

Facebook sampling methods: some methodological proposals

Ester Macri¹, Cristiano Tessitore²

¹University of Florence, e-mail: estermacri@gmail.com

²ISTAT - University of Florence, e-mail: tessitore@istat.it

Abstract

During the last decade, new social realities, like the social networks, urged social research methodology to reconsider many techniques in order to face the new reality. As a consequence, a new kind of survey has been developed, exploiting the Net's power and involving internet users: the "e-survey". One of the main challenges of e-surveys is represented by the sampling procedure: this one needs to be reconsidered in order to avoid the risk of being biased and the lack of scientific accountability. Our work will concentrate on Facebook, one among the most famous social networks.

In the first part, the paper provides a review of the major available sampling techniques, by highlighting and underlining their strengths and weaknesses, especially in Facebook perspective. Afterwards, the paper illustrates some proposal of Facebook sampling techniques.

First of all, Facebook sampling techniques can be divided into two groups:

- 1) techniques aimed at sampling the whole Facebook population
- 2) techniques aimed at sampling known groups (Facebook subpopulation).

In our opinion, one of the most problematic issues on Facebook surveys concerns the contact strategy that should be adopted in order to maximize the response rate. In particular, the difficulty arises in the first contact strategy: different strategies can be identified. The respondent can be contacted through:

- an individual researcher's profile, created ad - hoc;
- an impersonal profile regarding the Research Institute or the survey title.

Despite all, we would like to do a very important remark. We can use a Facebook sampling method when the respondent population and the target Facebook population are quite the same.

Keywords: [Facebook, Sampling, E-survey]

1. The context

The global diffusion of Internet involves economic, political, cultural and geographical factors and it is interesting for sociologists and policy makers. In the last few years, big changes in internet usage have occurred. In particular, during the last decade, new social realities, like the social networks, urged social research methodology to reconsider many techniques in order to face the new reality.

As a consequence, a new kind of survey has been developed, exploiting the Net's power and involving Internet users: the "e-survey".

This happened also with reference to sampling methods, which can rely on new techniques.

As well-known, social networks developed from the point of view of functionalities and reached a vast population of consumers. The world wide web is full of this kind of sites, like Twitter, Netlog, Myspace, but one of the most famous social networks is Facebook.

The use of the Facebook platform improves the potentialities of e-surveys, because it allows researchers to access and more easily sample individuals from a reference population. Facebook has now [Facebook statistics, 2010] more than 500 million users, and the 50% of them are active [an active user is a person who logs on to Facebook at least once a day]. The Facebook average user has 130 friends, is connected to 80 community pages, groups and/or events, creates 90 pieces of content each month. People spend over 700 billion minutes per month on Facebook and more than 30 billion pieces of content (web links, news stories, blog posts, notes, photo albums, etc.) are shared each month. According to the web statistics site "Experian Hitwise" [Experian Hitwise Statistics, 2011], Facebook ranks at third place on 2009 and at ninth place in 2008 and it became the most popular site on the Internet this year, jumping to the top spot in Internet searches (10,4% of unique visitors, against 7,4% of unique visitors of Google, 3,09% of Yahoo Mail and 3,04% of Youtube).

One of the main challenges of e-surveys is represented by the sampling procedure, which needs to be reconsidered in order to avoid the risk of being biased and the lack of scientific accountability. Are the classic sample methods, like random sampling, systematic sampling, stratified sampling, quota sampling or snowball sampling effective to investigate new web realities?

The key question of this paper is about one of the critical issues regarding samples drawn from Facebook: their representativeness. Biases can arise because of difficulties in defining the reference population. Can Facebook users be representative of internet users?

For example, examination of US data from comScore shows that Facebook reported 112 million unique visitors in December 2009. Given that the estimated active US Internet population consists of 205 million, that means that 54% of all Internet users visited Facebook in December [Prescott, 2010].

2. The inefficiency of traditional sampling methods in a Facebook perspective

The most common form of probability sampling is Random sampling; with this kind of sampling method each member of the population has a chance (not equal to zero) of being selected. In the Simple Random Sampling each individual has the same probability of being chosen at any stage during the sampling process. The Simple Random Sampling is highly representative if all subjects participate, but it is not applicable without complete list of population members and it is a method potentially uneconomical to achieve. In a Facebook perspective, this kind of Sampling Technique cannot be used because there isn't a complete list of the population. Only in the case of researches aimed at sampling

the whole Facebook population, the researcher can adopt a similar technique, called “random walk sampling technique” (par. 3).

Another common method is the Stratified Sampling. It is a commonly used probability method that is superior to random sampling because it reduces sampling error. A stratum is a subset of the population that shares at least one common characteristic. First of all, the researcher identifies the relevant strata and their actual representation in the population of interest. Random sampling is then used to select a sufficient number of subjects from each stratum. The principal advantage of this kind of sampling method is that it can ensure that specific groups are represented, even proportionally, in the sample, by selecting individuals from strata list. On the other hand, the disadvantages lie in the fact that it is a very complex method, it requires greater effort than the simple random sampling and strata must be carefully defined. This sampling technique is very complex in a Facebook perspective, because identifying the strata could be very difficult.

The Quota sampling is a non-probabilistic method, consisting in selecting individuals as they come to fill a quota by characteristics proportional to the population. The principal advantage of this kind of sampling method is that it ensures the selection of an adequate number of subjects with appropriate characteristics. By the way, a disadvantage of Quota Sampling is that it is not possible to prove that the sample is representative of a designated population. In Facebook, the researcher can select individuals following a quota scheme but he will never know the real distributions of the Facebook users.

Another non-probabilistic method is named “snowball sampling”: it is used when the desired sample characteristic is rare and it may be extremely difficult or prohibitively costly to locate respondents in these situations. In this kind of sampling method, subjects will refer the researcher to people they know who have the same characteristics. An advantage is that it is possible to include members of groups where no lists or identifiable clusters even exist (e.g., drug abusers, criminals). A disadvantage is that there is no way of knowing whether the sample is representative of the population. Snowball sampling is applicable in Facebook but, since there is not so much interaction on the web, it is difficult that the respondent refers you to another respondent.

Convenience sampling is another nonprobability method; it consists in asking for volunteers to participate in the survey. It is a cheap way of ensuring sufficient numbers of a study, but it can be highly unrepresentative.

Snowball sampling and convenience sampling should be adopted in a Facebook survey, with the recommendations exposed in the next paragraph.

3. Facebook sampling techniques

Facebook sampling techniques can be distinguished in two groups:

- 1) techniques aimed at sampling the whole Facebook population
- 2) techniques aimed at sampling known groups (Facebook subpopulation).

If the research project aims at surveying attitudes and opinions of Facebook users, the techniques belonging to the first group can be applied (e.g., an explorative survey about the time spent reading e-zines).

The techniques belonging to the second group can be used in exploring specific groups of people showing a common interest (e.g.: a survey concerning online role play gamers).

These people can be identified looking at their subscription to Facebook fan pages or groups.

When the research is focused on the whole population, a random walk sampling technique on Facebook profiles can be used, providing an unbiased random sample [Gjoka et al, 2010].

One of the techniques proposed by Gjoka et al. (2010) consists in modeling the Facebook social graph as an undirected graph $G = (V, E)$, where V is a set of nodes (users) and E is a set of edges (mutual friendship relationships). They suggest some methods based on random walk. These consist, basically, in choosing an user to start from and then randomly select a friend of him/her. The friend enters the sample and he/she becomes the starting point for the next step. The procedure ends when the needed quantity of subjects is reached.

Instead, when the research is focused on subgroups, groups and fanpages belonging to the research field (not forgetting to translate searches in the main languages) should be identified.

Once the sampling technique has been chosen according to the population of interest, another problem lies in trying to maximize the response rate by urging users to answer the survey. In the following table (which, however, is not exhaustive) we suggest some different contact strategies, according to the applied survey approach.

Population	Contact type	Strategy
Whole facebook population	User by user	Personal Message Without Friendship Request (PMWOFR)
		Personal Message With Friendship Request (PMWFR)
Known Group	User by user	Personal Message Without Friendship Request (PMWOFR)
		Personal Message With Friendship Request (PMWFR)
	Massive	Group Email Message (GEM)
		Fanpage Status Change (FSC)
		Wall Message in [Groups and Fanpages] (WMG – WMF)
		2 stage: Friendship and request (2S)

Table 1 – Different contact strategies

Even though the table is not exhaustive, the proposed classification tries to supply a review of the principal approaches allowing the sample selection and it illustrates the available strategies aiming at minimizing the nonresponse rate.

One of the main issues on Facebook surveys dwells in the contact strategy to be adopted in order to maximize the response rate. In particular, the problem arises in the choice of the first contact strategy. Different strategies can be identified: the respondent can be contacted through:

- an individual researcher's profile, created ad-hoc;
- an impersonal profile regarding the Research Institute or the survey title.

The latter strategy can appear more professional, giving a sense of science to the participation request.

A part from the adopted strategy, we suppose to provide the contacted person with a link to an external e-survey platform.

If the survey concerns the whole Facebook population, two different contact strategies can be identified:

- a personal message without friendship request (PMWOFR strategy);
- a personal message with friendship request (PMWFR strategy).

In the first case, PMWOFR strategy, the contacted user is selected randomly and he/she is asked to participate in the survey through a personal message. The message illustrates the survey's aim and reports the link to the questionnaire.

In the latter case, PMWFR strategy, the user is contacted through a personal message and the procedure will go on only if the friendship request is accepted. Very likely, the respondents selected through this strategy are those more prone to participate in the e-survey.

If the e-survey is focused on a known group, a massive contact approach can be adopted (GEM strategy, FSC strategy, WMG – WMF strategy, 2S strategy).

In the case of GEM strategy, at first the Facebook groups and fanpages concerning the investigated topic should be identified. This can be done through a simple research: the researcher can use the Facebook search bar inserting words related to the investigated topic.

Subsequently, the administrators of these pages can be contacted and asked to promote the survey with a personal message to all group members. If this step is successfully accomplished, the researcher can quickly reach a huge amount of people. Otherwise, this kind of contact might be problematic in the following cases:

- 1) Multi admin groups: sometimes a group is administrated by several people; this situation might cause troubles and represent an obstacle for the research because the participation request at the investigation will be discussed by a large number of persons.
- 2) For profit groups: some groups are created for profit and have commercial finality. In this case the administrators might not be collaborative because they already send email to the subscribers but they aim at making profit, not research.
- 3) Groups without admin: a group may lack of administrators, because they leaved the platform or they log in it very rarely. It might be an obstacle for the research because the request of participation in the investigation could remain unread for very long time.

Another strategy, FSC strategy, consists in:

- creating a fanpage
- or
- requesting permission to an administrator

in order to promote the survey through the page status. This kind of strategy presents a lot of critical situations. In the first case, it could require a lot of time to reach a sufficient number of fans; in the latter case, there might not be a way to know the identity of the fanpage administrators, so it could be impossible to contact them.

In both cases, the main trouble is that the page status remains on the top of the home page only for few minutes. Because of this problem, the less active users couldn't read the status. This kind of strategy may penalize the less active users and could exclude them from the research.

The 2S strategy consists in creating an ad hoc profile sharing some interests with potential respondents. When the profile achieves a sufficient number of friends, the survey can be promoted by publishing it on the page's wall status. Even though this kind of strategy is time-consuming, it turns out to be effective to maximize the survey response rate.

The WMG – WMF strategy consists in publishing a link on the fanpages or groups' wall not directly controlled by the researchers. This technique has not a long-lasting visibility and can be marked as spam by the group administrators.

Like FSC, these strategies penalize less active users, excluding them 'de facto' from the research.

Finally, if the population is stratified according to the characteristics of interest, contacting single persons can be managed in the same way as described for the sampling procedure on Facebook total population (PMWORF and PMWRF strategies).

4. Conclusions

In the end, we can highlight some points:

- 1) Facebook is an emergent web phenomenon. In particular, the surpass of Google concerning the number of unique visitors makes day by day Facebook population more similar to the global internet population.
- 2) Despite all, the researcher can use Facebook sampling methods only when the respondent population and the target Facebook population are quite the same.
- 3) The future (and the present too) of Social Survey is the E-Survey. It's necessary to investigate about Facebook from a methodological point of view to obtain high quality samples to use in E-Surveys.
- 4) The traditional sampling method are not appropriate for Facebook sampling (par. 2). It's necessary to find new solutions concerning Facebook sampling, like the ones we have shown in this work.
- 5) It is necessary to experiment those techniques on the field to collect empirical evidences about the real usefulness of the proposed instruments.

References

Experian Hitwise Statistics (2011), <http://www.hitwise.com/us/datacenter/main/dashboard-10133.html>

Facebook statistics (2010), <http://facebook.com/press/info.php?statistics>
Gjoka et al (2010), Walking in Facebook: A Case Study of Unbiased Sampling of OSNs
Prescott L. (2010), Social Networking by the Numbers [http://www.typepad.com/
services/trackback/6a010534a19f97970c0120a880c80d970b](http://www.typepad.com/services/trackback/6a010534a19f97970c0120a880c80d970b).