

WEB-BASED RDS: DINAMICS AND QUALITY

“WEB SURVEY METHODOLOGY” RESEARCH GROUP

Stanislav Moiseev
Yuliya Savinkova

INTRODUCTION

RDS for researching closed groups or online-communities is based on the presumption that if researcher is following basic assumptions network sample wouldn't be bias (1, 2). However, critics insist that these assumptions are non-realistic, and impugn quality of the research conducted with RDS (3). Another complexity is there is no information about the consequence of personal acquaintance of seeds with researcher. It's unknown if loyalty to the researcher or the project may influence final sample.

RESEARCH QUESTIONS

Q1 Does final sample comply with the RDS assumption about random selection of the recruiter's peers; reciprocal relationships between the respondent and members of target population; and ability of the respondent to accurately report their degree?

Q2 What is an influence of personal acquaintance on sample quality?

METHOD & DATA

Two waves of research were conducted. During the first one 6 seeds were chosen from people who just wanted to participate for reward or their own interest and had no connection to researchers. During the second wave 5 seeds were people among researchers peers. Personal acquaintance was another stimulating factor on a par with reward.

To collect data for RDS we used online-questionnaire. To construct online-questionnaire the software Unipark was utilized.

NEED FOR PERSONAL INFORMATION

Respondents' personal data was required to construct network. They were asked to provide their and recruiter's name, their and recruiter's e-mail and mobile number. The necessity of information was explained as required to receive the reward for participating in the research.

Visit our web site



References

1. Heckathorn D. Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling // M. Salganik, D. Heckathorn // Sociological Methodology, Vol.34: 193-239, 2004.
2. Heckathorn D. Web-Based Network Sampling Efficiency and Efficacy of Respondent-Driven Sampling for Online Research // D. Heckathorn, C. Wejnert // Sociological Methods & Research, 2008, #37: 105-134, first published on June 10, 2008.
3. Goel S. Assessing Respondent-Driven Sampling // S. Goel, M. J. Salganik, A. Raftery // Proceedings of the National Academy of Sciences of the United States of America, Vol. 107 #15: 6743-6747, 2010.

RESULTS

R1 Both final samples did correspond to 2 Heckathorn assumptions: reciprocal relationships between the respondent and members of target population; ability of the respondent to accurately report their degree. However, neither of our samples did meet the third requirement about random selection of respondents to the sample.

R2 Sample quality was measured by two parameters: (1) representation of statistical population and (2) network growth dynamics.

Controlled Variables		Statistical Population	RDS 1	RDS2
SEX	Male	0,46	0,32	0,22
	Female	0,54	0,69	0,78
EMPLOYMENT	Employed	0,34	0,35	0,13
	Unemployed	0,67	0,65	0,87

(1) The method didn't lead to random selection in our case, so both of the samples were not representing statistical population. However, the first sample was much more closer to statistical data then the second one.

(2) Regardless the fact the first sample was better at representation, the second one appeared to be much more effective in growth dynamics. The same amount of respondents was reached during 3 month plus reminders at first case and only one month without reminders at the second case. We suppose that personal acquaintance don't need any additional force to recruit respondents.

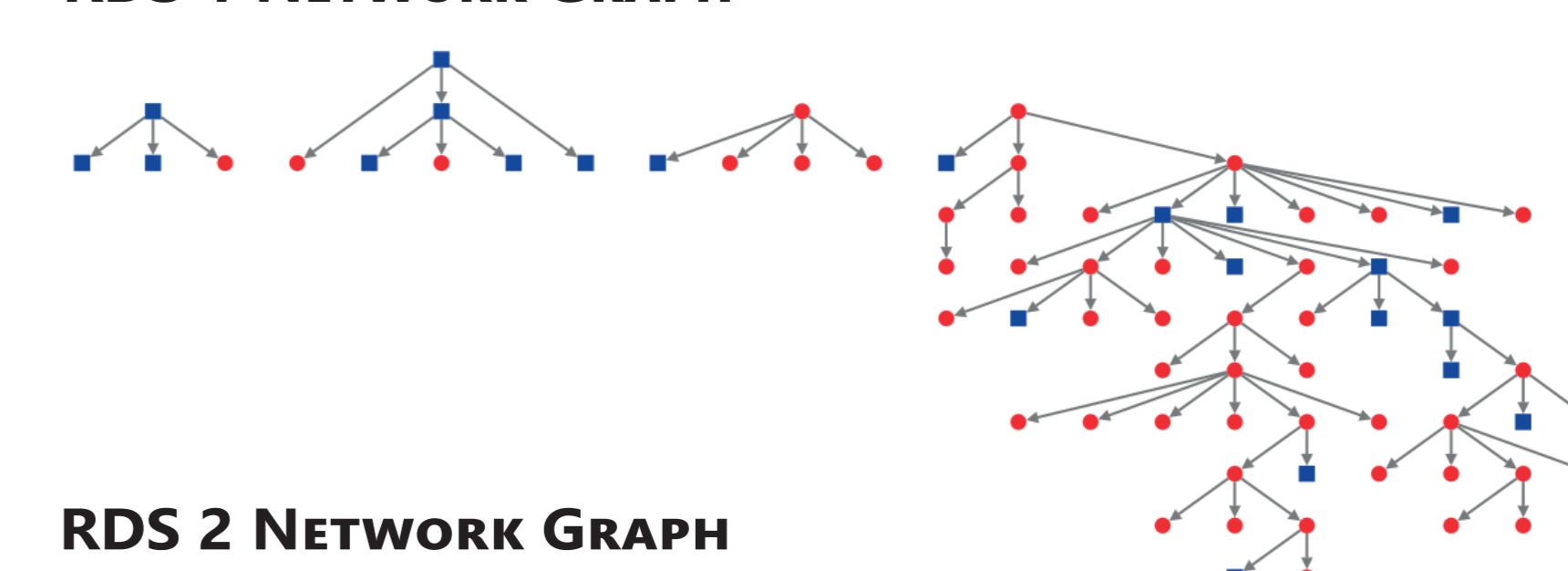
Graph Metrics	RDS Graph 1	RDS Graph 2
Vertices	72	79
Maximum Vertices in a Connected Component	56	56
Maximum Geodesic Distance (Diameter)	13	12
Average Geodesic Distance	5,285183	5,814577

DISCUSSIONS

R2 (1) RDS showed itself ineffective while used on students population in Russia. As a possible reason we can discuss absence of close relationships between students and hardness for them to communicate with other faculties which are located in different buildings across Moscow. Probably, some other groups can be represented better.

R2 (2) Both of our samples had super-seeds, and both of them brought a chain of 56 people to final sample. The most interesting part is that longest chain in RDS 1 has 13 edges and in RDS 2 12 edges. Is it a limit for network growth with used models of stimulation or just a coincidence?

RDS 1 NETWORK GRAPH



RDS 2 NETWORK GRAPH

