# Network Sampling with Memory: A Proposal for more efficient sampling from networks

Ted Mouw
Ashton Verdery

Department of Sociology
University of North Carolina, Chapel Hill

March 25, 2012

# Outline

# motivation

- immigration is a network process
- lack of data on social networks and international migration–>collect network data
- mixed method community studies
- vs. large scale seconadary data sets

# is it feasible?

- is it feasible to collect network data on hidden populations?
- (1) 2010 NSIT
- (2) cost effectiveness of gains in precision.

# 2010 NSIT

- This research is a pilot study of migrant networks connecting a town in Guanajuato, Mexico with Chapel Hill, North Carolina and Houston, Texas (with Sergio Chavez)
- despite the distance, strong ties exist between these geographically separate communities
- we want to sample from the network itself

Trip: **2104.5 mi, 34 hr 47 min**

# Survey

- 2010 Network Survey of Immigrant Transnationalism
- pilot study for a larger project–>longitudinal with 5 destinations.
- Network + Community study of immigration in Chapel Hill, Houston, and Guanajuato, Mexico
- N=140 in N.C., 50 in Houston, and 420 in Guanajuato.
- Over 8,000 nominations in the network
- questions about migration history, employment, social incorporation and attitudes, happiness, and health.

**ACULTURACION Y BIENESTAR EMOCIONAL**

**LEER AL PARTICIPANTE:** Ahora hablaremos de su experiencia de adaptación en los Estados Unidos.

| 49. | Generalmente, usted diría que su salud es: | |
|---|---|---|
| | ○ | 5. Excelente |
| | ○ | 4. Muy buena |
| | ○ | 3. Buena |
| | ○ | 2. Regular |
| | ○ | 1. Mala |

| 50. | ¿Tiene usted familiares que viven en México? (Marcar todos los que que aplican) | |
|---|---|---|
| | ☐ | Esposo(a) |
| | ☐ | Padres |
| | ☐ | Hijos |
| | ☐ | Otro _____ |

| 51. | Usando una escala de 1 a 5, describe qué tanto extrañas a tus familiares y amigos que viven en México. | |
|---|---|---|
| | ○ | 5. Los extraño muchísimo...es como un dolor en el corazón |
| | ○ | 4. |
| | ○ | 3. |
| | ○ | 2. |
| | ○ | 1. Solo un pequeño _____ |

| 52. | Reflexionando en su actual felicidad, ¿donde piensa usted que estaría más feliz? ¿En Texas o en su lugar de origen en México? | |
|---|---|---|
| | ○ | aquí en Texas |
| | ○ | en mi lugar de origen en Mexico |
| | ○ | Igual _____ |

| 53. | ¿Que feliz esta usted? [leer categorías] |
|---|---|
| | ○ 5 Extremadamente feliz |
| | ○ 4 Muy feliz |
| | ○ 3 Feliz |
| | ○ 2 Un poco feliz |
| | ○ 1 Infeliz |

| 54. | ¿Que es su opinión sobre la cultura de los Estados Unidos? [donde 5 es "me encanta" y 1 es "no me gusta" ] | |
|---|---|---|
| | ○ | 5. me encanta |
| | ○ | 4. |
| | ○ | 3. |
| | ○ | 2. |
| | ○ | 1. no me gusta _____ |

| 55. | Usando una escala de 1 a 5, describe qué tanto deseas adaptarte a la sociedad, cultura, y tipo de vida en Texas | |
|---|---|---|
| | ○ | 5. quiero adaptarme completamente |
| | ○ | 4. |
| | ○ | 3. |
| | ○ | 2. |
| | ○ | 1. quiero adaptarme solo lo necesario _____ |

| 56. | ¿Para ti qué es mas interesante, las noticias sobre los Estados Unidos o noticias sobre México? | |
|---|---|---|
| | ○ | sobre los Estados Unidos |
| | ○ | sobre México |
| | ○ | igual _____ |

| 57. | Usando una escala de 1 a 5, describe que tanto te identificas como Mexicano. [donde 5 es 100% y 1 es 0% ] | |
|---|---|---|
| | ○ | 5. 100%, completamente |
| | ○ | 4. 75% |
| | ○ | 3. 50% |
| | ○ | 2. 25% |
| | ○ | 1. 0% _____ |

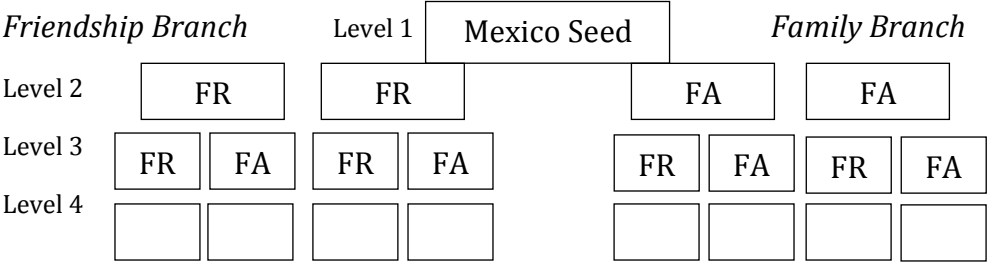
| 58. | Si usted pudiera arreglar o ya arregló su estado migratorio, ¿preferiría vivir permanente en los Estados Unidos? | ○ | Si | ○ | No |
|---|---|---|---|---|---|

**HOJA 1: APUNTAR LOS NOMBRES DE AMIGOS O CONOCIDOS MAYORES DE 18 AÑOS QUE VIVEN HOUSTON, TEXAS.**

Empezamos. ¿Podría apuntar información básica acerca sus redes sociales en la hoja número uno? Además, es importante que anote los nombres y apodos, la información sobre ocupación, edad, y otras características para evitar confundir personas que tengan nombres similares.

| # | Nombre (completo) | Apellido (si lo sabe)<br><br>(solo primeras 4 letras) | Apodo (si lo sabe) | Ocupación (tipo de trabajo) | Sexo<br>H= hombre<br><br>M= mujer | Edad | ¿Vive aquí con niños propios? (indique cuantos) | Lugar de origen: Estado o ciudad | ¿Cuántos años hace que conoció a esta persona? | Aproxima-damente, ¿cuántos años vivió en Houston, Texas? | ¿Con que frecuencia usted se comunica con esta persona?<br>1=cada día<br>2=cada semana<br>3=cada mes<br>4=cada año<br>5=menos que cada año |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A1 | | | | | H M | | 0 1 2 3+ | | | | |
| A2 | | | | | H M | | 0 1 2 3+ | | | | |
| A3 | | | | | H M | | 0 1 2 3+ | | | | |
| A4 | | | | | H M | | 0 1 2 3+ | | | | |
| A5 | | | | | H M | | 0 1 2 3+ | | | | |
| A6 | | | | | H M | | 0 1 2 3+ | | | | |
| A7 | | | | | H M | | 0 1 2 3+ | | | | |
| A8 | | | | | H M | | 0 1 2 3+ | | | | |
| A9 | | | | | H M | | 0 1 2 3+ | | | | |
| A10 | | | | | H M | | 0 1 2 3+ | | | | |

**Diagram 1: Selection of Friend and Family Members in Mexico**



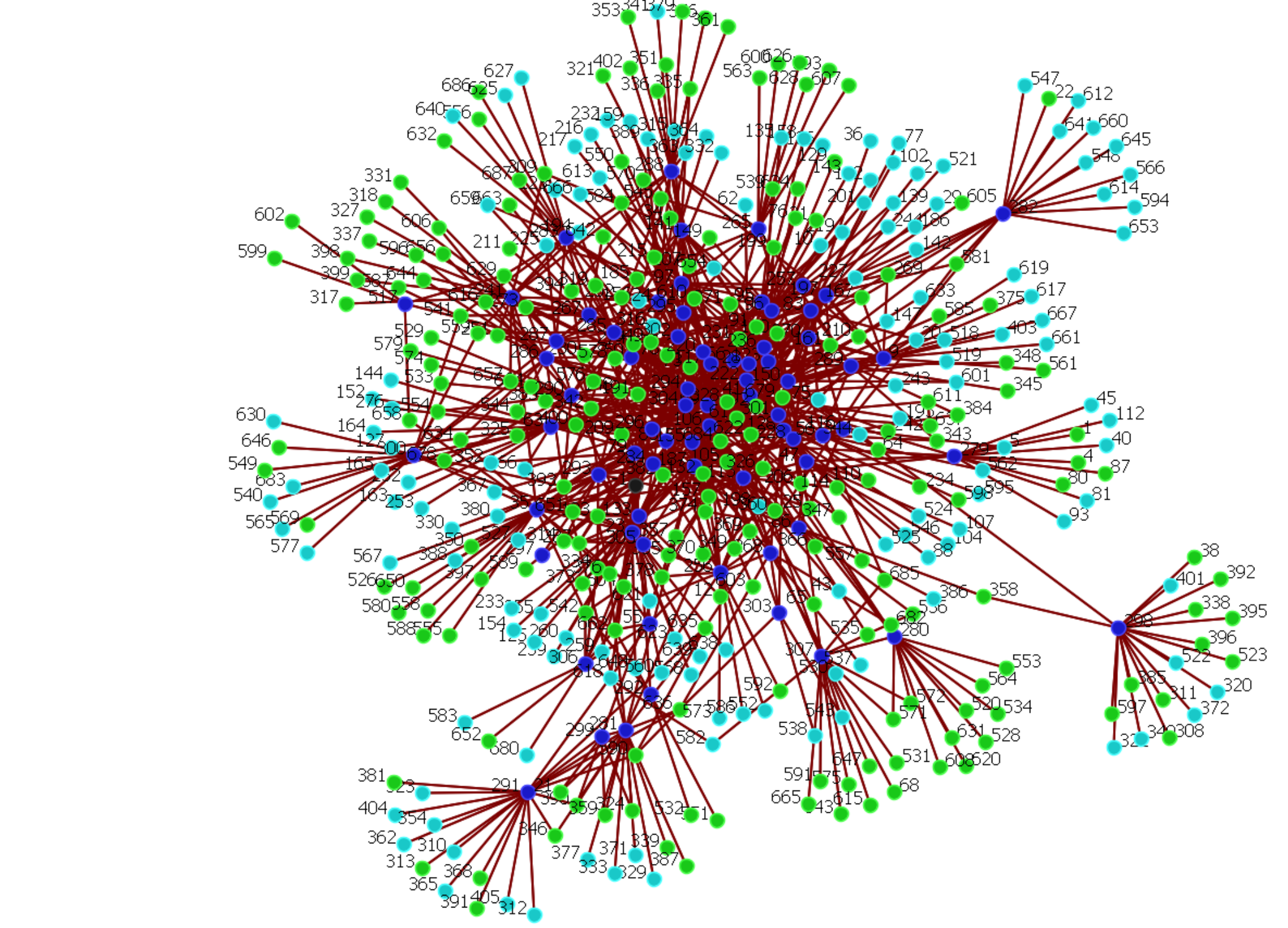
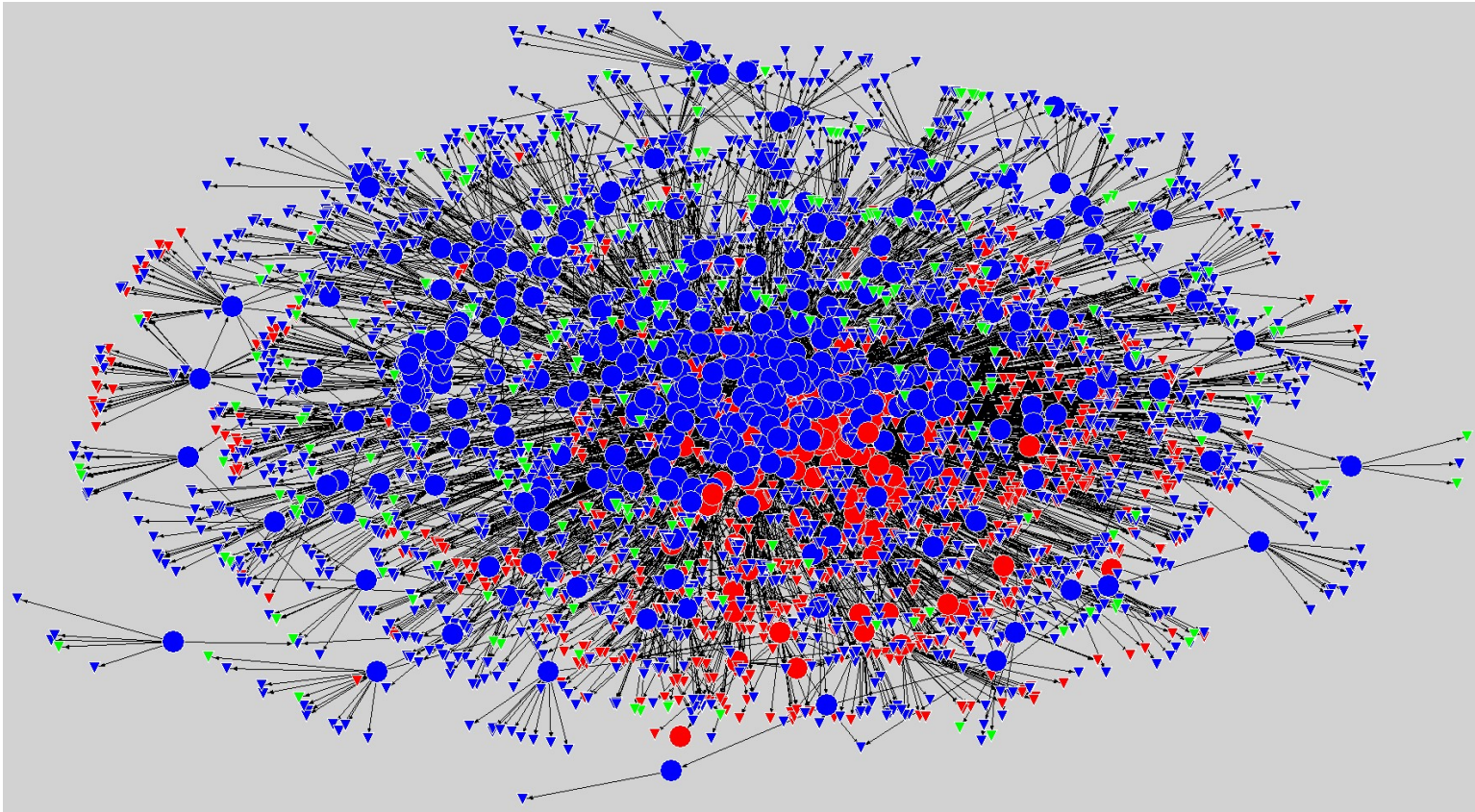


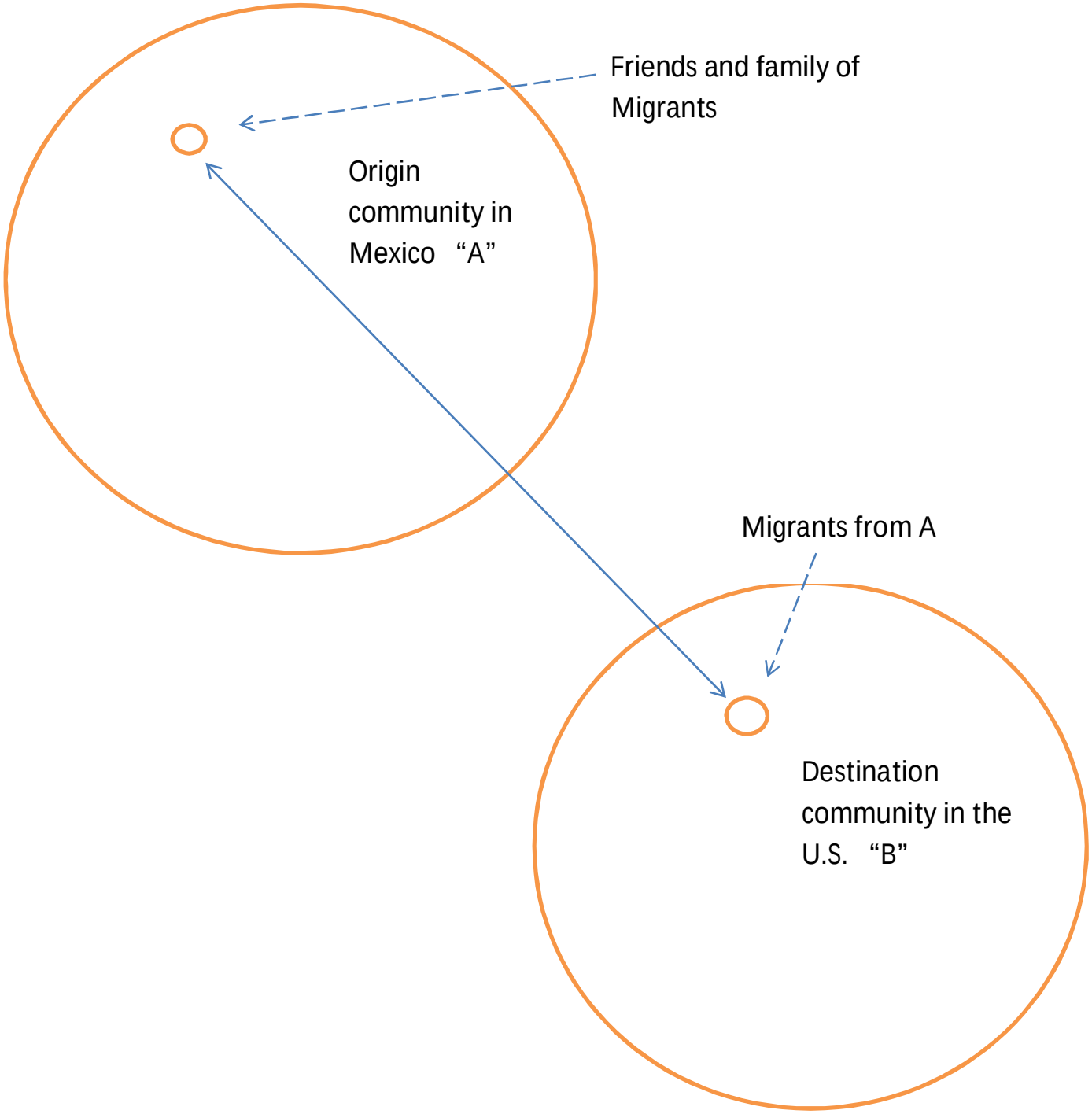
**Figure 1: The binational network of sampled and nominated individuals in the 2010 Network Survey of Immigrant Transnationalism.**



**Notes: Red nodes are located in North Carolina, green nodes are located in Houston, and blue nodes are located in Mexico. Large circles indicate interviewed cases, and small triangles indicate nominated but not interviewed cases.**

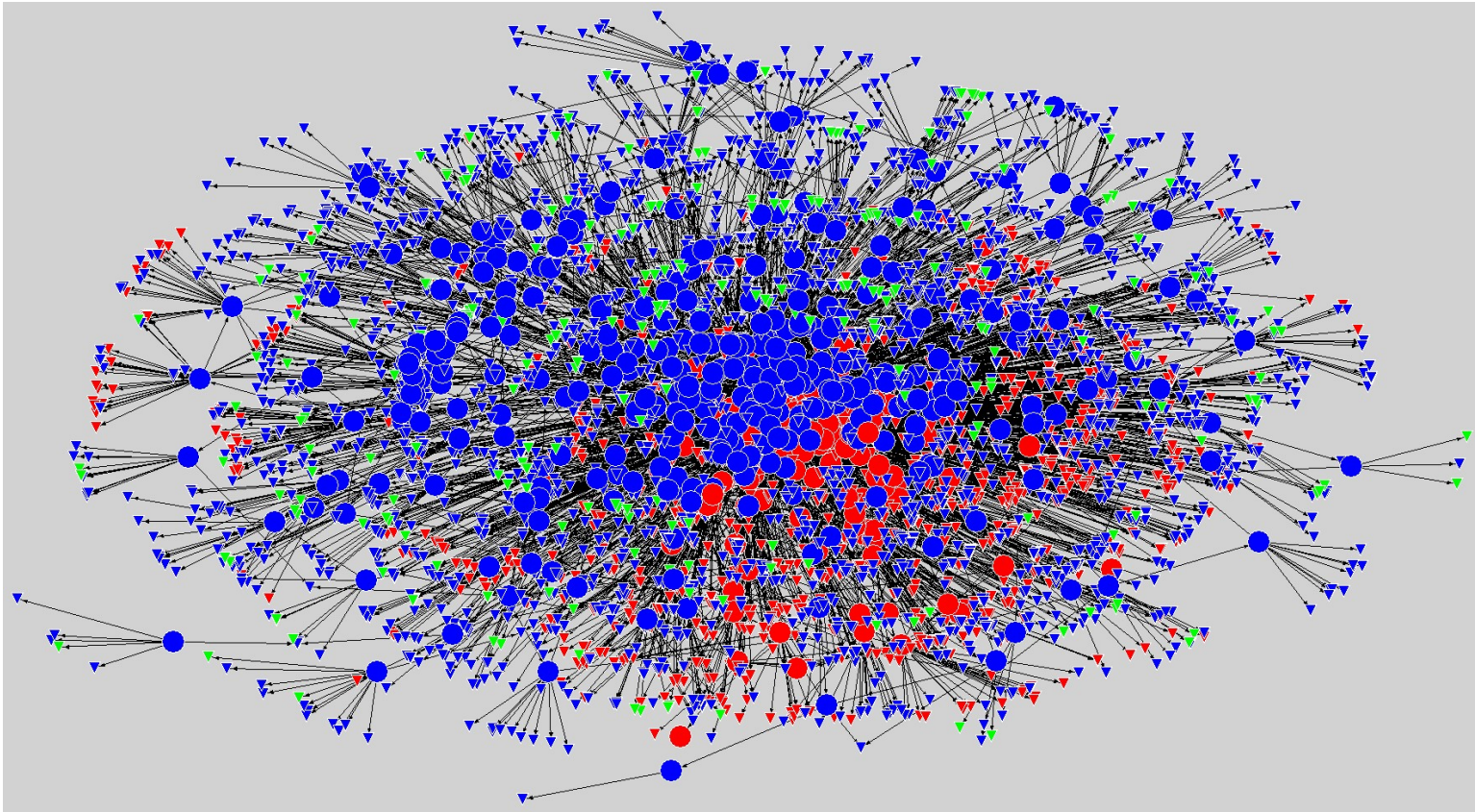
# basic problem

Sampling migrant networks—the needle in a haystack problem



# Hidden and rare Populations

**Figure 1: The binational network of sampled and nominated individuals in the 2010 Network Survey of Immigrant Transnationalism.**



**Notes: Red nodes are located in North Carolina, green nodes are located in Houston, and blue nodes are located in Mexico. Large circles indicate interviewed cases, and small triangles indicate nominated but not interviewed cases.**

# Hidden and rare Populations

Collecting data from a hidden population is difficult because of the absence of a sampling frame

# Hidden and rare Populations

Collecting data from a hidden population is difficult because of the absence of a sampling frame

"Respondent Driven Sampling" (RDS)–a random walk (RW) based approach

# Hidden and rare Populations

Collecting data from a hidden population is difficult because of the absence of a sampling frame

"Respondent Driven Sampling" (RDS)–a random walk (RW) based approach

current respondent gives 1-3 coupons to friends, who become the next wave of respondents

# accuracy of network sampling

.

Bias

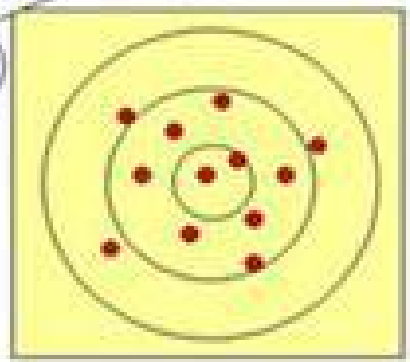
–RWs are unbiased in large samples

–exhibit bias in finite samples

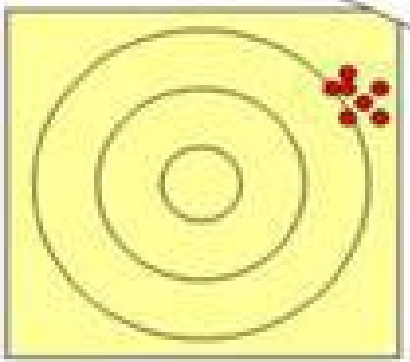
# Accuracy vs.Precision

*Measure of bias*

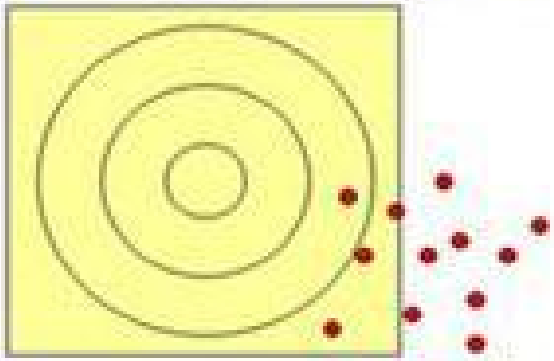
*Measure of spread*



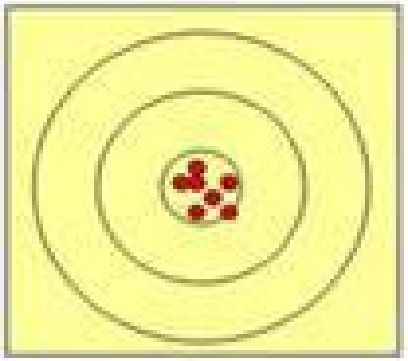
**Accurate but not precise**



**Precise but not accurate**



**Not accurate or precise**



**Both accurate and precise**

# sampling variance

Sampling variance

–Design effect (DE) is the ratio of sampling variance of the network sample to simple random sampling.

# sampling variance

Sampling variance

–Design effect (DE) is the ratio of sampling variance of the network sample to simple random sampling.

–the DE of RWs and RDS is a function of the structure of the network (in addition to sample size)

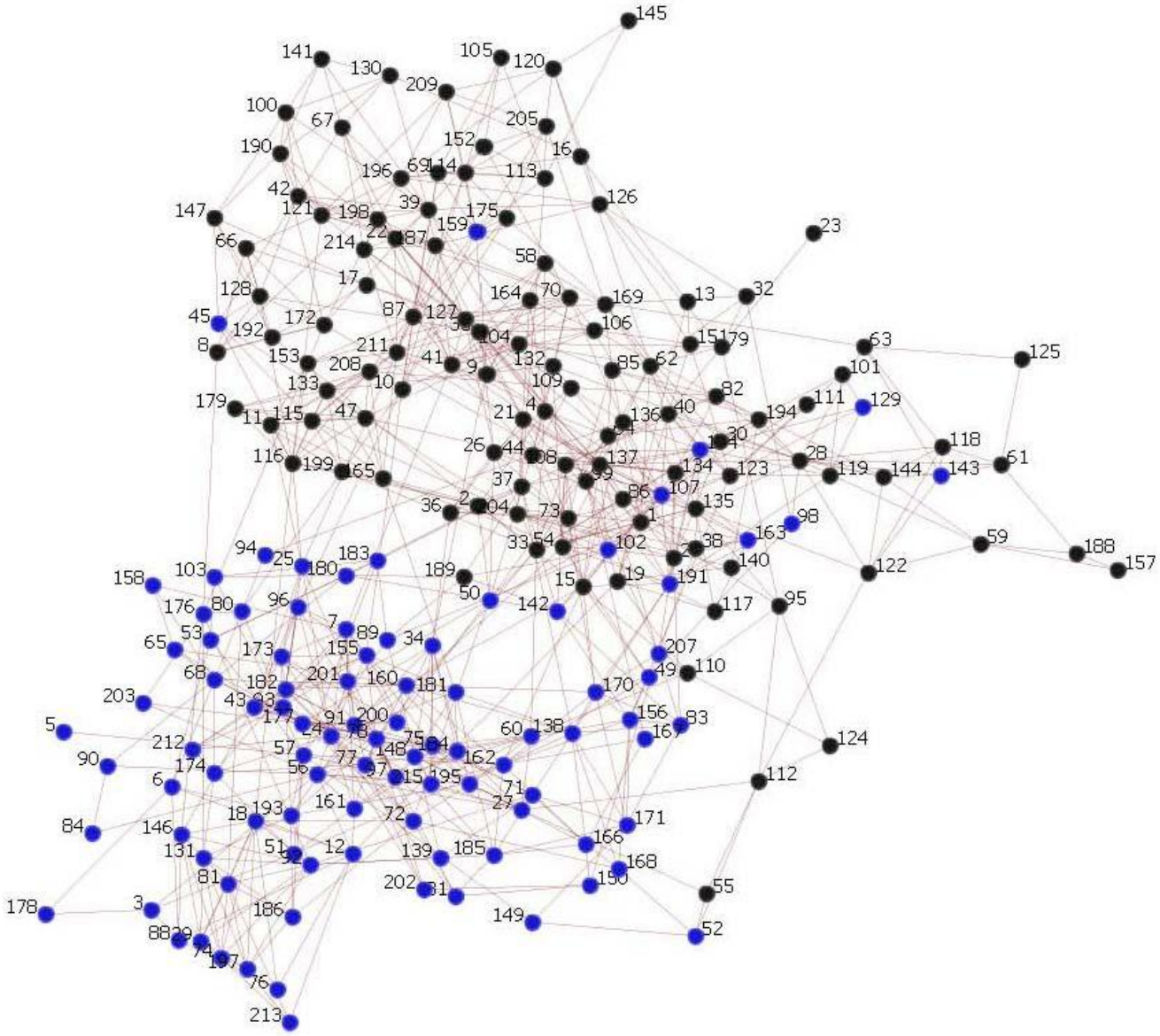
# sampling variance

Sampling variance
–Design effect (DE) is the ratio of sampling variance of the network sample to simple random sampling.

–the DE of RWs and RDS is a function of the structure of the network (in addition to sample size)

Dartboard analogy of accuracy (sampling bias) and precision (sampling variance)

Figure 4: Add Health Network # 112.



Notes: Nodes colored by student race. [Black = White , Blue = Non-white]
Node ID numbers have been randomly assigned.

# Network Sampling with Memory

.

Mouw and Verdery. 2011 "Network Sampling with Memory: A proposal for more efficient sampling from social networks"

Collect network data (example: 2010 NSIT)

Use the network data to sample more efficiently.

# Network Sampling

two sampling modes, List and Search

# Network Sampling

two sampling modes, List and Search

List mode: (a) keep a list, L, of all nominated network members

# Network Sampling

two sampling modes, List and Search

List mode: (a) keep a list, L, of all nominated network members
(b) sample with replacement from L

# Network Sampling

two sampling modes, List and Search

List mode: (a) keep a list, L, of all nominated network members
(b) sample with replacement from L
(c) "Even sampling" of new nodes–sample new nodes at the current cumulative sampling rate

# Network Sampling

two sampling modes, List and Search

List mode: (a) keep a list, L, of all nominated network members
(b) sample with replacement from L
(c) "Even sampling" of new nodes–sample new nodes at the current cumulative sampling rate
very simple

# Network Sampling

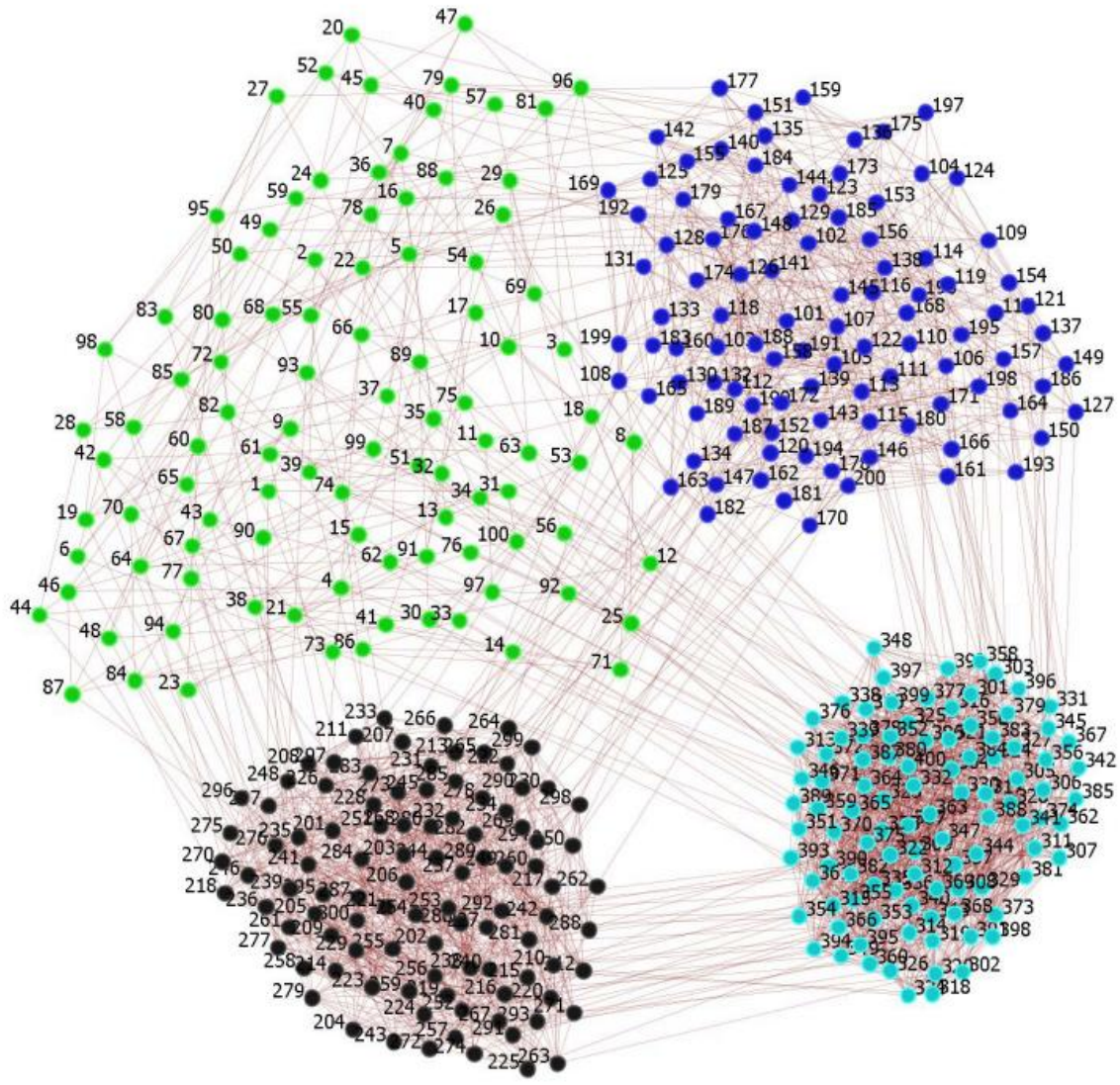
two sampling modes, List and Search

List mode: (a) keep a list, L, of all nominated network members
(b) sample with replacement from L
(c) "Even sampling" of new nodes–sample new nodes at the current cumulative sampling rate
very simple
converges to simple random sampling

# list mode

- sample with replacement from L
- “naive” list mode: early nodes are over sampled
- “even sampling”: when L>200–>
- sample newly discovered nodes at “even sampling rate”
- –>the cumulative sampling rate since the start of even sampling

Figure 2: 400-Node Test  Network.



**Example of sampling from the test network: step by step sampling rates.**

| | | | | | | Cumulative sampling rate of Nodes: . | | | |
|---|---|---|---|---|---|---|---|---|---|
| "Step" | ID of interviewed node | Network size | # of Eligible nodes | Cumulative Even sampling Rate | Sampling rate for step | Node 1 | Node 101 | Node 201 | Node 301 |
| 1 | 25 | 8 | 8 | 0 | 0.125 | 0 | 0 | 0 | 0 |
| 2 | 76 | 14 | 14 | 0 | 0.071429 | 0 | 0 | 0 | 0 |
| 3 | 21 | 20 | 20 | 0 | 0.05 | 0 | 0 | 0 | 0 |
| 4 | 21 | 20 | 20 | 0 | 0.05 | 0 | 0 | 0 | 0 |
| 5 | 14 | 25 | 25 | 0 | 0.04 | 0.04 | 0 | 0 | 0 |

After 5 steps, the revealed network size is 25 [i.e., 25 different nodes have been nominated]
Node 1 is nominated for the first time in step 5, and Node 301 is nominated for the first time in step 7

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 6 | 378 | 46 | 46 | 0 | 0.021739 | 0.061739 | 0 | 0 | 0 |
| 7 | 323 | 63 | 63 | 0 | 0.015873 | 0.077612 | 0 | 0 | 0.015873 |
| 8 | 56 | 68 | 68 | 0 | 0.014706 | 0.092318 | 0 | 0 | 0.030579 |
| 9 | 15 | 73 | 73 | 0 | 0.013699 | 0.106017 | 0 | 0 | 0.044278 |
| 10 | 20 | 78 | 78 | 0 | 0.012821 | 0.118837 | 0 | 0 | 0.057098 |
| 11 | 369 | 87 | 87 | 0 | 0.011494 | 0.130331 | 0 | 0 | 0.068592 |
| 12 | 344 | 99 | 99 | 0 | 0.010101 | 0.140432 | 0 | 0 | 0.078693 |

After 39 steps, Nodes 1 and 301 have large cumulative sampling rates. They will be temporarily excluded from the pool of eligible nodes once even sampling is turned on.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 39 | 322 | 194 | 194 | 0 | 0.005155 | 0.323978 | 0 | 0 | 0.262239 |
| 40 | 81 | 199 | 199 | 0 | 0.005025 | 0.329004 | 0 | 0 | 0.267264 |
| 41 | 79 | 203 | 100 | 0.01 | 0.01 | 0.329004 | 0 | 0 | 0.267264 |

"Even sampling" is turned on in step 41 because the network size > 200. Now, all newly nominated nodes are sampled at the current even sampling rate. Node 101 is nominated for the first time in step 43.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 42 | 138 | 211 | 100 | 0.02 | 0.01 | 0.329004 | 0 | 0 | 0.267264 |
| 43 | 177 | 213 | 100 | 0.03 | 0.01 | 0.329004 | 0.03 | 0 | 0.267264 |
| 44 | 122 | 218 | 100 | 0.04 | 0.01 | 0.329004 | 0.04 | 0 | 0.267264 |

After 44 steps, the even sampling rate is .04. Nodes 1 and 301 are still excluded from sampling eligibility because their cumulative sampling rate is greater than the even sampling rate.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 79 | 199 | 357 | 270 | 0.205092 | 0.003704 | 0.329004 | 0.205092 | 0 | 0.267264 |
| 80 | 294 | 357 | | 0.205092 | | 0.329004 | 0.205092 | 0 | 0.267264 |
| 81 | 287 | 358 | | 0.205092 | | 0.329004 | 0.205092 | 0 | 0.267264 |
| 82 | 217 | 358 | | 0.205092 | | 0.329004 | 0.205092 | 0.205092 | 0.267264 |

Node 201 is nominated for the first time in step 82. Steps 80-83 are "catch-up" interviews of newly nominated cases.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 83 | 224 | 358 | | 0.205092 | | 0.329004 | 0.205092 | 0.205092 | 0.267264 |
| 84 | 270 | 369 | 282 | 0.208638 | 0.003546 | 0.329004 | 0.208638 | 0.208638 | 0.267264 |
| 85 | 305 | 369 | 282 | 0.212184 | 0.003546 | 0.329004 | 0.212184 | 0.212184 | 0.267264 |

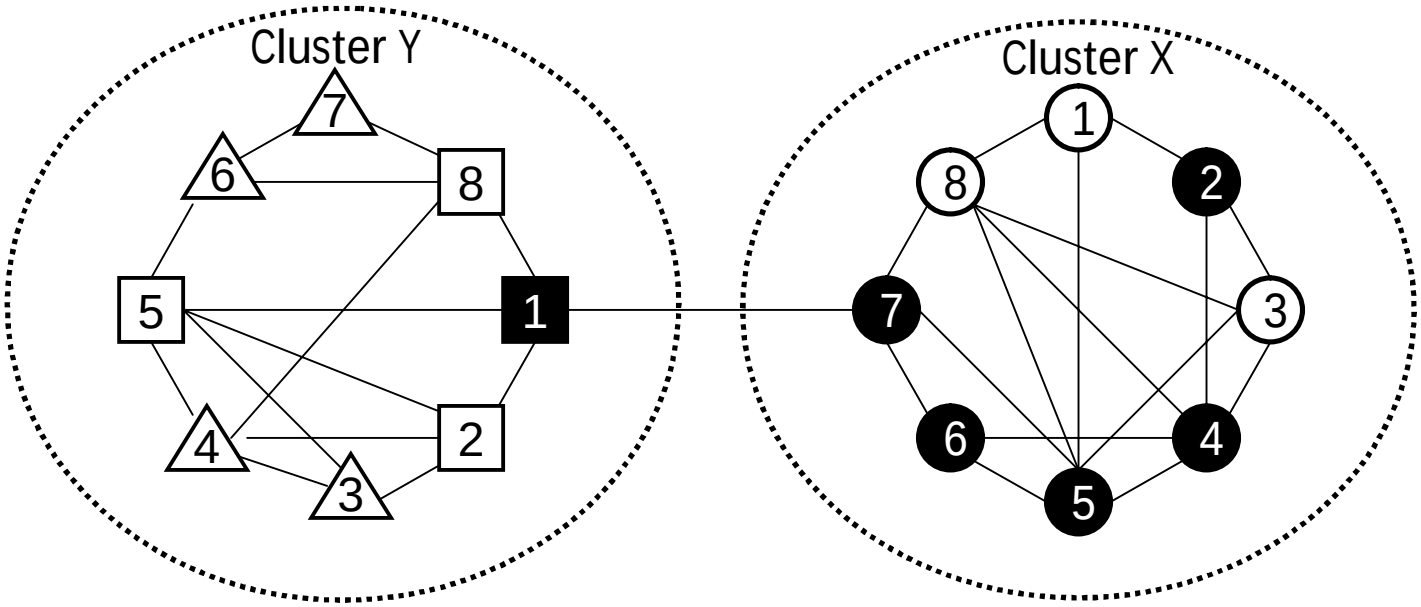
In Step 251, all 400 nodes have been nominated, and the sample is $\cong$ a simple random sample

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 248 | 101 | 399 | 399 | 0.633599 | 0.002506 | 0.632375 | 0.633599 | 0.633599 | 0.633053 |
| 249 | 389 | 399 | 399 | 0.636105 | 0.002506 | 0.634881 | 0.636105 | 0.636105 | 0.63556 |
| 250 | 300 | 399 | 399 | 0.638612 | 0.002506 | 0.637388 | 0.638612 | 0.638612 | 0.638066 |
| 251 | 275 | 400 | 400 | 0.641112 | 0.0025 | 0.639888 | 0.641112 | 0.641112 | 0.640566 |

# search mode

- Search mode: push the sample to explore the network
- Bridge tie: a node that connects two clusters of nodes

**Figure 1. Illustrative network with two clusters.**




**Notes: Hollow nodes are unsampled, dark nodes are sampled. Circles indicate nodes nominated 2+ times, squares indicate nodes nominated 1 time, triangles indicate nodes nominated 0 times.**

# hybrid approach

- List mode [sample from list L]
- Search mode [sample friends of bridge ties]
- NSM hybrid approach: start in Search mode
- –switch to List mode as the network is explored

Calculating Sampling weights

The estimated probability that node ii has been nominated after S steps is:

$$\text{p\_nom}_i \cong 1\text{-}(1\text{-}\frac{d_i}{G})^S$$

Where $d_i$ is the degree of node i, and G is the estimated size of the network.

Example: what is the probability that a node with degree 5 has been nominated after 200 steps in a 1092 node network?

$$\text{p\_nom}_i \cong 1\text{-}(1\text{-}\frac{5}{1092})^{200} = 1 - (0.99542)^{200} = 0.6$$

$$\text{Sampling weight} = \frac{1}{\left[\text{p\_nom}_i \times \text{cumulative node-specific sampling rate}_i\right]}$$

Figure 2: 400-Node Test  Network.

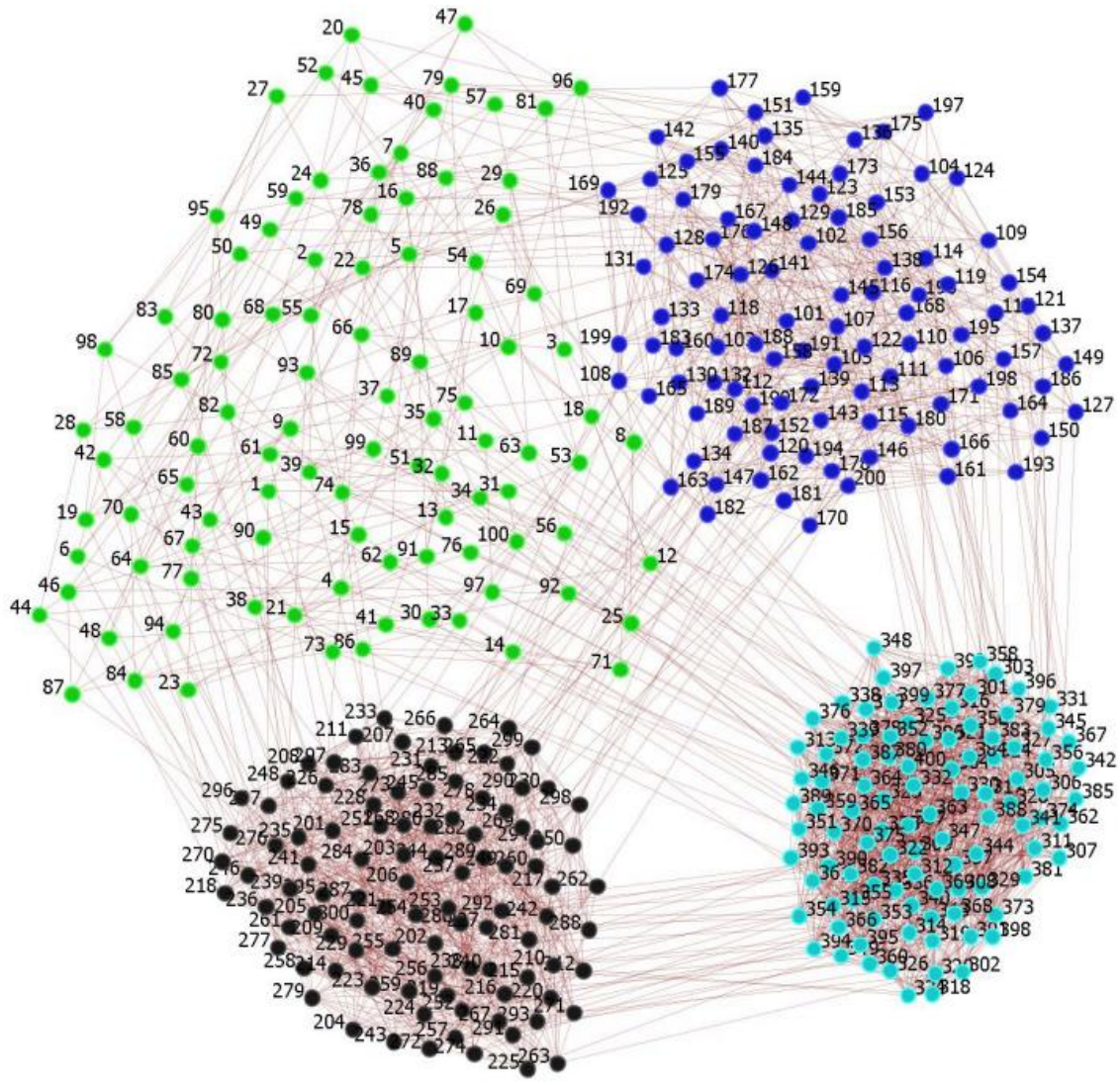
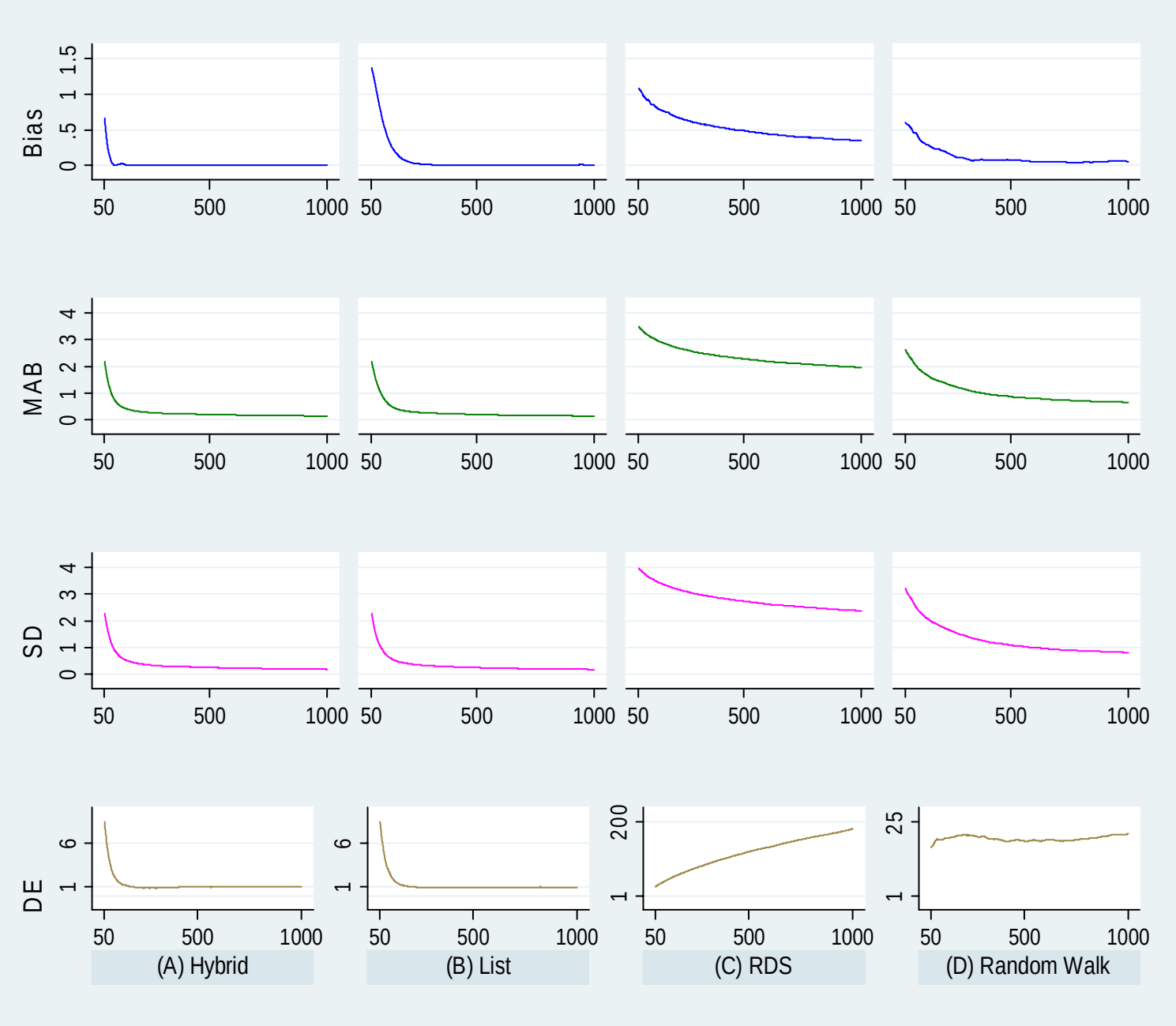


Figure 3: Test Network sampling results,
(400 nodes, dependent variable: average degree)




**Notes: Bias = average bias; MAB = mean absolute bias;** Design effect (DE) = $\frac{\sigma^2_{method}}{\sigma^2_{SRS}}$

**Table 2: Test network sampling results**
**(400 nodes, dependent variable: average degree)**

| | | Method | | | | |
|---|---|---|---|---|---|---|
| Variable | Steps | Hybrid | List | Naive List | RDS | Random Walk |
| Average Bias | 250 | .00036 | .00472 | .1472 | .1278 | .02825 |
| | 500 | .000084 | .000014 | .07651 | .09775 | .01458 |
| | 1000 | 5.00e-06 | .001203 | .03781 | .06936 | .01095 |
| Mean Absolute Bias | 250 | .05281 | .05526 | .1831 | .5254 | .2555 |
| | 500 | .04001 | .03938 | .09757 | .4568 | .173 |
| | 1000 | .02756 | .02808 | .05165 | .3933 | .1304 |
| Design Effect | 250 | .8986 | .9615 | 5.49 | 76.95 | 20.72 |
| | 500 | 1.015 | .962 | 3.221 | 119.6 | 18.8 |
| | 1000 | .9963 | .9845 | 2.066 | 181 | 21.1 |
| Standard Deviation | 250 | .06711 | .06942 | .1659 | .6211 | .3223 |
| | 500 | .05045 | .0491 | .08985 | .5476 | .2171 |
| | 1000 | .03533 | .03512 | .05088 | .4762 | .1626 |

# Results

- Test NSM vs RWs and RDS using 162 university and school networks from Facebook and Add Health
- size ranges from 300 to 16,500 nodes
- estimate % white (Add Health) and % first year students (Facebook)
- start from a randomly selected student, repeat 500 times for each network
- calculate bias, design effects, and mean absolute bias

.

# Results

- NSM has a 97.5% reduction in the design effects on these 162 networks
- 1.16 for NSM
- 77.38 for RDS

**Figure 6: Sampling results the largest Facebook university network (16,280 nodes, dependent variable: proportion freshman)**

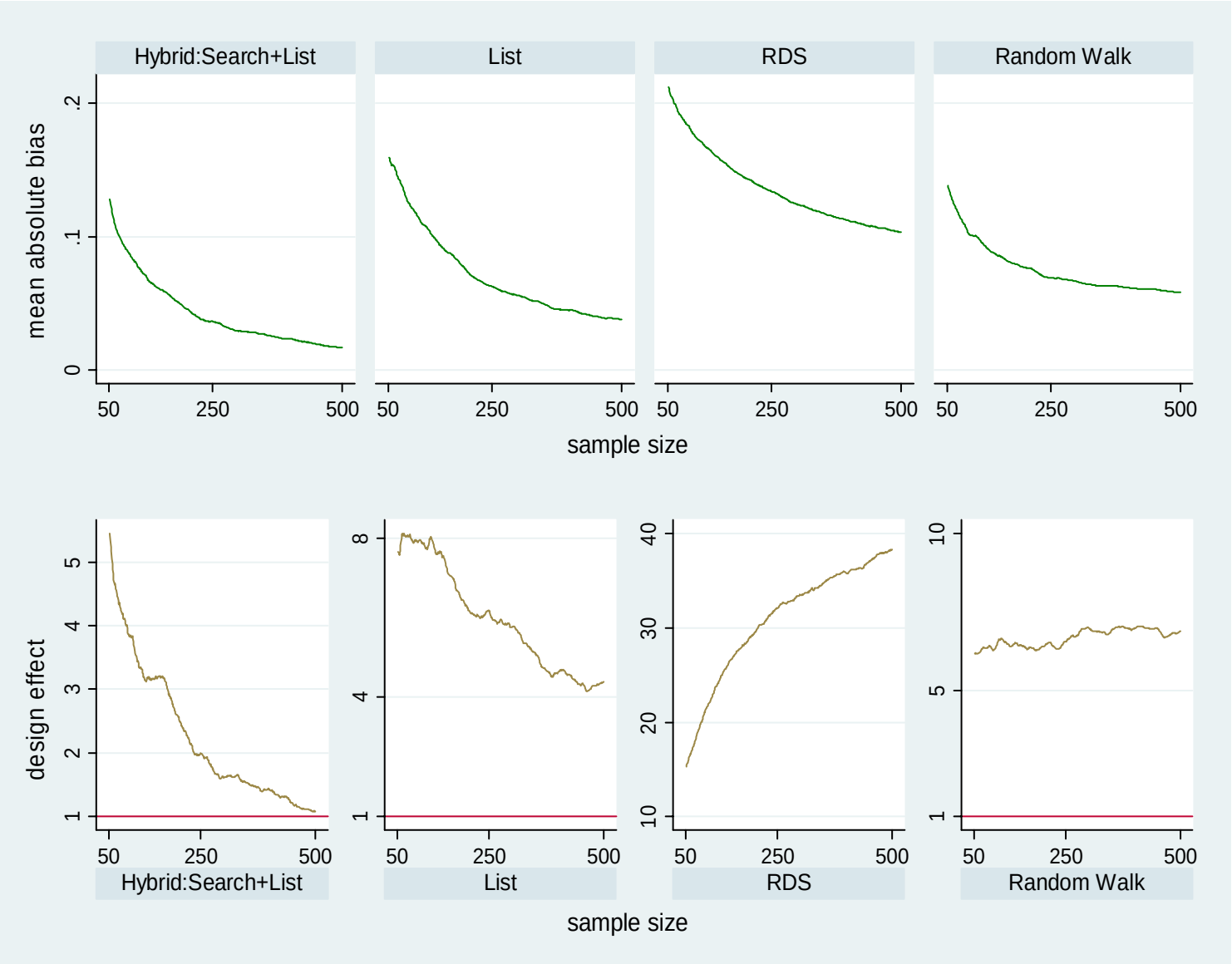


**Table 6A. Overall Add Health results for proportion white in the school.**

| | Network Size | | |
|---|---|---|---|
| | **Micro network** | **Small network** | **Overall avg.** |
| **Avg. # of nodes** | 264 | 741 | 495 |
| **Average Design effects** | | | |
| NSM: hybrid | 1.03 | 1.20 | 1.11 |
| NSM: list | 1.03 | 1.33 | 1.18 |
| Naive list | 3.75 | 7.59 | 5.61 |
| RDS | 67.80 | 65.91 | 66.88 |
| RW | 13.43 | 11.92 | 12.70 |
| **Median Design Effects** | | | |
| NSM: hybrid | 1.02 | 1.17 | 1.06 |
| NSM: list | 1.02 | 1.20 | 1.09 |
| Naive list | 2.31 | 4.64 | 3.32 |
| RDS | 47.41 | 49.26 | 47.77 |
| RW | 9.27 | 9.24 | 9.27 |
| **Average bias** | | | |
| NSM: hybrid | 0.0011 | 0.0028 | 0.0019 |
| NSM: list | 0.0012 | 0.0029 | 0.0020 |
| Naive list | 0.0122 | 0.0217 | 0.0168 |
| RDS | 0.0094 | 0.0088 | 0.0091 |
| RW | 0.0034 | 0.0029 | 0.0032 |
| **Median Average. Bias** | | | |
| NSM: hybrid | 0.0008 | 0.0018 | 0.0011 |
| NSM: list | 0.0011 | 0.0020 | 0.0013 |
| Naive list | 0.0073 | 0.0201 | 0.0116 |
| RDS | 0.0050 | 0.0049 | 0.0050 |
| RW | 0.0021 | 0.0023 | 0.0022 |

**Table 8. Overall Facebook Results.**

| | Network Size | | | |
|---|---|---|---|---|
| | **Small network** | **Medium network** | **Large network** | **Overall avg.** |
| **Average # of nodes** | 1,752 | 4,937 | 9,218 | 4,637 |
| **Average Design effects** | | | | |
| NSM: hybrid | 1.22 | 1.31 | 0.96 | 1.20 |
| NSM: list | 1.34 | 2.42 | 3.64 | 2.27 |
| Naive list | 11.56 | 32.03 | 38.11 | 22.85 |
| RDS** | 78.79 | 93.15 | 76.69 | 83.93 |
| RDS full data* | 90.82 | 97.11 | 58.38 | 86.14 |
| RW** | 11.59 | 13.04 | 11.05 | 12.03 |
| RW full data* | 20.45 | 20.15 | 14.88 | 19.11 |
| **Median Design Effects** | | | | |
| NSM: hybrid | 1.20 | 1.26 | 0.92 | 1.19 |
| NSM: list | 1.23 | 2.35 | 3.65 | 1.97 |
| Naive list | 9.34 | 31.23 | 40.04 | 17.51 |
| RDS** | 78.25 | 83.17 | 74.38 | 78.15 |
| RDS full data* | 86.42 | 93.70 | 33.73 | 85.99 |
| RW** | 11.15 | 12.95 | 10.08 | 11.52 |
| RW full data* | 21.20 | 19.66 | 14.83 | 18.29 |
| **Average bias** | | | | |
| NSM: hybrid | 0.0031 | 0.0057 | 0.0063 | 0.0048 |
| NSM: list | 0.0045 | 0.0152 | 0.0183 | 0.0117 |
| Naive list | 0.0405 | 0.0544 | 0.0602 | 0.0483 |
| RDS** | 0.1289 | 0.1335 | 0.1542 | 0.1363 |
| RDS full data* | 0.0194 | 0.0187 | 0.0146 | 0.0180 |
| RW** | 0.1690 | 0.0632 | 0.0613 | 0.1047 |
| RW full data* | 0.0062 | 0.0045 | 0.0035 | 0.0049 |
| **Median Average. Bias** | | | | |
| NSM: hybrid | 0.0018 | 0.0044 | 0.0052 | 0.0037 |
| NSM: list | 0.0038 | 0.0127 | 0.0191 | 0.0091 |
| Naive list | 0.0394 | 0.0544 | 0.0627 | 0.0470 |
| RDS** | 0.1015 | 0.0945 | 0.1280 | 0.1060 |
| RDS full data* | 0.0191 | 0.0170 | 0.0096 | 0.0164 |
| RW** | 0.0701 | 0.0681 | 0.0566 | 0.0691 |
| RW full data* | 0.0056 | 0.0038 | 0.0028 | 0.0044 |

Notes: ** indicates that RDS and RW were run on the same truncated data (maximum degree of 20) as the NSM approaches.
* indicates that RDS and RW were run on the full data, with no limit on maximum degree. See text for details.

**Figure 8: Diagnostic Properties for Facebook network #10 by sample size (step)**

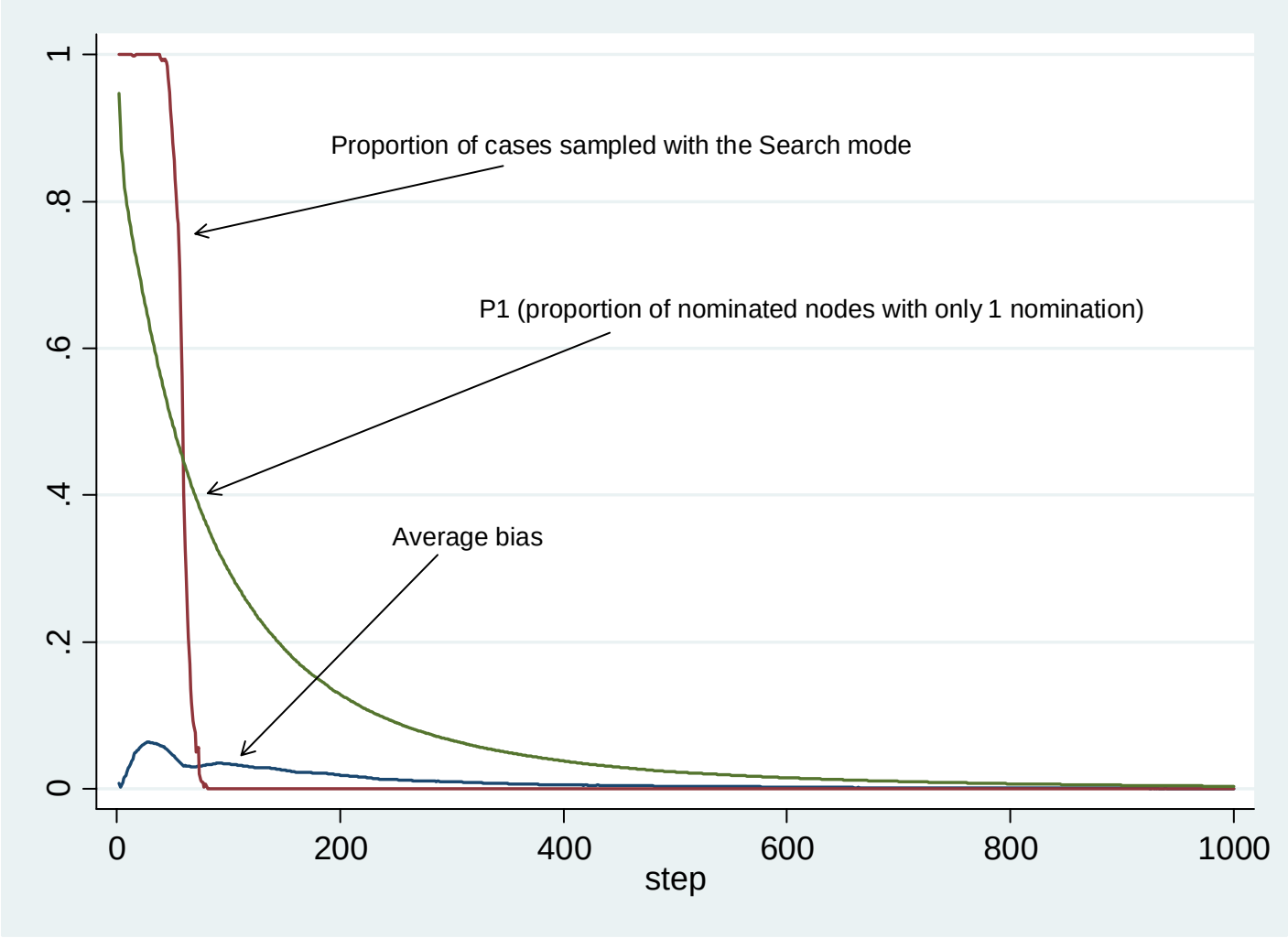
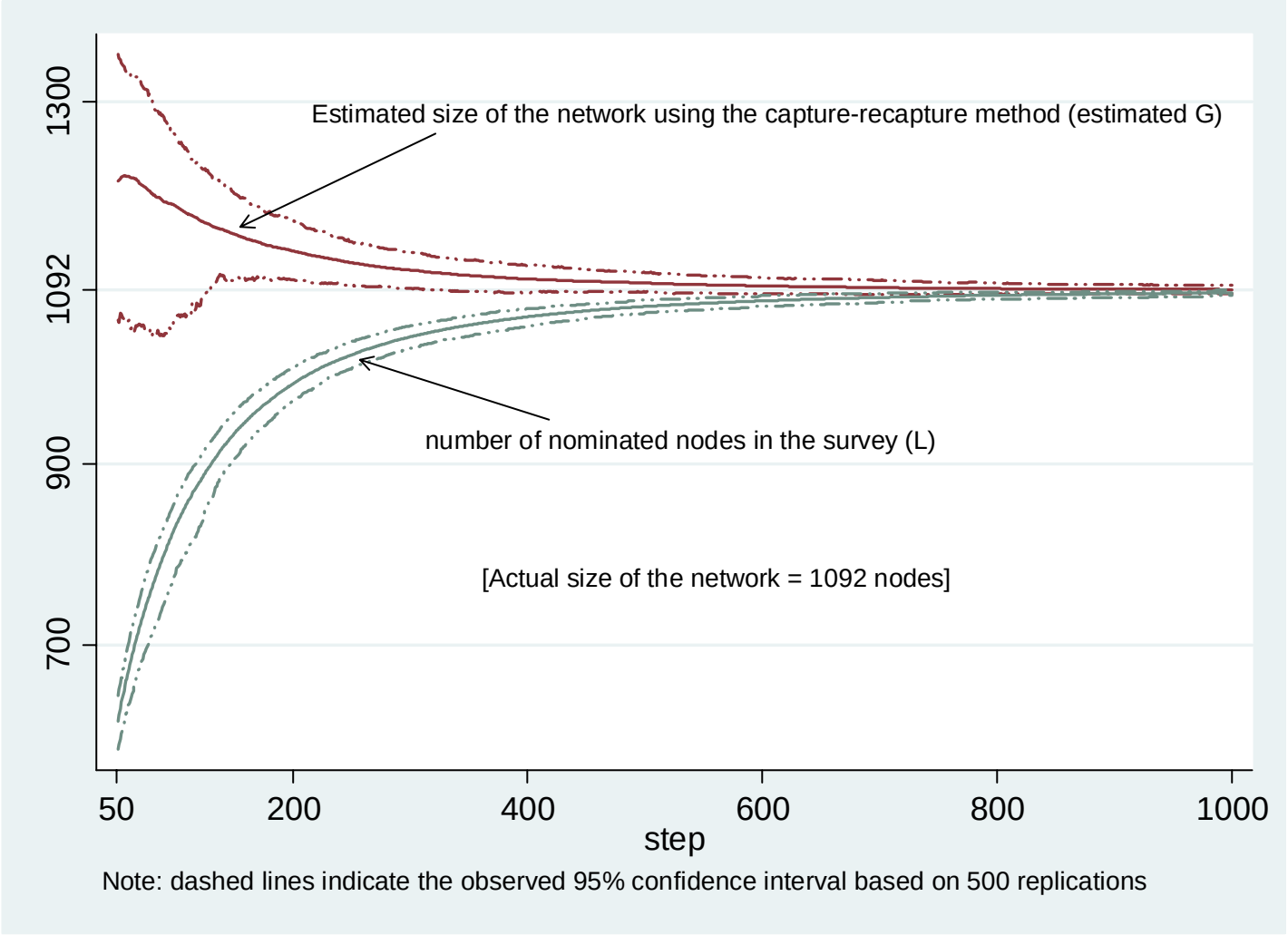


**Figure 9: Number of nominated nodes and estimated network size by sample size (step), Facebook network #10**




Note: dashed lines indicate the observed 95% confidence interval based on 500 replications

**Table 9: OLS regression results for the Design Effect, combined Add Health and Facebook networks**

| | Method | | | | |
|---|---|---|---|---|---|
| | Hybrid | List | Naive List | RDS | Random Walk |
| VARIABLES | | | | | |
| Homophily | 0.158 | -0.136 | 28.36*** | 284.1*** | 40.41*** |
| | (0.144) | (0.338) | (5.563) | (15.16) | (2.757) |
| Add Health network | -0.182 | 0.471 | 16.67*** | 71.37*** | 6.933** |
| | (0.114) | (0.268) | (4.427) | (11.99) | (2.181) |
| Ln(nodes) | -0.00576 | 0.964*** | 12.14*** | -8.917** | -2.030*** |
| | (0.0278) | (0.0654) | (1.252) | (2.930) | (0.533) |
| Average Degree | -0.00111 | -0.00451 | -0.0159 | -0.0950 | 0.0436 |
| | (0.00121) | (0.00285) | (0.0494) | (0.127) | (0.0232) |
| Y-Mean degree difference | 0.00714 | 0.0130 | 0.00201 | -0.139 | 0.207** |
| | (0.00376) | (0.00886) | (0.144) | (0.397) | (0.0722) |
| Constant | 1.202*** | -5.133*** | -94.72*** | -56.61* | 0.0898 |
| | (0.236) | (0.555) | (9.751) | (24.88) | (4.525) |
| Observations | 162 | 162 | 162 | 162 | 162 |
| R-squared | 0.076 | 0.698 | 0.658 | 0.718 | 0.687 |
| Mean design effect | 1.16 | 1.85 | 15.48 | 78.76 | 16.65 |
| S.D. design effect | 0.25 | 1.05 | 15.90 | 48.54 | 8.38 |

Notes: Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

# transnationalism

- recent changes in telecomunications
- Transnationalism: maintaining long term ties to the origin
- living in two places at once
- "social life increasingly takes place across borders" (Levitt and Jaworsky 2007).

# critics

- assimilation perspective:
- transnationalism is not new
- not widespread
- does not meaningfully affect the trajectory of assimilation.

# lack of adequate data

- current tests (entreprenuership, political activity)
- Transnationalism is about a cross-border network of social relations
- need network data

# Binational Study of Migrant Networks

- How does information about jobs flow through binational migrant networks?
- How are these networks strucured?
- How do they affect migration decisions?

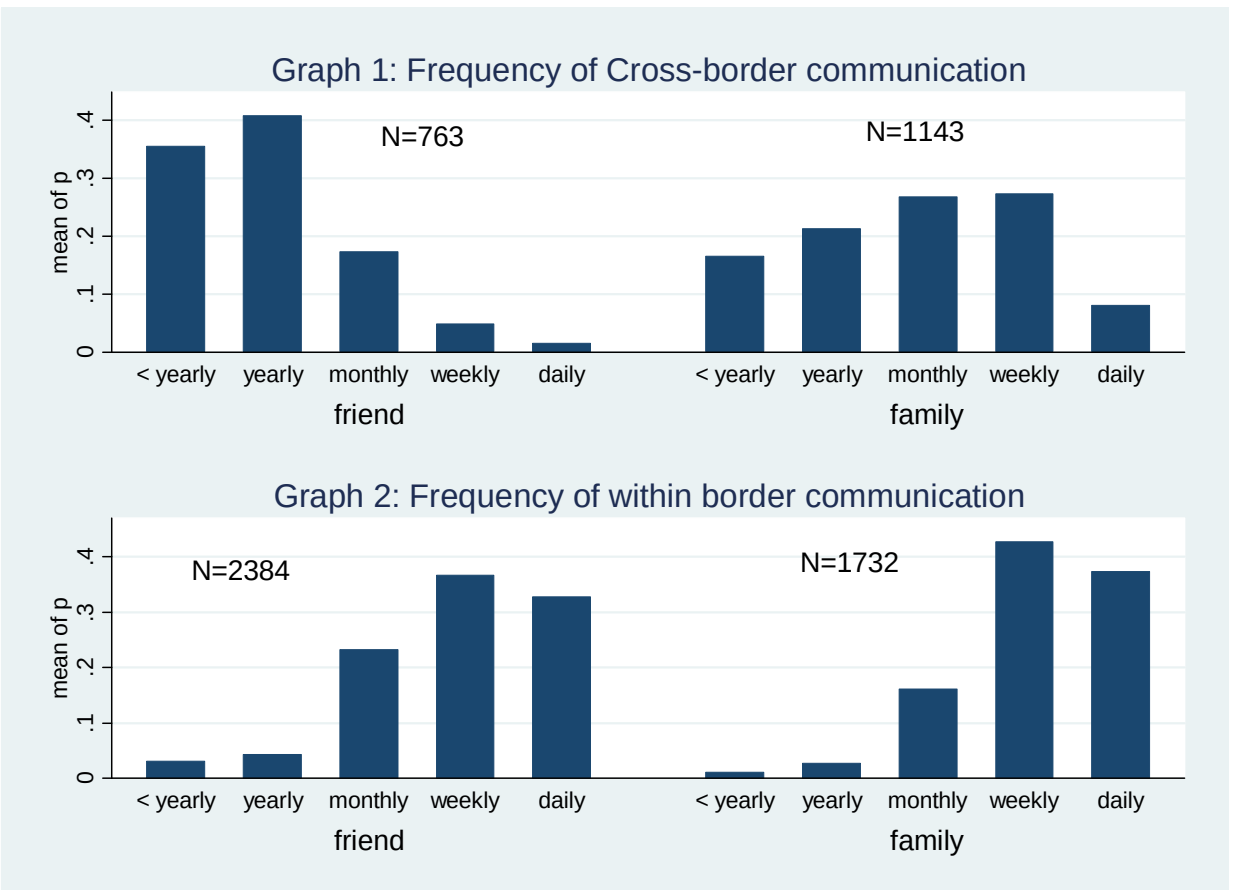
# test of transnationalism

- much debate about the relevance of transnationalism
- no one has actual network data to test it
- Research question:
- how much do social networks and cross-border communication affect the desire for permanent residence in the U.S. ?
- ...and the respondent's opinion about where he/she would be happiest?

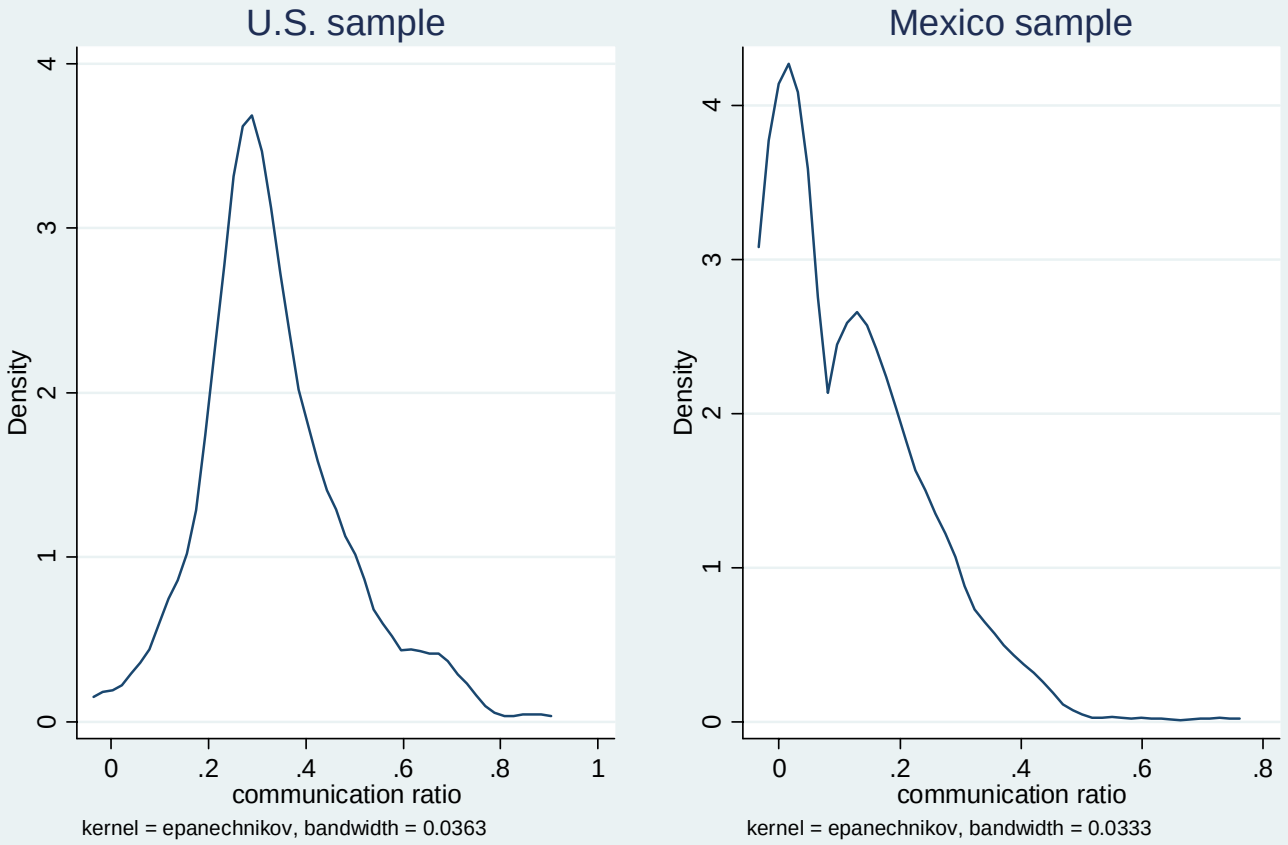
# network quesitons

[show page from survey]
[show sampling strategy in Mexico]
[show network]

Communication Frequency with Friends and Family



# Network Measure of Transnational Social Field



How much the U.S. based respondent misses friends and family in Mexico, % by column

| | gender | |
|---|---|---|
| | male | female |
| 1 A little | 12.24 | 2.00 |
| 2 | 4.08 | 4.00 |
| 3 | 16.33 | 26.00 |
| 4 | 20.41 | 18.00 |
| 5 “A lot ….like a a pain” | 46.94 | 50.00 |
| Total | 100.00 | 100.00 |
| | 101 | 389 |

Self-reported happiness by location, % by column

| how happy are you? | Location | |
|---|---|---|
| | Chapel Hill | Guanajuato |
| unhappy | 0.00 | 2.31 |
| a little happy | 15.84 | 8.48 |
| happy | 55.45 | 56.04 |
| very happy | 20.79 | 21.59 |
| extremely happy | 7.92 | 11.57 |
| Total | 100.00 | 100.00 |
| Sample size | 101 | 389 |

Where do you think you would be happiest? ( by location, column %)

| | R's Location | |
|---|---|---|
| | Chapel Hill | Guanajuato |
| in the U.S. | 25.74 | 4.63 |
| in Mexico | 48.51 | 88.69 |
| equal | 25.74 | 6.68 |
| | | |
| Total | 100.00 | 100.00 |

What is your opnion of U.S. culture? (by location, column %)

| | R's Location | |
|---|---|---|
| | Chapel Hill | Guanajuato |
| 1. I don't like it | 11.00 | 46.37 |
| 2. | 12.00 | 14.51 |
| 3. | 47.00 | 24.61 |
| 4. | 18.00 | 6.74 |
| 5. like it (me encanta | 12.00 | 7.77 |
| | | |
| Total | 100.00 | 100.00 |

If you could, would you choose to live permanently in the U.S.?

| | R's Location | |
|---|---|---|
| | Chapel Hill | Guanajuato |
| | | |
| No | 54.00 | 85.57 |
| yes | 46.00 | 14.43 |
| | | |
| Total | 100.00 | 100.00 |

```
. tab anyone_nc

do you know |
  anyone in |
        nc? |      Freq.     Percent        Cum.
------------+-----------------------------------
        yes |        190       48.72       48.72
         no |        200       51.28      100.00
------------+-----------------------------------
      Total |        390      100.00

. tab anyone_houston

do you know |
  anyone in |
   houston? |      Freq.     Percent        Cum.
------------+-----------------------------------
        yes |        135       35.53       35.53
         no |        245       64.47      100.00
------------+-----------------------------------
      Total |        380      100.00
```

**Table 6: R's self identification as Mexican (U.S. respondents only)**

| Degree of identification as Mexican | Percent |
|---|---|
| 25% | 1 |
| 50% | 6 |
| 75% | 11 |
| 100% | 82 |
| Total | 100 |
| Number of cases | (100) |

**Table 9: Summary statistics of variables used in the analysis**

| Variable | U.S. Sample (N) | U.S. Sample Mean | U.S. Sample Std. Dev. | Mexico Sample (N) | Mexico Sample Mean | Mexico Sample Std. Dev. |
|---|---|---|---|---|---|---|
| **Contacts' average of:** | | | | | | |
| 1. Desire for permanent residence ("zperm_us") | 188 | 0.469 | | 370 | 0.167 | |
| 2. Happier living in Mexico ("zhapmex") | 192 | 0.528 | | 370 | 0.845 | |
| 3. Opinion of U.S. culture ("zopinion_us") | 192 | 3.05 | 0.648 | 370 | 2.25 | 0.82 |
| 4. Transnational Communication Ratio ("TCR") | 198 | 0.337 | 0.161 | 407 | 0.118 | 0.137 |
| 5. Number of years in the U.S.: current migrants | 202 | 11.52 | 9.48 | | | |
| 6. Ln(years in the U.S.): current migrants | 202 | 2.11 | 1.05 | | | |
| 7. Number of years in the U.S.: returned migrants[1] | | | | 93 | 4.28 | 4.65 |
| 8. Ln(years in the U.S.): returned migrants[1] | | | | 93 | 1.26 | 0.93 |
| 9. Education: High school or higher | 202 | 0.534 | | 407 | 0.64 | |
| 10. Age | 189 | 36.4 | 12.30 | 391 | 39.1 | 16.28 |
| 11. Age: greater than 50 years old | 189 | .099 | | 391 | 0.24 | 0.430 |
| **Network roster count variables:** | | | | | | |
| 12. Number of other respondents who nominated R. | 202 | 3.06 | 4.67 | 407 | 2.24 | 2.72 |
| 13. Number of other respondents R nominated. | 198 | 3.96 | 3.20 | 407 | 2.96 | 2.98 |
| 14. # of nominations R made in the network roster. | 198 | 18.41 | 5.96 | 407 | 12.07 | 6.06 |
| 15. Female | 202 | 0.440 | | 407 | 0.558 | |
| English ability (excluded category: very good) | | | | | | |
| 17. Good | 194 | 0.155 | | 407 | 0.061 | |
| 18. Not good | 194 | 0.485 | | 407 | 0.204 | |
| 19. Not at all | 194 | 0.253 | | 407 | 0.722 | |

Notes: [1] There are 93 returned migrants in the data. The number of years in the U.S. is 0 for the non-migrants in the Mexico sample

**Table 10: Logit and Ordered Probit models of three measures of destination country incorporation**

| Model | Model 1 (Logit) | Model 2 (Logit) | Model 3 (Ordered Probit) |
|---|---|---|---|
| Variables | Desire for permanent residence | Happier living in Mexico | Opinion of U.S. culture |
| Contacts' average of: | | | |
| Desire for permanent residence ("zperm_us") | 1.036*** | | |
| | (0.367) | | |
| Happier living in Mexico ("zhapmex") | | 0.821** | |
| | | (0.395) | |
| Opinion of U.S. culture ("zopinion_us") | | | 0.183*** |
| | | | (0.0635) |
| Location: in Mexico | -1.313* | 4.078*** | -0.591* |
| | (0.722) | (0.683) | (0.313) |
| Transnational Communication Ratio ("TCR") | 0.286 | 2.794** | 0.198 |
| | (1.085) | (1.195) | (0.390) |
| Interaction of location=Mexico and TCR | 0.641 | -3.670** | 0.0452 |
| | (1.453) | (1.716) | (0.656) |
| Ln(years in the U.S.): current migrants | 0.405** | 0.320* | 0.179** |
| | (0.191) | (0.169) | (0.0846) |
| Ln(years in the U.S.): returned migrants | 0.913*** | -0.730*** | 0.648*** |
| | (0.193) | (0.203) | (0.0937) |
| Female | 0.355 | -0.263 | 0.191* |
| | (0.253) | (0.244) | (0.0980) |
| English ability (excluded category: very good) | | | |
| Good | -0.160 | -0.214 | 0.263 |
| | (0.618) | (0.556) | (0.244) |
| Not good | -0.491 | 0.886* | 0.0351 |
| | (0.574) | (0.494) | (0.193) |
| Not at all | -0.695 | 1.087** | -0.165 |
| | (0.592) | (0.528) | (0.203) |
| Education: High school or higher | -0.254 | 0.348 | 0.223** |
| | (0.252) | (0.260) | (0.109) |
| Age: greater than 50 years old | -0.406 | 0.429 | 0.148 |
| | (0.337) | (0.362) | (0.137) |
| Number of nominations in the survey | -0.0777** | 0.0299 | -0.0242** |
| | (0.0309) | (0.0319) | (0.0117) |
| Interaction of location=Mexico and nominations | 0.0593 | -0.0179 | 0.0520** |
| | (0.0677) | (0.0776) | (0.0240) |
| Constant | -0.518 | -3.341*** | |
| | (0.812) | (0.837) | |
| (N) | (609) | (609) | (609) |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
##(estimated cutpoints for the ordered probit in Model 3: 0.178, 0.425, 1.512, 2.060)