

# Network Size: Measurement and Errors in Respondent-Driven Sampling

JPSM/MPSDS Seminar

Ai Rene Ong, Yibo Wang

March 8, 2023



INSTITUTE FOR SOCIAL RESEARCH  
PROGRAM IN SURVEY AND DATA SCIENCE  
UNIVERSITY OF MICHIGAN

# Outline

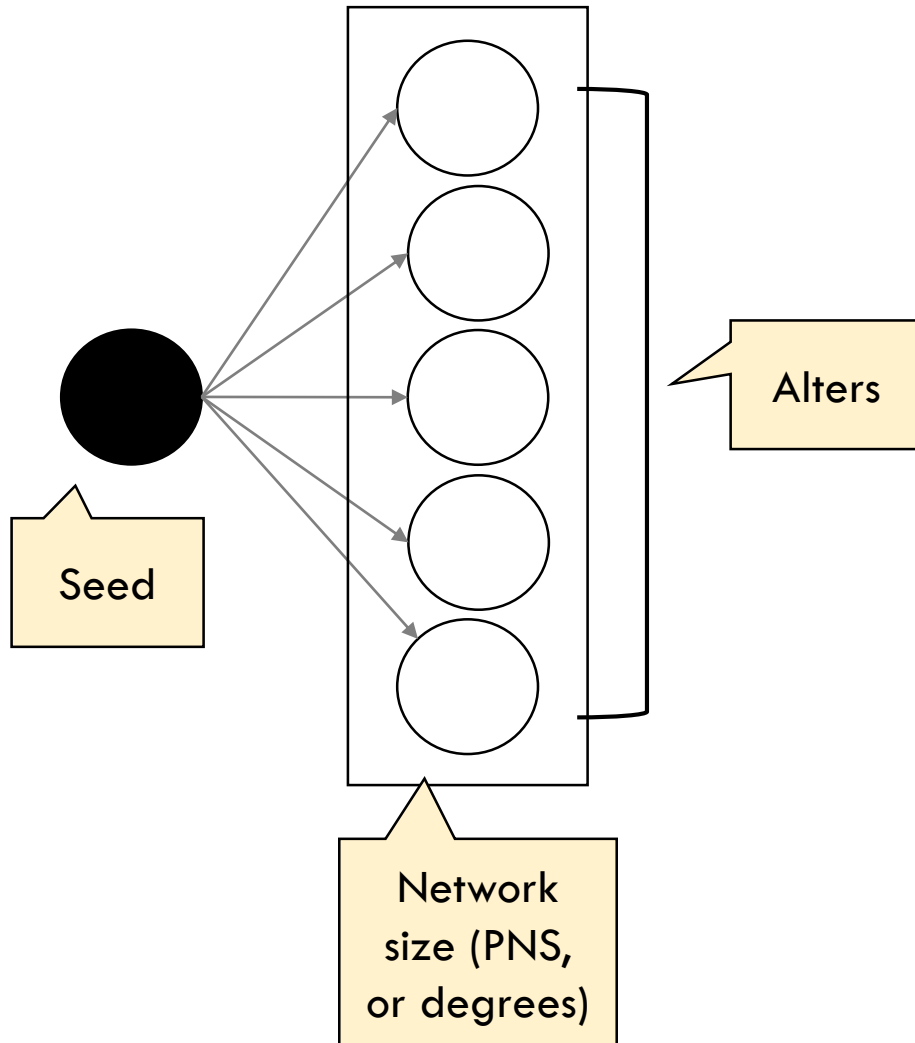
- Overview
- Study I: Measuring Degree in Respondent Driven Sampling
- Study II: A Latent Variable Model for Individual Degree Estimation in Respondent-Driven Sampling

# Overview

# Background

- Respondent driven sampling (RDS) is commonly used to sample hard-to-reach populations, e.g., people who inject drugs, sex workers
- RDS leverages respondents' social networks to reach the target population
- RDS uses self-reported personal network size (also known as egocentric network size, degrees) to adjust for the probability of being invited

# RDS terminology



Example of a question measuring degree/network size:

"How many transgender people do you know, who knows you and who live in this city?"

# Issues with degree measurement (1)

- Degree question is not standardized across RDS studies
  - Some uses a single question, some uses a sequence of questions
  - Many uses the word “knowing”, measuring the network of people *known* to the respondent

*“How many MSM do you know personally by name and how many MSM know you personally by name?” – Aung et al. (2013)*

*“How many transgender and MSM do you know, who know you and who live in this city?  
Of these, how many are 18 years of age or older?*

*Of those, how many have you seen or spoken to in the past 30 days?” – Paz-Bailey et al. (2013)*

## Issues with degree measurement (2)

- Test-retest reliability ranged from low to very high ( $r=0.17$  to  $r=0.98$ ) (Gile, Johnston & Salganik 2015; Yamanis et al., 2013)
- Large egocentric network size leads to more reporting error (Iguchi, Ober & Berry, 2009)
- Degree question in RDS needs to measure the *recruitable network*

- Study I: Measuring degree in Respondent Driven Sampling
  - Presents qualitative and quantitative findings in an attempt to understand how respondents report their degree
- Study II: A Latent Variable Model for Individual Degree Estimation in Respondent-Driven Sampling
  - Presents a proposal on a new method to estimate individual degree



# Study 1: Measuring degree in Respondent Driven Sampling

# Research questions

- What do different degree questions (“knowing someone”, “being close to someone”) mean to the respondents?
- Which degree questions are the most useful for RDS?
  - i.e., reliable, and measures the network of interest which is the *recruitable network*

# Data: In-depth interview

- Eligibility:  $\geq 18$  years old, identify as LGBTQ, White or Black, reside in the United States,
  - Advertised in University of Michigan LGBT groups listserv; Michigan Institute for Clinical & Health Research participants portal
- Interviewed  $n = 19$  respondents in 1-hour Zoom interviews
- Topics of interview: definition of “know”, “being close to”; the alters they will consider inviting to an RDS survey; motivation for cooperating with recruitment request

# Data: Web-RDS

- Eligibility:  $\geq 18$  years old, identify as LGBT, reside in the United States
  - Seeds recruited from the Michigan Institute for Clinical & Health Research participants portal; Max. recruits is two
- 10-15 minutes survey about the people in the respondents' LGBT networks
  - Respondents asked to describe up to 4 alters
- n seeds = 68; n respondents = 394; n recruiters = 185
- Follow-up survey two weeks later for recruiters (n=166)
- Recruitment restricted to two waves

# Data: Degree questions asked in web-RDS (1)

- **PNS (all):** How many [TARGET POPULATION] do you know, who knows you and who live in the United States?
- **PNS (18+):** Of these [PNS (all)], how many are 18 years of age or older? If you are not sure, just count them as adults.
- **PNS (close):** Now I am going to ask you questions about the [PNS (18+)] people you know that are [TARGET POPULATION]. How many do you feel close to (that is, you feel at ease with, can talk to about what is on your mind, or call on for help)?
- **PNS (interact):** Of those [PNS (18+)] people you know who are [TARGET POPULATION], how many do you interact with (including talking to, visiting with, calling, emailing, texting, Facebook, etc.) personally more than once week?

# Data: Degree questions asked in web-RDS (2)

Scenario:

*"LGBT Health Study is an online survey which is 15-20 minutes long. This survey asks questions about various topics such as physical and mental health and healthcare access. Participants are compensated a \$10 gift card for their time upon completing the survey."*

- **PNS (interest):** Of those [PNS (18+)] people you know who are [TARGET POPULATION], how many how many do you think will be interested in participating in this study?
- **PNS (invite):** Of those [PNS (18+)] people you know who are [TARGET POPULATION], how many do you think you can invite individually to participate in this study?

# Data: Questions about alters

- Age (*18-29; 30-39; 40-49; 50-59; 60-69; 70+*)
- Race (*White; Black; Asian; NHPI; Other; Don't know*)
- Alter's Hispanicity (*Yes; No; Don't know*)
- Relationship (*family; partner; friend; others*)
  - Influence over alter (*Yes; No*)
- Length known in years
- Closeness (*1: Not at all close, 7: Very close*)
- Likelihood of responding to invitation (*1: Least likely, 5: Most likely*)

# Analytic methods

- Thematic analysis of the transcripts from the in-depth interviews (knowing; being close to someone; likely recruits)
- Descriptive analysis of relationship with alters in the web-RDS
- Consistency of the PNS questions assessed with test-retest reliability (Spearman's rho) using main web-RDS survey and follow-up survey data
  - Restricted to recruiters who responded to the follow-up survey
- The amount of estimation is measured by heaping responses (multiples of 5 or 10)



# Study 1: Results

# What do different degree questions (“knowing someone”, “being close”) mean to the respondents?

- The degree of intimacy of the definition of “knowing” someone varied greatly
  - For some, “knowing” is someone they recognize, for others, it is someone they are close to:

*“Oh yeah like people who could recognize me, or I would say ‘Hi’, I would say ‘Hi’ and greet.” – Age 28, White, Bisexual, Transgender man*

*“They know how I think, they know what I like and what I dislike.” – Age 50, Black, Gay, Transgender woman*

# What do different degree questions (“knowing someone”, “being close”) mean to the respondents?

- Definition of “being close” to someone is less varied
- Generally being “close” is to be able to confide in them, lean on for support:

*“Like I can talk to you, and tell my problems, explain my problems on all levels, like our body problems.” – Age 23, Black, Lesbian, Transgender woman*

*“I guess it mostly boils down to a transaction of information between each people and the specificity of that information let’s say.” – Age 34, White, Bisexual, Cisgender woman*

# Which degree questions are the most useful for RDS? (1)

From the interview transcripts:

- Respondents mentioned alters who they know have time to respond to the survey, have interest in the survey, frequently interact with, are close to, and may have a monetary need
- This implies that they have more than a passing acquaintance with the alters they are likely to invite
- The general degree that asks about the people they “know” is not specific enough — using the close degree is more accurate

# Which degree questions are the most useful for RDS? (2)

	<b>Est (SE)</b>
	Overall n = 1012
<b>Relationship: Family</b>	1.3% (0.4%)
<b>Relationship: Partner</b>	18.2% (1.2%)
<b>Relationship: Friend</b>	80.1% (1.3%)
<b>Relationship: Other</b>	0.4% (0.2%)
<b>Closeness (1: Not at all close, 7: Very Close)</b>	5.02 (0.04)
<b>No. of years respondents have known the alter</b>	4.27 (0.16)

# Which degree questions are the most useful for RDS? (3)

Personal network size type	Mean (Range)	Reliability	Correlation with PNS (invite)	Proportion of heaping
PNS (all)	6.29 (1-318)	.74	.84	10.3%
PNS (18+)	6.08 (1-318)	.71	.85	9.2%
PNS (close)	3.58 (1-36)	.66	.90	3.8%
PNS (interact)	3.54 (0-50)	.64	.88	4.3%
PNS (interest)	3.36 (0-20)	.67	.96	4.9%
PNS (invite)	3.24 (0-20)	.68	-	3.8%

Note: Analytic sample for test-retest reliability restricted to respondents who completed the follow-up survey

# Summary

- “Knowing” has a very varied interpretation across respondents – it is likely to lead to inconsistent reporting across respondents
- The *recruitable network* is better targeted with specific degree questions which asks for closer relationships
  - This is limited to LGBTQ population — other populations might need different questions
- Specific degree questions have more realistic ranges and less heaping