size of social networks is 44 people, with a standard deviation of 80. The smallest network has 5 people, and the largest has 746 people. The distribution of network degrees in our sample closely approximates what would be determined by the power law, to which the distributions of friendship networks in real life often adhere (Watts & Strogatz, 1998). This distribution, in addition to the hold-out method's confirmation, leads us to trust the estimates of network degrees D_i as reliable and realistic.

With reliable D_i values estimated, the sizes of hidden populations can be estimated in NSUM's classic form per equation 2, and in NSUM's generalized form after adjusting the barrier and

transmission biases per equation 3. We will first elaborate how correctors of the barrier and transmission biases are obtained.

Recall equation 3, barrier bias δ_f is the fraction of the unknown group's average network size consisting of outsiders over the average network size among outsiders themselves. If this fraction is larger than unity, it implies that the unknown group is more socially proactive, its members may have acquainted a good number of outsiders and these outsiders would report more people from this unknown group in a survey, that is, less barrier effect. If this fraction is smaller than unity, the unknown group is more secluded, and thus greater barrier effect. In table 2, we listed the average network sizes between religious believers and non-believers, as well as the network size among non-believers. On average, a Buddhist's social network consists of 40.8 people, and 1.6 among them are also Buddhists, which yields an average network size of 39.3 between Buddhists and non-Buddhists. Meanwhile, the average network size among non-Buddhists is 35.5. The δ_f is 1.105 for the Buddhists in this sample. On the other hand, although Christians have larger social networks (42.6), there are more coreligionists in their networks (6.2), resulting in fewer connections between themselves and non-believers (36.4). Accordingly, the δ_f is 1.047 for the Christians in this sample. Again, with the representativeness assumption of the sampling, we can assume that the ϕ_f from equation 3 is negligible and δ_f should suffice as an expression of the barrier effect.

Transmission bias τ_f is assessed by the question "what percentage of the following people clearly know your religious identity". We used two categories of reference: good friends, and people in your major. The average transmission rate (visibility) is .36 for Buddhists, and .52 for Christians. Chinese Christians are more visible to their network alters in comparison to Buddhists.

The classic NSUM estimates of Buddhists and Christians in the population and their respective percentages are presented in the second row of table 3. We have applied 300 bootstrap resampling to derive unbiased means. The distributions of the 300 estimates generated by bootstrap resampling are presented in figure 3. According to the classic NSUM, there are 71

Buddhists in this population (1.7%), and 219 Christians (5.3%). In the next row are the estimates weighted by respondents' poststratification weights. The overrepresented demographic groups in the sample were slightly more likely to report higher numbers of both Buddhists and Christians, leading the post-weighted sizes of Buddhists and Christians to be slightly smaller. About 1.7% of the population are Buddhists, and 4.7% are Christians. Finally, after adjusting for δ_f and τ_f per equation 3, the estimates in fifth row are 176.1 for Buddhists (4.3% of the population) and 253.8 for Christians (8.6% of the population). In comparison, rate of self-reported "totally believe" is 1.2% for Buddhism and 4.7 for Christianity. In other words, using the self-reported method, there would be 1.2 percent Buddhists and 4.7 percent Christians among the Chinese international students in the United States in 2016. However, using generalized network scale-up estimation, we find that among this Chinese population there were 4.3 percent Buddhists and 8.6 percent Christians.

Concluding Remarks

One of the persistent issues that pesters the social scientific study of religions for many decades is the question of how to appropriately measure and identify the size of religious believers in a society. It has long been recognized by scholars that the self-reported religious attendance rate in the United States is unreliable and must have been inflated by some underlying mechanisms. Comparing attendance count and the number of religious congregations in America, Hadaway and Marler (2005) found fewer than 22 percent of the Americans weekly attend services, although more than 40 percent claimed so (C Kirk Hadaway et al., 1998). The similar overreporting behavior exists in many Muslim countries, and is associated with the importance of religious identity (Brenner, 2011). On the other hand, neither do subjective religious beliefs make a good indicator of being a religious believer. The unaffiliated people, a demographic group on the rise in the U.S., often report conventional religious beliefs not dissimilar to those hold by the church-going populace (Hout & Fischer, 2002).

There seems to be a very strong preference for appearing as religious among Americans, hence their self-reports on religious attendance and identity are inflated by the social desirability bias. But the contrary trend may exist in other cultural contexts, such as the countries that hoard hostility against certain religions or all religions. In Egypt, Christians underreport their religiosity and adherence to Christianity when the interviewer is visibly Muslim (Blaydes & Gillum, 2013). As a country that elevates the religious issue as a national security concern, China's official religious demographics are questioned by many scholars (Potter, 2003; F. Yang, 2006), but the survey-based estimates are also suspected to suffer from social stigma underreports.

To counter the problems associated with self-reported estimates, primarily that caused by biased reporting behaviors, we have introduced the network scale-up method (NSUM) to estimate the size of religious groups. The NSUM has several advantages over traditional survey-based self-reports. First, this methodology shifts the burden of information from the target group to their network connections, thus effectively reducing the required sample size to identify the members of the target group. Second, by shifting the burden of information to network alters, social stigma and social desirability is a non-issue when it comes to estimating the size of the persecuted/promoted religions. Third, because the method is an unbiased maximum likelihood estimation, its results are more stable and less susceptible to idiosyncratic reporting behaviors of the respondents. Fourth, the estimated network sizes can be examined by comparing the estimates with the true population sizes, aka the hold-out method, which was demonstrated in the result section. Doing so, users of NSUM can be certain that at least the network sizes D_i are unbiased. Finally, researchers may compare the estimated sizes of the target religious groups using different adjustment factors, and test the sensitivity of their results.

In this article, we have also demonstrated the utility NSUM with a practical example. We have offered a step-wise guide to estimate the sizes of Buddhists and Christians in a sample of Chinese international students. We hope this could facilitate interested scholars to implement

NSUM in their own studies. We conclude that 4.3% of the population are Buddhists and 8.6% are Christians.

The contribution of NSUM is not only its technical reliability, because several researchers have come up with other innovative methods of gauging accurate information on religious behaviors, such as asking the respondent's time-use allocation, comparing known congregation sizes, etc. (Brenner, 2011; C. Kirk Hadaway & Marler, 2005). More importantly, NSUM answers to the question of the very nature of religious identity: what makes a person a believer of X religion? The growing population of "the spiritual but not religious", privatized belief systems epitomized by "Sheilaism", the revival of non-institutional religions, all challenge the assumption that religious identity can be assessed by some direct and unambiguous questions in self-reports and that a respondent's recognition of own religious identity equates his/her actual position in the religious world. Lim, MacGregor and Putnam (2010) showed that many liminal religious-nones sporadically reported religious affiliation when their overall religious involvements changed little. A non-attending person may claim membership in a religious organization, but does the authority of that religion acknowledge him? Or in another case, a non-attending Catholic who has never admitted being Christian, but the Church still count him in the herds since he was infant-baptized, is he really Catholic? Estimates of religious identity based on self-reports may not suffice to offer a satisfactory answer.

A hundred years ago, Durkheim clearly stated that "Religious representations are collective representations which express collective realities. They should be social affairs and the product of collective thought" (Durkheim, 1912 (2008), p. 9). For Durkheim and largely the sociology of religion, there may be religions with no concept of god and only an amorphous belief system, but all religions have to exist as a collective of distinct identities in social interactions. A religion creates important functions of social solidarity by imposing recognizable marks to its believers, so that the believers are distinguished from outsiders, and vice versa. Beliefs, attendance, religiosity, are but parts of this collective identity. A previous method by Mockabee, Monson, and Grant (2001) on religious commitment shows that how religious identity was socially

defined influences the salience of individual religious behaviors. After weighted by the normative value placed on different activities by a respondent's religious tradition, they found little variation in religious commitment across different denominations.

The NSUM we introduced in this study is a methodological realization of this sociological perspective on the essence of religion: the priority is given to the question how are religious believers perceived by their network alters, rather than how do religious believers define themselves. Although we do not intend to exaggerate the conceptual superiority of the NSUM over self-report, our method may be less vulnerable to the instability of religious identity among believers of non-institutional religions and "unaffiliated" believers. Because as long as a person's religious identity is a sufficiently perceptible signal to others, his/her existence can be modeled with adjustments. For example, Lim, MacGregor and Putnam's (2010) "liminal religious nones" would show no unstable religious identity in a network-based approach because these people's chance of being reported by network alters wouldn't change much if their overall religious involvements change little. We also suggest that NSUM and self-reports can be used in combination rather than seen as conflictual, because each of them represents a conceptually different definition of religious identity. While NSUM emphasizes the social definition of religious identity, self-reports may tell us about religious believers' self-awareness. Unlike behaviors and events, religious identity is sometimes a multi-dimensional social construct, and there is hardly any true value for it. We can at best assume there is a religious group whose real size is unknown but estimable. Based on this assumption, NSUM offers an easy and technically reliable alternative to self-reports.

References:

- Bernard, H. R., Hallett, T., Iovita, A., Johnsen, E. C., Lyerla, R., McCarty, C., . . . Stroup, D. F. (2010). Counting hard-to-count populations: the network scale-up method for public health. *Sexually Transmitted Infections*, 86(Suppl 2), ii11-ii15. doi:10.1136/sti.2010.044446
- Bian, Y. (1997). Bringing strong ties back in: Indirect ties, network bridges, and job searches in China. *American Sociological Review*, 366-385.

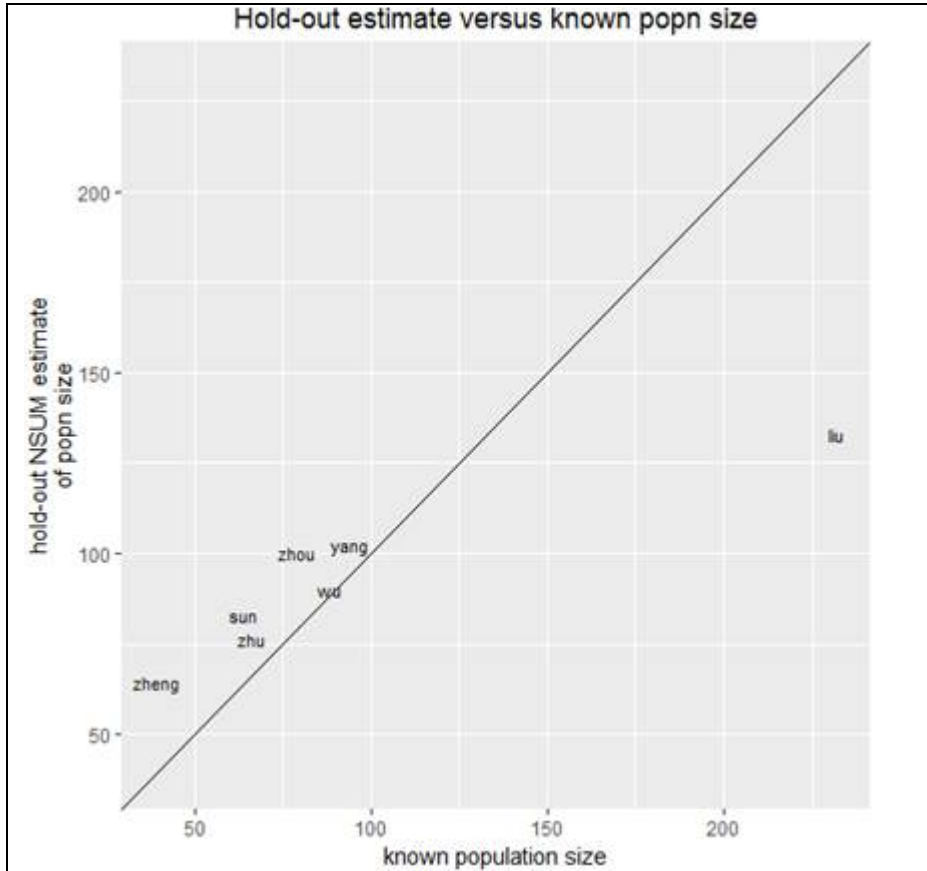
- Blaydes, L., & Gillum, R. M. (2013). Religiosity-of-Interviewer Effects: Assessing the Impact of Veiled Enumerators on Survey Response in Egypt. *Politics and Religion*, 6(03), 459-482. doi:doi:10.1017/S1755048312000557
- Brenner, P. S. (2011). Identity importance and the overreporting of religious service attendance: Multiple imputation of religious attendance using the American Time Use Study and the General Social Survey. *Journal for the Scientific Study of Religion*, 50(1), 103-115.
- Brenner, P. S. (2014). Testing the Veracity of Self-Reported Religious Practice in the Muslim World. *Social Forces*, 92(3), 1009-1037. doi:10.1093/sf/sot120
- Burt, R. S. (2000). The Network Structure Of Social Capital. *Research in Organizational Behavior*, 22(0), 345-423. doi:[http://dx.doi.org/10.1016/S0191-3085\(00\)22009-1](http://dx.doi.org/10.1016/S0191-3085(00)22009-1)
- Chen, J. (2014). 警惕 “祭祖热” 对父权文化的 “传承”. *山西师大学报: 社会科学版*, 41(2), 130-133.
- Clogg, C. C., Massagli, M. P., & Eliason, S. R. (1989). Population undercount and social science research. *Social Indicators Research*, 21(6), 559-598. doi:10.1007/bf02217994
- Coughlin, S. S. (1990). Recall bias in epidemiologic studies. *Journal of Clinical Epidemiology*, 43(1), 87-91. doi:[http://dx.doi.org/10.1016/0895-4356\(90\)90060-3](http://dx.doi.org/10.1016/0895-4356(90)90060-3)
- Durkheim, E. (1912 (2008)). *The elementary forms of the religious life*. London: Dover Publications.
- Edgell, P., Gerteis, J., & Hartmann, D. (2006). Atheists As “Other”: Moral Boundaries and Cultural Membership in American Society. *American Sociological Review*, 71(2), 211-234. doi:10.1177/000312240607100203
- Feehan, D., & Salganik, M. (Producer). (2014). The networkreporting package. *R Foundation for Statistical Computing*. Retrieved from <http://cran.r-project.org/package=networkreporting>
- Feehan, D. M., & Salganik, M. J. (2016). Generalizing the Network Scale-up Method. *sociological methodology*, 46(1), 153-186. doi:doi:10.1177/0081175016665425
- Feehan, D. M., Umubyeyi, A., Mahy, M., Hladik, W., & Salganik, M. J. (2016). Quantity Versus Quality: A Survey Experiment to Improve the Network Scale-up Method. *American Journal of Epidemiology*, 183(8), 747-757. doi:10.1093/aje/kwv287
- Finke, R., Bader, C. D., & Polson, E. C. (2010). *Faithful measures: Developing improved measures of religion*: Association of Religion Data Archives.
- Frykenberg, R. E. (2008). *Christianity in India: from beginnings to the present*. New York: Oxford University Press.
- Gladney, D. C. (1998). *Ethnic identity in China: the making of a Muslim minority nationality*: Harcourt Brace College Publishers Fort Worth, TX.
- Graham, C. A., Catania, J. A., Brand, R., Duong, T., & Canchola, J. A. (2003). Recalling sexual behavior: A methodological analysis of memory recall bias via interview using the diary as the gold standard. *The Journal of Sex Research*, 40(4), 325-332. doi:10.1080/00224490209552198
- Grim, B. J., & Finke, R. (2007). Religious Persecution in Cross-National Context: Clashing Civilizations or Regulated Religious Economies? *American Sociological Review*, 72(4), 633-658. doi:10.1177/000312240707200407
- Hadaway, C. K., & Marler, P. L. (2005). How Many Americans Attend Worship Each Week? An Alternative Approach to Measurement. *Journal for the Scientific Study of Religion*, 44(3), 307-322. doi:10.1111/j.1468-5906.2005.00288.x
- Hadaway, C. K., Marler, P. L., & Chaves, M. (1993). What the polls don't show: A closer look at US church attendance. *American Sociological Review*, 741-752.
- Hadaway, C. K., Marler, P. L., & Chaves, M. (1998). Overreporting church attendance in America: Evidence that demands the same verdict. *American Sociological Review*, 63(1), 122-130.
- Hout, M., & Fischer, C. S. (2002). Why More Americans Have No Religious Preference: Politics and Generations. *American Sociological Review*, 67(2), 165-190. doi:10.2307/3088891
- Howard, M. M. (2003). *The weakness of civil society in post-communist Europe*. Cambridge: Cambridge University Press.

- Kadushin, C., Killworth, P. D., Bernard, H. R., & Beveridge, A. A. (2006). Scale-up methods as applied to estimates of heroin use. *Journal of Drug Issues*, 36(2), 417-440.
- Killworth, P. D., Johnsen, E. C., McCarty, C., Shelley, G. A., & Bernard, H. R. (1998). A social network approach to estimating seroprevalence in the United States. *Social Networks*, 20(1), 23-50.
- Killworth, P. D., McCarty, C., Bernard, H. R., Shelley, G. A., & Johnsen, E. C. (1998). Estimation of seroprevalence, rape, and homelessness in the United States using a social network approach. *Evaluation Review*, 22(2), 289-308.
- Li, J., & Lavelly, W. (2003). Village Context, Women's Status, and Son Preference Among Rural Chinese Women*. *Rural Sociology*, 68(1), 87-106. doi:10.1111/j.1549-0831.2003.tb00130.x
- Lim, C., MacGregor, C. A., & Putnam, R. D. (2010). Secular and Liminal: Discovering Heterogeneity Among Religious Nones. *Journal for the Scientific Study of Religion*, 49(4), 596-618. doi:10.1111/j.1468-5906.2010.01533.x
- Liu, M., & Stainback, K. (2013). Interviewer gender effects on survey responses to marriage-related questions. *Public Opinion Quarterly*, nft019.
- Lohr, S. (2010). *Sampling: Design and Analysis* (2nd Ed.). Boston, MA: Brooks/Cole, Cengage Learning.
- Maltiel, R., Raftery, A. E., McCormick, T. H., & Baraff, A. J. (2015). Estimating population size using the network scale up method. *The annals of applied statistics*, 9(3), 1247.
- Marsden, P. V. (2002). Egocentric and sociocentric measures of network centrality. *Social Networks*, 24(4), 407-422. doi:[http://dx.doi.org/10.1016/S0378-8733\(02\)00016-3](http://dx.doi.org/10.1016/S0378-8733(02)00016-3)
- McCormick, T. H., Salganik, M. J., & Zheng, T. (2010). How Many People Do You Know?: Efficiently Estimating Personal Network Size. *Journal of the American Statistical Association*, 105(489), 59-70. doi:10.1198/jasa.2009.ap08518
- Mockabee, S. T., Monson, J. Q., & Grant, J. T. (2001). Measuring Religious Commitment among Catholics and Protestants: A New Approach. *Journal for the Scientific Study of Religion*, 40(4), 675-690.
- Mulligan, K., Grant, J. T., Mockabee, S. T., & Monson, J. Q. (2003). Response Latency Methodology for Survey Research: Measurement and Modeling Strategies. *Political Analysis*, 11(3), 289-301.
- Pew Research Center. (2016). *Trends in Global Restrictions on Religion*. Washington, DC: Pew Research Center.
- Peyrouse, S. (2004). Christianity and nationality in Soviet and post - Soviet Central Asia: mutual intrusions and instrumentalizations. *Nationalities Papers*, 32(3), 651-674.
- Pollard, M. S., Tucker, J. S., Green, H. D., Kennedy, D., & Go, M.-H. (2010). Friendship networks and trajectories of adolescent tobacco use. *Addictive Behaviors*, 35(7), 678-685.
- Potter, P. B. (2003). Belief in Control: Regulation of Religion in China. *The China Quarterly*, 174, 317-337. doi:10.1017/S0009443903000202
- Presser, S., & Stinson, L. (1996). *Estimating the bias in survey reports of religious attendance*. Paper presented at the Proceedings of the Section on Survey Research Methods. Alexandria, VA: American Statistical Association.
- Salganik, M. J., Fazito, D., Bertoni, N., Abdo, A. H., Mello, M. B., & Bastos, F. I. (2011). Assessing Network Scale-up Estimates for Groups Most at Risk of HIV/AIDS: Evidence From a Multiple-Method Study of Heavy Drug Users in Curitiba, Brazil. *American Journal of Epidemiology*, 174(10), 1190-1196. doi:10.1093/aje/kwr246
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological bulletin*, 133(5), 859.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge & Melbourne: Cambridge university press.
- Watanabe, M., Olson, K., & Falci, C. (2016). Social isolation, survey nonresponse, and nonresponse bias: An empirical evaluation using social network data within an organization. *Social Science Research*, 63. doi:<http://dx.doi.org/10.1016/j.ssresearch.2016.09.005>
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *nature*, 393(6684), 440-442.

- Wu, J., & Yang, J. (2014). 张, 王, 李, 赵谁最多---2010 年人口普查姓氏结构和分布特点. *中国统计*(6), 21-22.
- Wuthnow, R. (2015). *Inventing American Religion: Polls, Surveys, and the Tenuous Quest for a Nation's Faith*. New York: Oxford University Press.
- Yang, F. (2006). The red, black, and gray markets of religion in China. *The Sociological Quarterly*, 47(1), 93-122.
- Yang, F., & Hu, A. (2012). Mapping Chinese folk religion in mainland China and Taiwan. *Journal for the Scientific Study of Religion*, 51(3), 505-521.
- Yang, X. Y., Kelly, B. C., & Yang, T. (2015). Together we have fun: native-place networks and sexual risk behaviours among Chinese male rural-urban migrants. *Sociology of Health & Illness*, 38(4), 559-575. doi:10.1111/1467-9566.12380
- Zheng, T., Salganik, M. J., & Gelman, A. (2006). How Many People Do You Know in Prison? *Journal of the American Statistical Association*, 101(474), 409-423. doi:10.1198/016214505000001168

Tables and figures

Figure 1. Validity check of the known population estimate using the “hold-out” method, before vs. after.



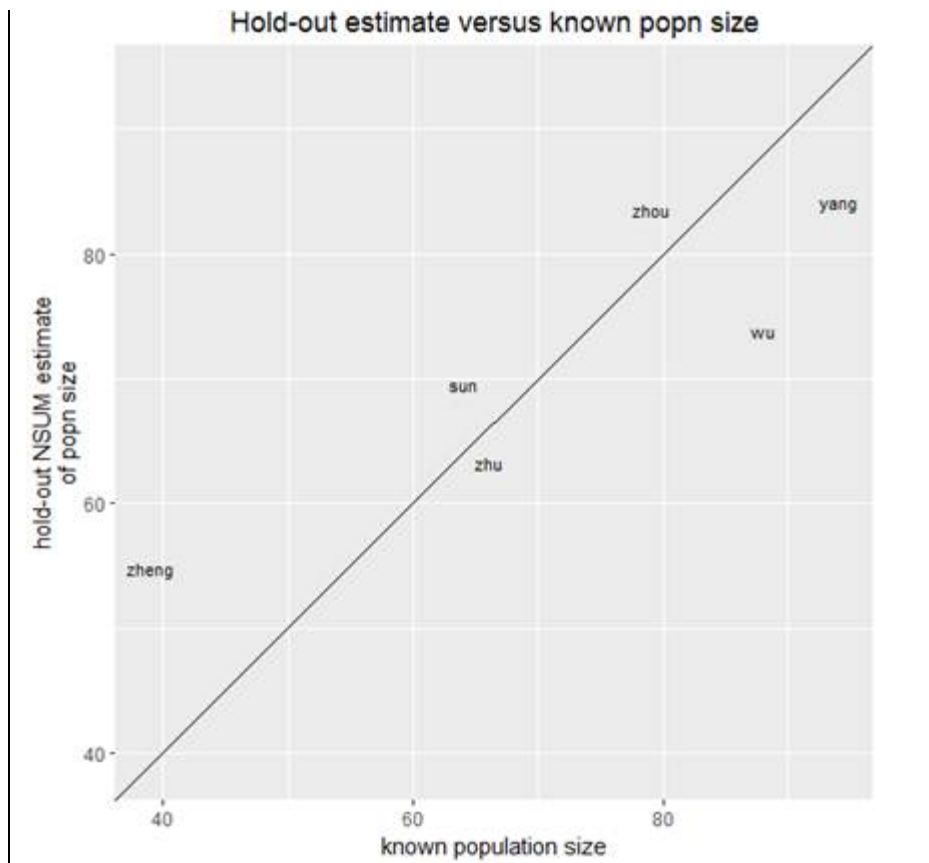


Figure 2. Distribution of network degrees from equation 1. Minimum=5, Max=746.3, Mean=43.6, s.d.=79.9

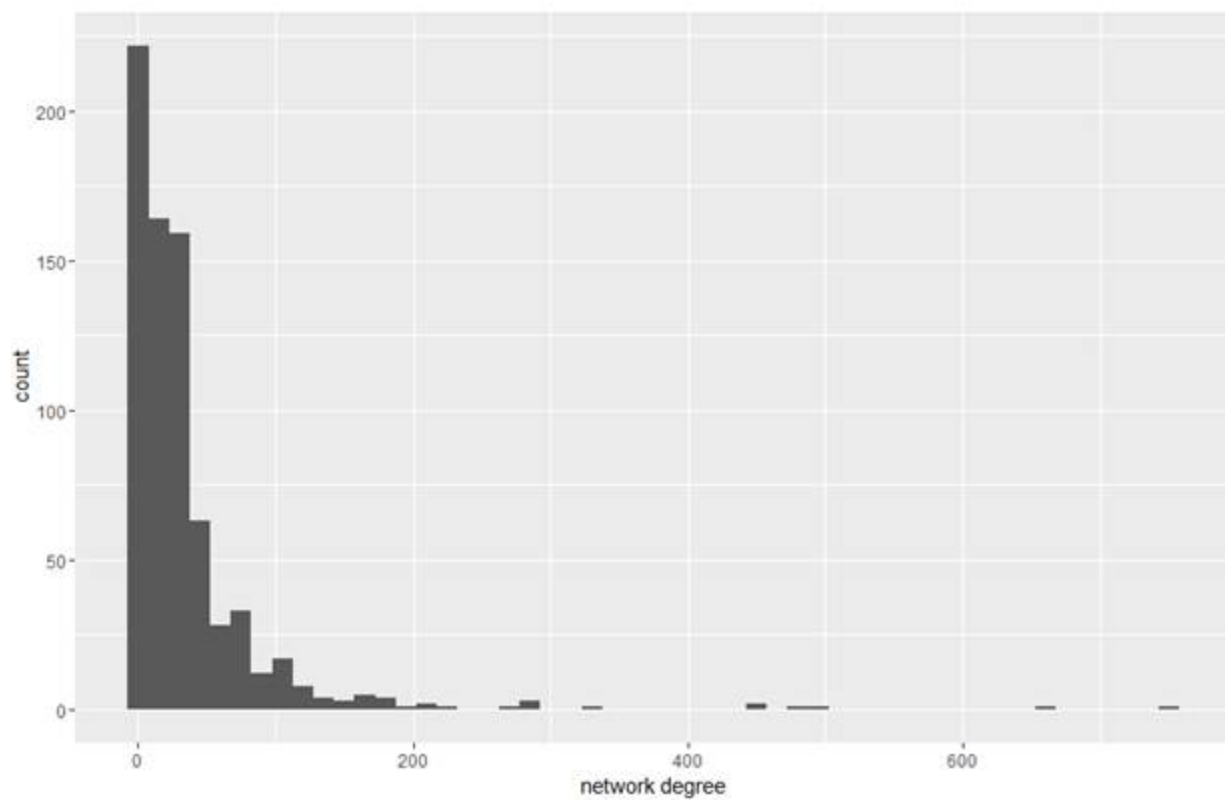
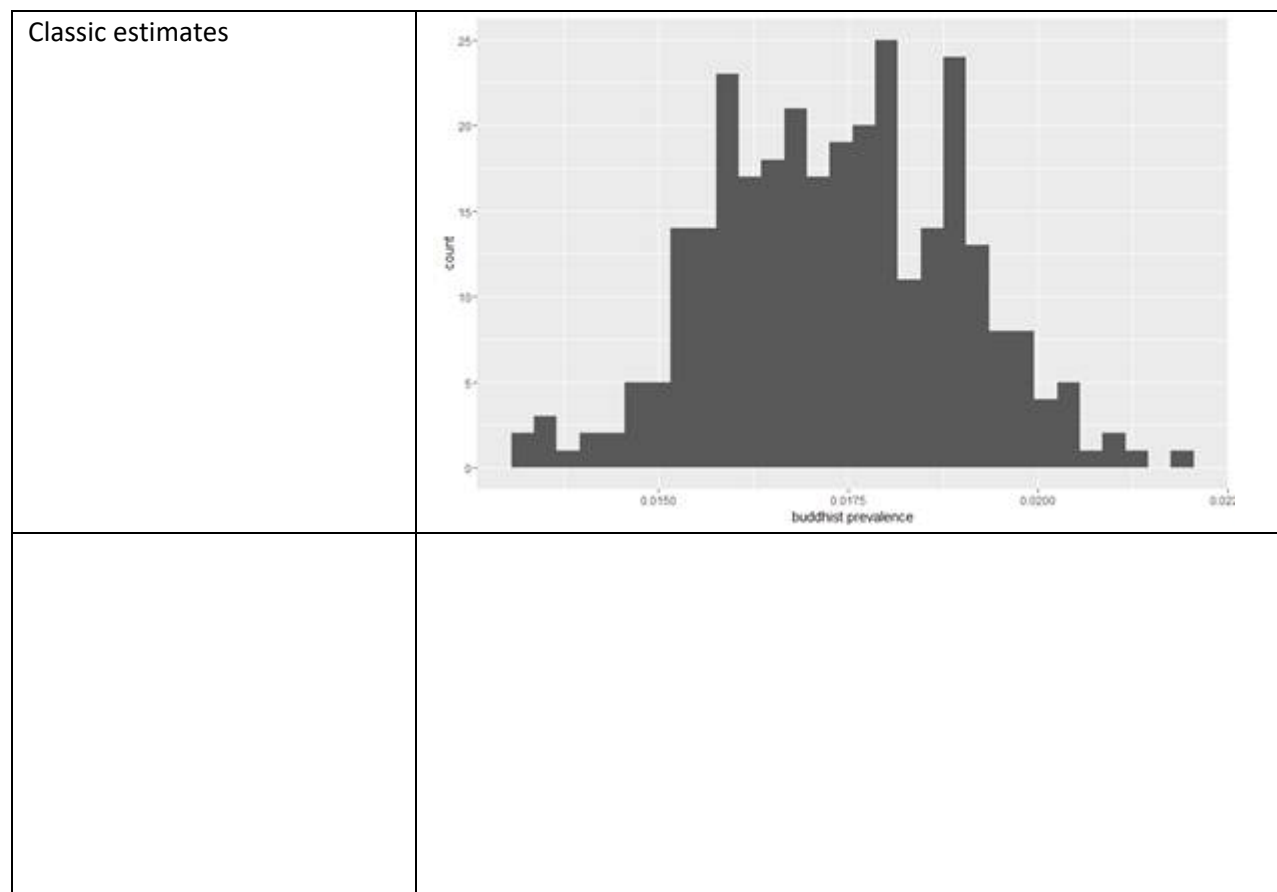
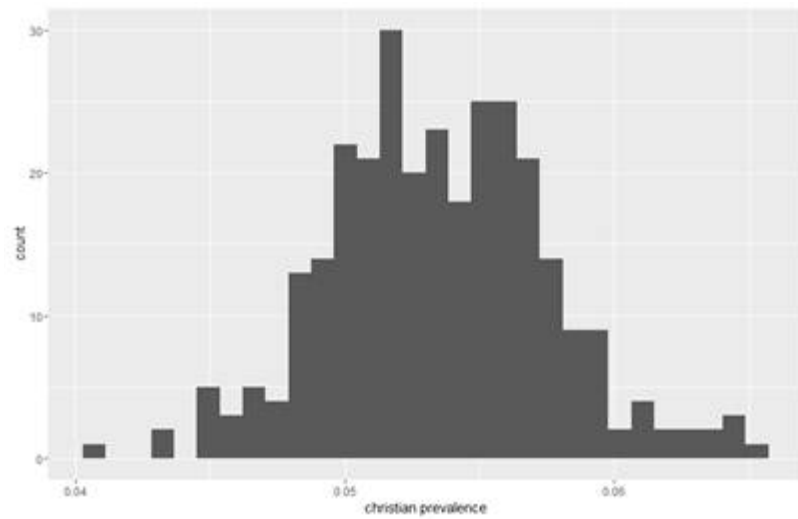


Figure 3. Selected distributions of NSM estimates with 300 bootstrap resampling





Estimate with sampling weights and adjustment for transmission bias

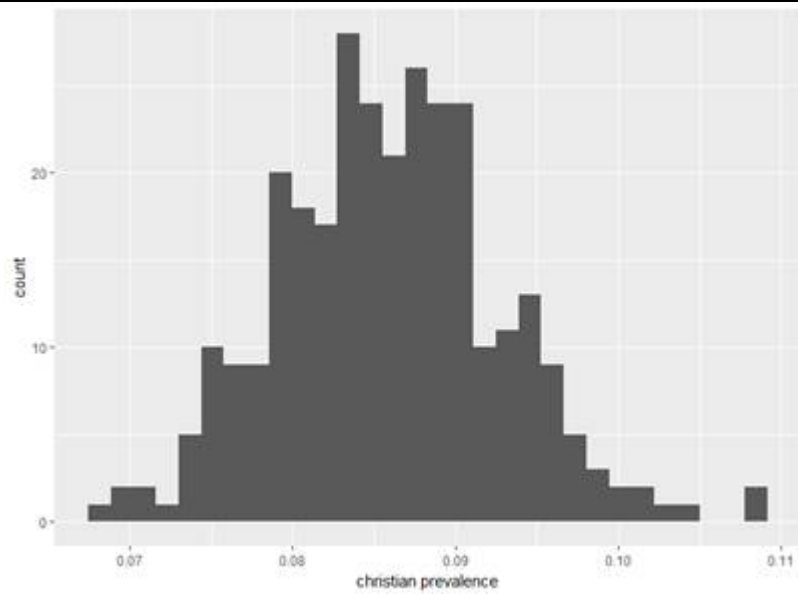
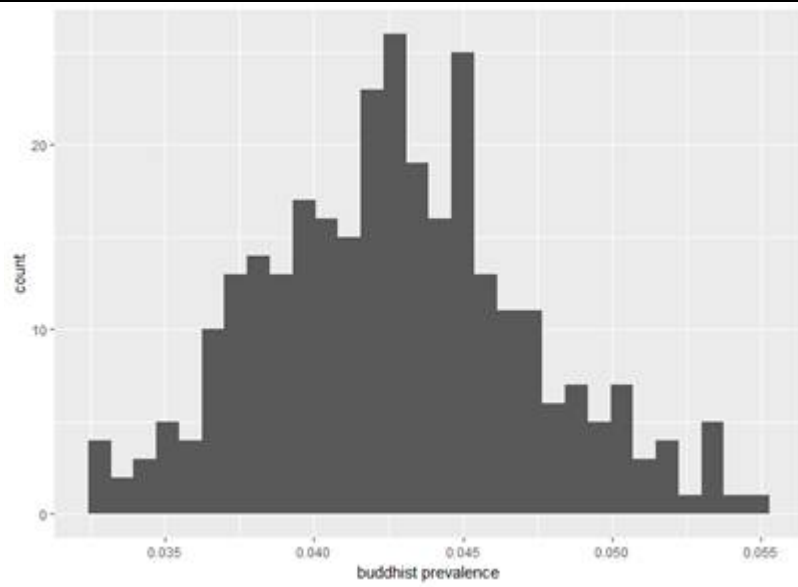


Table 1. Accuracy of the estimate of surnames using the “hold-out” method

Known Population types	Before		After taking out inaccurate predictor “Liu”	
	Estimated vs. known size	Killworth S.E. ^[1]	Estimated vs. known size	Killworth S.E.
Zhu	76:66	3.41	63:66	2.82
Wu	90:88	3.68	73:88	3.02
Liu	132:232	4.08	-	-
Sun	82:64	3.58	69:64	3.01
Yang	102:94	3.94	84:94	3.25
Zhou	100:79	3.93	83:79	3.29
Zheng	64:39	3.18	54:39	2.71

Table 2. Descriptions for barrier bias δ_f and transmission bias τ_f in reporting Buddhist and Christian acquaintances.

	Buddhists	Christians
Average # connections between believers and the whole sample	40.84	42.56
Average # connections between a believer and believers	1.58	6.15
Average # connections between believers and non-believers	39.26 (40.84-1.58)	36.41 (42.56-6.15)
Average # connections between non-believers and the whole sample	35.97	36.29
Average # connection between a non-believers and believers	0.44	1.54
Average # connection among non-believers	35.53 (35.94-0.44)	34.75 (36.29-1.54)
δ_f	1.105 (39.26/35.53)	1.047 (36.41/34.75)
τ_f (among good friends)	.37	.61
τ_f (in the same major)	.36	.45
τ_f (average)	.36	.52

Table 3. Classic and generalized network scale-up estimates of Buddhist and Christian Chinese in the target population, mean and standard deviation estimated with 300 bootstrap resampling

	Buddhists (% in population)		Christians (% in population)	
	Mean	s.d.	mean	s.d.
Classic estimate	71 (1.7%)	6.6 (.2%)	219 (5.3%)	17 (.4%)
Estimate with sampling weights	70 (1.7%)	7.5 (.2%)	194 (4.7%)	16 (.4%)
	63.9 (1.6%)	7.0 (.2%)	183 (4.4%)	15 (.4%)

Estimate with sampling weights and adjustment for δ_f				
Estimate with sampling weights and adjustment for τ_f	195.5 (4.7%)	20.6 (.5%)	372.8 (9.1%)	31.09 (.8%)
Estimate with sampling weights and adjustment for δ_f, τ_f	176.1 (4.3%)	18 (.4%)	253.8 (8.6%)	27.9 (.7%)
Self-reported from the sample (more believing)	.094 (9.4%)	.011 (1.1%)	.092 (9.2%)	.011(1.1%)
Self-reported from the sample (totally believing)	.012 (1.2%)	<u>.047 (4.7%)</u>	.008(.8%)	

Killworth, P. D., McCarty, C., Bernard, H. R., Shelley, G. A., & Johnsen, E. C. (1998). Estimation of Seroprevalence, Rape, and Homelessness in the United States Using a Social Network Approach. *Evaluation Review*, 22(2), 289-308. doi:10.1177/0193841x9802200205

[1] $\sqrt{N \frac{\delta_i}{\sum_j n_j}}$ (Killworth, McCarty, Bernard, Shelley, & Johnsen, 1998, p. 294).