

Article

Incremental Recall: An Efficient Method for Estimating Egocentric Network Density

Chad A. Davis *  and Caimiao Liu 

Department of Psychology, University of Virginia, Charlottesville, VA 22903, USA; uxn6fy@virginia.edu

* Correspondence: knj5ft@virginia.edu

Abstract

Accurate estimation of network density is central to egocentric social network analysis, yet existing survey-based methods require researchers to balance accuracy against participant burden and systematic recall bias. Traditional approaches, such as fixed-list name generators, tend to overrepresent salient ties. Although the more recent random sampling method yields better accuracy, it relies on exhaustive free recall, which can be cognitively demanding and impractical for researchers. In this study, we introduce and evaluate an alternative approach—incremental recall—that structures alter nomination across relationship categories to improve coverage of differing tie strengths while reducing respondent burden. Using a large-scale Monte Carlo simulation encompassing over 9 million egocentric networks, we compare incremental recall against traditional fixed-list recall and random sampling across a wide range of network sizes, compositions, and recall bias assumptions. Results show that the incremental recall method consistently outperforms traditional fixed-list recall and performs comparably to or better than random sampling under unbiased and moderately biased recall conditions. Performance advantages persist even when respondents are unable to provide the full number of alters specified by design. We further validate these findings using empirical egocentric network data from 103 participants. Treating observed networks as proxy ground truths, empirical results closely mirror the simulation patterns, confirming the robustness of incremental recall under real-world reporting conditions. These findings demonstrate that incremental recall addresses a central practical challenge in egocentric social network research: balancing feasibility and accuracy in density estimation. The proposed method maintains strong performance while substantially reducing respondent burden and simplifying administration for applied studies. For researchers conducting large scale surveys where network density is one of several measures, incremental recall provides a practical and validated alternative to exhaustive recall that maintains robustness to realistic reporting biases.

Academic Editor: Francesco
Cauteruccio

Received: 1 February 2026

Revised: 14 February 2026

Accepted: 18 February 2026

Published: 2 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.**Keywords:** egocentric networks; network density; social network methodology; Monte Carlo simulation; survey-based network analysis

1. Introduction

An egocentric social network is defined as the set of individuals (the “alters”) directly connected to a focal person (the “ego”) [1]. The use of egocentric social networks has become increasingly popular across fields such as psychology, economics, information science, and public health [2–5]. From these networks, researchers can derive structural indicators such as size, homophily, and network density. In particular, density describes

the extent to which a person's contacts are also connected to one another, quantifying a network's cohesion and connectivity.

Having an accurate measure of density is crucial theoretically as it is linked to numerous important outcomes including norm enforcement, information circulation, and mobilization of support. For instance, high-density networks are associated with less loneliness and more acceptance of substance abuse and are more commonly found in interdependent cultures. In contrast, low-density networks predict access to diverse opportunities, novel information, and are more commonly found in independent cultures [6,7]. More broadly, accurate estimation of network density is foundational for social network analysis. Firstly, density provides a common scale for comparing networks that differ in size, composition, or context. Secondly, it serves as a structural baseline to facilitate interpretation of other network metrics such as centrality, which quantifies a node's importance in a network. Underestimation of density can artificially inflate a node's importance, while overestimation may compress centrality differences.

However, it is difficult to measure network density accurately [1,8]. The challenge arises because such measurement requires information on all possible ties among alters, and the number of ties increases exponentially with network size. For example, a 10-person network requires reports on 45 ties, whereas a 50-person network requires 1225. Collecting such data is both time-consuming and cognitively taxing for participants. In addition, self-reported networks are subject to systematic biases. Respondents tend to recall and report close friends more readily than weak ties [9]. Past research has found that up to 20% of alters can go forgotten, and such forgetting is especially pronounced for peripheral contacts [10,11]. This dual challenge, participant burden and recollection bias, has direct consequences for density estimates. In practice, density is often inflated because close friends are more likely to be reported and are also more likely to know each other compared to weaker ties. Using such biased estimates as predictors of outcomes raises concerns about the validity of inferences drawn from them.

To mitigate these challenges, researchers have developed several strategies to sample egocentric networks. The most common is the fixed-list name generator, in which participants nominate a set number of alters (typically 5–10) and provide tie information for each. This method is quick and easy to implement but systematically biased toward central figures in respondents' lives [11]. Other methods include instructed-recall sessions or cue-based recall (e.g., phonebook), which guide participants through a structured process to enumerate as many alters and alter–alter ties as possible [12–14]. While thorough, this approach is labor-intensive and impractical for large-scale studies where network measures are only one component. Graphical elicitation tools [15] can facilitate reporting but require specialized programming and may not be easy to implement in a survey-based study. More recently, random sampling methods have been proposed: participants generate a free list of alters (with no cap, encouraged to recall as exhaustively as possible) and then provide tie information for a randomly selected subset [16,17]. Prior work generally suggests that sampling 10 alters from such lists can yield accurate estimates of network characteristics, as free recall often captures both strong and weak ties. This method reduces centrality bias and better represents the composition of individuals' networks, leading to more valid density estimates. However, although less burdensome than instructed recall, it remains time-consuming and rests on the problematic assumption that free lists fairly represent different kinds of relationships, despite evidence that weaker ties are more readily forgotten.

In brief, there remains a lack of research on efficient approaches to measuring egocentric network density that balance accuracy, participant burden, and recall bias, particularly approaches that can be easily implemented in survey-based designs, which remain the most common method of data collection in the social sciences [18]. (We acknowledge that

when researchers are interested solely in estimating network density within a specific, restricted environment, such as a classroom or workplace, traditional fixed-list recall may be sufficient [19].)

Drawing on an analogy to stratified sampling, we propose a middle ground: incremental recall. Just as stratified sampling improves representativeness by dividing a population into meaningful subgroups before selection, our incremental recall method asks respondents to nominate only a few alters from three widely recognized relationship categories in the literature: close friends, casual friends, and acquaintances. The distinction between strong, moderate, and weak ties is foundational in social network research [20–24]. This general categorization appears cross-culturally, although the threshold definition for what qualifies as a truly close friend may vary [25]. Importantly, incremental recall does not impose fixed tie strength definitions. Rather, it leverages respondents' subjective relationship categorizations. This approach is supported by research showing that perceived network structure predicts psychosocial outcomes more strongly than objectively measured ties [26,27] (For a visual illustration, see Figure 1). To reduce participant burden, we limit nominations to 3–4 alters per category, yielding a total of 9–12 alters, a number comparable to existing methods. The design is intuitive for participants, reduces cognitive load, and ensures coverage across tie-strength categories. At the same time, we acknowledge that it remains subject to the same fundamental biases that prior approaches have struggled to address, that is, participants may select the most salient alters within each category. Nonetheless, if the incremental recall method is shown to be more robust to these biases in simulations and validation studies, it would represent a practical advance over existing approaches by offering a promising balance between accuracy and ease of implementation.

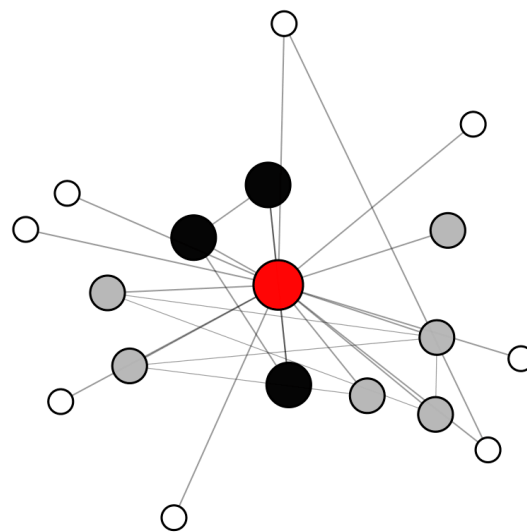


Figure 1. Example Ego-centric Network. Note: Each node represents either the ego (shown in red) or an alter, with node shading indicating their relationship to the ego: close friends (black), casual friends (gray), and acquaintances (white). Assuming this figure represents a participant's complete network, application of the three-alter incremental recall method would result in the nomination of all three close friends, three of the six casual friends, and three of the eight acquaintances.

Therefore, the present study evaluates the performance of the proposed incremental recall method against the traditional fixed-list recall method and the most recent random sampling method. We focus on traditional recall and random sampling as comparison points because they represent, respectively, the most widely used baseline in psychological research and the most recent methodological innovation with empirical validation [16]. To do so, we first use simulated networks to compare the proposed method against traditional fixed-list recall and random alter sampling under varying conditions of network size,

relationship composition, and intra- versus inter-category connectivity. Second, we test performance under realistic respondent biases, such as when participants recall only a subset of their actual network or disproportionately nominate salient alters within each category. Third, we validate the simulation findings using a real-world dataset, treating observed networks as proxy ground truths and applying each method to estimate density. Together, these analyses are intended to provide researchers with empirical guidance for selecting egocentric network density measures that balance accuracy, robustness to bias, and participant burden across diverse study contexts, with a particular focus on approaches that can be easily implemented in survey-based research, the dominant mode of data collection in the social sciences [18].

2. Materials and Methods

2.1. Simulation

We evaluated the performance of different ego-network sampling methods using a Monte Carlo simulation: traditional fixed-list nomination [4,19,28,29], random sampling from free recall [16,17,30,31], and the proposed incremental recall methods. Across a range of network sizes, network compositions, and recall bias assumptions, we assessed how accurately each method recovered the density of an underlying “true” network.

We simulated true population egocentric networks by varying four factors: (a) total network size (the number of individuals/alters in an ego’s network in addition to themselves), (b) the distribution of alters across relationship categories (i.e., varying how many alters are in each of the categories in the ego’s network: close, casual, and acquaintances), and the probabilities of (c) intra-category and (d) inter-category ties. These latter two factors reflect the likelihood that alters in the same relationship category will be connected to each other and the likelihood that alters in different relationship categories will be connected respectively. For each condition, 500 independent networks were generated, yielding a total of 661,500 population networks. Because intra- and inter-category ties were specified probabilistically, networks with identical sizes and alter distributions differed in their realized structures. We then applied three sampling approaches under a range of recall bias assumptions, resulting in 14 sampling conditions. Across all conditions, this procedure produced 9,261,000 sampled egocentric networks.

2.2. Network Generation

Simulated networks consisted of one ego node and a varying number of alters. Total network size ranged from 21 to 151 nodes (21, 31, 41, 51, 71, 101, and 151). Ego size was fixed at one node, with the remaining nodes representing alters. Alters were assigned to one of three relationship categories: close friends, casual friends, and acquaintances. To capture heterogeneity in network composition, we varied nine proportions of category configurations, ranging from networks dominated by close ties to those dominated by weaker ties. In addition, we included more extreme network compositions, such as networks containing no close friends or no acquaintances, to evaluate method performance under rare but theoretically informative conditions. Category sizes were computed by multiplying the specified proportions by the total number of alters and adjusting for rounding so that group sizes summed to the correct total.

2.3. Tie Formation

Alter–alter ties were generated probabilistically as a function of relationship category membership. For each simulated network, we specified a matrix of tie probabilities governing both intra- and inter-category connections varying across high, moderate and low connectivity to represent a diverse range of possible network compositions. Connectivity

parameters were chosen to reflect theoretically plausible gradients in tie strength, with higher probabilities among close friends and lower probabilities among acquaintances. By default, the ego node was simulated to be connected to every alter. Simulation parameters are summarized in Table 1.

Table 1. Simulation Parameters for Population Egocentric Networks.

Parameter	Values
Network size (including ego)	21, 31, 41, 51, 71, 101, 151
Alter composition (close, casual, acquaintance)	[2/3, 1/3, 0], [2/3, 1/6, 1/6], [1/2, 1/3, 1/6], [1/2, 1/6, 1/3], [1/3, 1/3, 1/3], [1/6, 1/2, 1/3], [1/6, 1/3, 1/2], [0, 2/3, 1/3]
Within-category tie probability	[0.5, 0.3, 0.1], [0.7, 0.5, 0.3]
Between-category tie probability	[0.7, 0.6, 0.5], [0.5, 0.4, 0.3], [0.3, 0.2, 0.1], [0.2, 0.1, 0.1]

Note. Alter composition vectors indicate the proportion of alters categorized as close friends, casual friends, and acquaintances, respectively. Within-category tie probabilities specify the probability that two alters belonging to the same relationship category are connected. Between-category tie probabilities specify the probability that two alters from different categories are connected. For example, a within-category vector of [0.5, 0.3, 0.1] indicates tie probabilities of 0.5 among close friends, 0.3 among casual friends, and 0.1 among acquaintances. A between-category vector of [0.5, 0.4, 0.3] indicates tie probabilities of 0.5 between close and casual friends, 0.4 between close friends and acquaintances, and 0.3 between casual friends and acquaintances.

2.4. Sampling Methods

For each complete network, we treated the network as a participant's true social network and simulated the process by which the participant self-reported their network under different survey procedures. Under each procedure, only the subset of alters reported by the participant and the ties among those alters were observed (see Figure 2 for an illustrative example).

Traditional Fixed-List Recall Nomination. The traditional fixed-list recall condition was designed to approximate the most common way social science researchers have generated ego-centric network data, where respondents are asked to nominate a fixed number of alters without structured recall aids. We examined three list lengths: five alters (Cap-5), nine alters (Cap-9), and 12 alters (Cap-12). Five is among the most commonly used number of alters in egocentric network studies, with typical ranges extending up to 10. We included Cap-9 and Cap-12 to enable direct comparison between the traditional fixed-list approach and the proposed incremental recall method (introduced below), as these list lengths correspond to the number of alters elicited by the proposed method. Consistent with evidence that such instruments disproportionately elicit salient or strong ties, alters were sampled exclusively from the close-friend category. Alters were sampled uniformly at random from the close-friend subset; if no alters were present in that category, sampling proceeded from the casual-friend category.

This procedure assumes that recall is constrained primarily by salience, resulting in samples that overrepresent densely interconnected alters and are therefore expected to yield upwardly biased density estimates.

Random Sampling from Free Recall. In the random sampling condition, respondents were assumed to first generate a free list of alters and then provide tie information for a random subset of 10 alters. We examined three recall assumptions to model varying degrees of recall completeness:

- Unbiased recall (100%), in which all alters in the true network were recalled.
- 75% truncation, in which only the top 75% of alters ranked by degree (interconnectedness) in the true network were recalled.
- 50% truncation, in which recall was limited to the top 50% most connected alters.

In all cases, 10 alters (as recommended in the previous literature [16]) were selected at random from the recalled set. In the truncated conditions, random sampling was

unbiased with respect to the recalled network but operated on a systematically incomplete representation of the true network, thereby overrepresenting more embedded alters and potentially inflating density estimates.

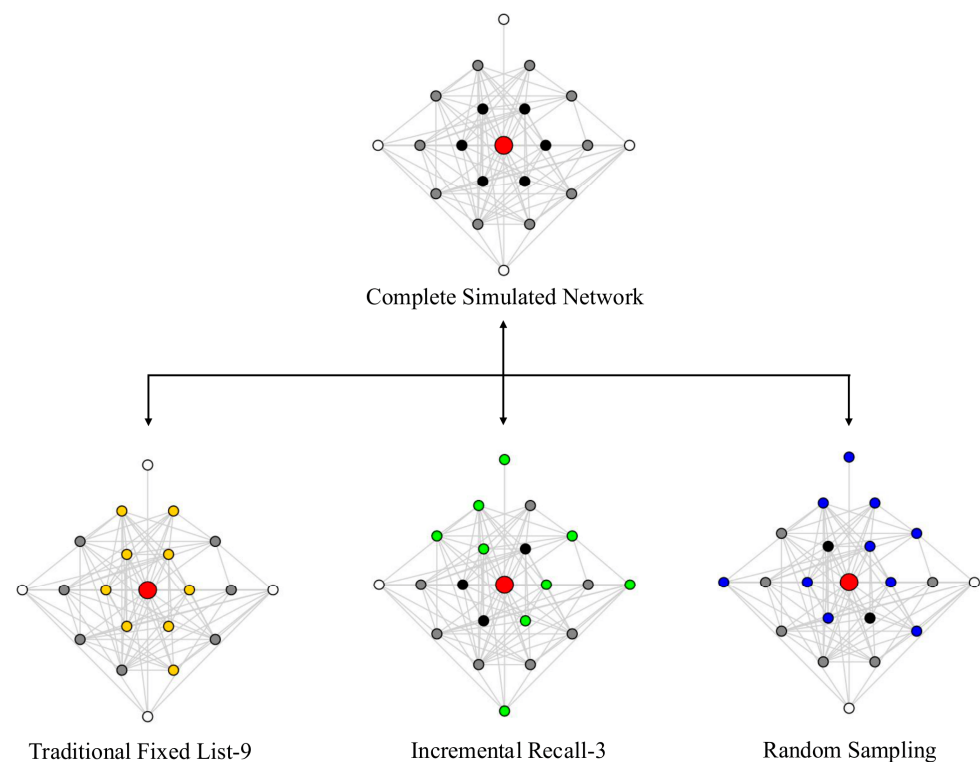


Figure 2. Illustration of Different Sampling Methods from True Ego-Centric Network. Note: The top panel displays the complete simulated egocentric network ($N = 21$), consisting of one ego (red) and 20 alters distributed across three relationship categories: close friends (black), casual friends (gray), and acquaintances (white). The lower panels illustrate how versions of the three sampling strategies previously discussed select alters from the same underlying network. Highlighted nodes indicate alters that would be nominated under each method: Traditional Fixed List-9 (yellow), Incremental Recall-3 (green), and Random Sampling (blue).

Incremental Recall (Proposed Method). The incremental recall method simulated a structured sampling procedure in which respondents reported alters separately within each relationship category (close friends, casual friends, acquaintances). Two alter limits were examined: up to three per category (Incremental-3; 9 total) and up to four per category (Incremental-4; 12 total). These limits were chosen to bracket the commonly recommended benchmark of 10 alters used in random sampling. This then allows us to assess whether sampling slightly fewer alters compromises accuracy and whether sampling modestly more alters provides additional gains. This design enables direct comparison with existing methods while informing practical decisions about alter limits in applied research.

Although incremental recall is designed to improve coverage of weaker ties, it may still reflect systematic recall biases. Specifically, it may overrepresent highly salient alters within each relationship category, as in fixed-list nomination, or biases associated with incomplete recall of the entire network in which weaker or more peripheral ties are less likely to be remembered prior to sampling, as in random sampling from free recall.

As such, we examined three incremental recall variants:

- Unbiased incremental recall, in which all alters within each category were equally likely to be nominated.

- Preferential recall within category, in which reporting within each category was restricted to the most connected alters (by degree in the full network), with higher-degree alters selected first. This mechanism mirrors the bias of fixed-list nomination but applies it separately within each relationship category. As a result, incremental recall may still overrepresent well-connected alters, particularly within the close-friend and casual-friend categories.
- Restricted recall prior to categorization, in which recall was incomplete before categorization. Only the top 75% or top 50% of alters by degree were assumed to be recalled, after which alters were assigned to categories and nominations were drawn from the available alters within each category. This procedure models a situation in which respondents are unable to recall their full network but nevertheless attempt to distribute nominations across categories. In this case, incremental recall inherits the bias of random sampling from an incomplete recall set while additionally constraining selection by category.

2.5. Density Estimation

For all sampling procedures, network density was calculated as the ratio of existing ties among alters to the total possible number of ties and was computed from the induced subgraph consisting of the ego and the sampled alters. These estimates were compared to the density of the full simulated network to assess bias and variability in density recovery under each sampling method and recall assumption.

To quantify systematic bias in density recovery, we calculated **relative bias** for each sampled network as:

$$\text{Relative Bias} = \frac{\bar{D} - D}{D}$$

where $\bar{D}$ denotes the estimated density obtained from a given sampling method and D denotes the true density of the full simulated network. Relative bias provides a scale-free measure of estimation error, facilitating comparison across networks that differ in size and baseline density.

To assess overall estimation accuracy, we additionally computed the **mean squared error (MSE)** for each method, defined as the squared difference between estimated and true density $(\bar{D} - D)^2$.

Whereas relative bias captures directional tendencies toward over- or underestimation, MSE incorporates both bias and variability, penalizing large deviations more heavily. As such, it is a comprehensive index of performance. For each combination of network parameters and sampling conditions, relative bias and squared error were computed at the level of individual simulation iterations. Results were then aggregated across iterations, and we report the mean, median, and standard deviation of relative bias, as well as the MSE, for each sampling method and recall assumption.

2.6. Empirical Validation

To validate the simulation results, egocentric network data were collected using an online survey administered in Qualtrics. The survey was designed to facilitate recall of a social network that was as complete as possible, with particular emphasis on encouraging the inclusion of weaker and more peripheral social ties.

Participants were 103 undergraduate students (40.8% male, 59.2% female) recruited from a large university participant pool and compensated with course credit. All participants provided informed consent prior to participation, and study procedures were approved by the university's institutional review board.

Participants were first presented with six structured social prompting questions intended to broaden recall beyond close relationships. These prompts explicitly referred to a range of interaction contexts and tie strengths (e.g., people participants occasionally greet, interact with casually, or encounter in routine social settings), encouraging participants to consider both strongly and weakly connected alters. The full set of prompts is provided in Appendix A.

Following this prompting, participants were asked to report up to 100 alters. No minimum number of alters was imposed, and participants were instructed to list as many individuals as they could recall from their social environment. For each alter, participants classified the relationship as one of three categories: close friend, casual friend, or acquaintance.

Data on alter–alter ties were then collected. For each alter, participants were presented with a list of all other reported alters and indicated, using a binary response (yes/no), whether the two individuals were friends to one another. This procedure yielded a complete adjacency matrix for the reported egocentric network. The number of alters reported per participant ranged from 2 to 100 ($M = 21$, $SD = 15$), reflecting substantial variability in network size.

For the purposes of empirical validation, the networks reported through this procedure were treated as participants' complete egocentric networks and used as proxy ground truths. Although self-reported networks are inevitably subject to recall limitations, the combination of multiple structured prompts, unlimited alter nominations, and exhaustive alter–alter tie reporting was intended to approximate as complete a representation of participants' social networks as is feasible in survey research.

To evaluate the performance of different ego-network sampling strategies under real-world recall conditions, we applied the same sampling procedures examined in the simulation study to each observed network and again calculated aggregated results for MSE and Relative Bias by method.

3. Results

3.1. Simulation

Across 9,261,000 simulated egocentric networks, we compared the performance of the proposed incremental recall methods against traditional fixed-list recall and random sampling approaches across a range of recall assumptions (see Table 2).

Table 2. Performance Metrics by Sampling Method and Bias Condition.

Method	MSE	Median Relative Bias	Mean Relative Bias	SD _{Mean Relative Bias}
Incremental-4 (unbiased)	0.043	0.10	0.25	0.61
Random sampling (unbiased)	0.048	0.18	0.32	0.61
Incremental-3 (unbiased)	0.049	0.17	0.33	0.65
Incremental-4 (75% truncation)	0.066	0.32	0.47	0.68
Random sampling (75% truncation)	0.072	0.40	0.53	0.66
Incremental-4 (preferential recall)	0.072	0.38	0.56	0.72
Incremental-3 (75% truncation)	0.074	0.37	0.53	0.71
Incremental 3 (preferential recall)	0.087	0.47	0.65	0.76
Traditional fixed-list (Cap-12)	0.090	0.58	0.68	0.49
Traditional fixed-list (Cap-9)	0.097	0.61	0.72	0.52
Random sampling (50% truncation)	0.100	0.58	0.71	0.68
Incremental-4 (50% truncation)	0.120	0.65	0.78	0.74
Traditional fixed-list (Cap-5)	0.124	0.70	0.84	0.60
Incremental-3 (50% truncation)	0.133	0.69	0.83	0.77

Note. Methods ordered by ascending MSE. Relative bias is calculated as the difference between the estimated density and the true density divided by the true density. MSE is mean squared error, calculated as the squared difference between estimated and true density.

Under unbiased recall assumptions, Incremental-4 exhibited the strongest performance, placing it at the top of all evaluated conditions. This was followed by Incremental-3 and random sampling across the full dataset, which showed comparable performances. When recall bias was introduced, including preferential recall and restricted recall, the incremental methods, particularly Incremental-4, continued to perform robustly. Under 75% truncation, Incremental-4 achieved the best overall performance, followed by Incremental-3 under preferential recall bias and random sampling under 75% truncation. Across unbiased and moderately biased conditions, Incremental-4 consistently ranked among the top-performing methods. Incremental-3 exhibited comparable performance as random sampling.

Performance declined substantially for methods in the lower half of the comparison. This group consisted primarily of traditional fixed-list recall methods (Cap-5, Cap-9, and Cap-12) and conditions assuming extreme recall bias (50% truncation). Although performance improved with longer lists for the traditional methods, even Cap-12 exhibited substantial bias (median relative bias = 58%; mean relative bias = 68%).

Under 50% truncation, random sampling appeared to outperform the incremental methods; however, this pattern reflected limitations in data availability rather than estimation accuracy. Under this extreme bias condition, recalled networks often contained fewer alters than required by the incremental designs. For example, in a network with 20 alters, 50% truncation yields only 10 recalled alters, fewer than required for Incremental-4. Consequently, only 6% of simulated networks contained sufficient alters for Incremental-4, and only 9% for Incremental-3. Density estimates were nevertheless computed using all available information.

When calculations were restricted to networks in which the incremental methods successfully sampled their intended number of alters (see Table 3), their performance under truncation bias improved markedly and exceeded that of competing approaches. Notably, even when the incremental methods failed to reach their nominal sample sizes, their aggregated performance remained strong. Incremental-4 under unbiased recall and Incremental-3 and Incremental-4 under 75% truncation ranked among the best-performing conditions overall, despite achieving full sampling in only 59%, 39%, and 31% of networks, respectively.

Table 3. Comparison of Sampling Methods Under Complete Sampling.

Method	MSE	Complete Sampling (%)	Median Relative Bias	Mean Relative Bias	SD _{Mean Relative Bias}
Incremental-4 (50% truncation)	0.034	5.6	−0.11	0.00	0.4
Incremental-3 (50% truncation)	0.039	8.1	−0.02	0.12	0.49
Incremental-4 (unbiased)	0.039	58.7	−0.04	0.10	0.5
Incremental-4 (75% truncation)	0.040	31.9	−0.02	0.12	0.5
Incremental-3 (unbiased)	0.040	66.7	0.01	0.15	0.52
Incremental-3 (75% truncation)	0.043	39.0	0.04	0.19	0.54
Random sampling (unbiased)	0.048	100.0	0.18	0.32	0.61
Incremental-4 (preferential recall)	0.053	58.7	0.21	0.36	0.57
Incremental-3 (preferential recall)	0.062	66.7	0.28	0.43	0.59
Random sampling (50% truncation)	0.072	100.0	0.40	0.53	0.66
Traditional fixed-list (Cap-12)	0.095	74.6	0.64	0.74	0.52
Random sampling (75% truncation)	0.100	85.7	0.58	0.71	0.68
Traditional fixed-list (Cap-9)	0.101	84.1	0.65	0.76	0.53
Traditional fixed-list (Cap-5)	0.125	96.8	0.71	0.85	0.6

Note. Methods ordered by ascending MSE. Complete sampling (%) indicates the percentage of simulated networks in which the method obtained the intended number of alters specified by design (e.g., 9 alters for Incremental-3, 12 alters for Incremental-4). In incomplete cases, density estimates were computed using all available alter–alter information.

3.2. Empirical Results

Empirical results closely matched the simulation findings. Both unbiased random sampling and Incremental-4 produced the most accurate density estimates, with near-zero median and mean relative bias and the lowest mean squared error (Table 4). Incremental-3 under unbiased recall also showed comparable accuracy but with slightly higher MSE and variability. When recall bias was introduced, performance for all methods was reduced. Under both 75% and 50% truncation, Incremental-3 and Incremental-4 exhibited lower bias and error than the corresponding random sampling method.

Traditional fixed-list (Cap-12) had performance only slightly worse than the top unbiased methods, but had substantially greater variance. This increase in variance is reflected in both the Traditional fixed-list (Cap-5) and the Traditional fixed-list (Cap-9), as well as the Incremental-3 and -4 (preferential recall) methods. This pattern suggests that when recall is driven by salience rather than coverage, density estimates become highly dependent on the particular subset of alters nominated, resulting in unstable and less reliable estimates (see Figure 3 for an illustration of sampling methods performance across conditions).

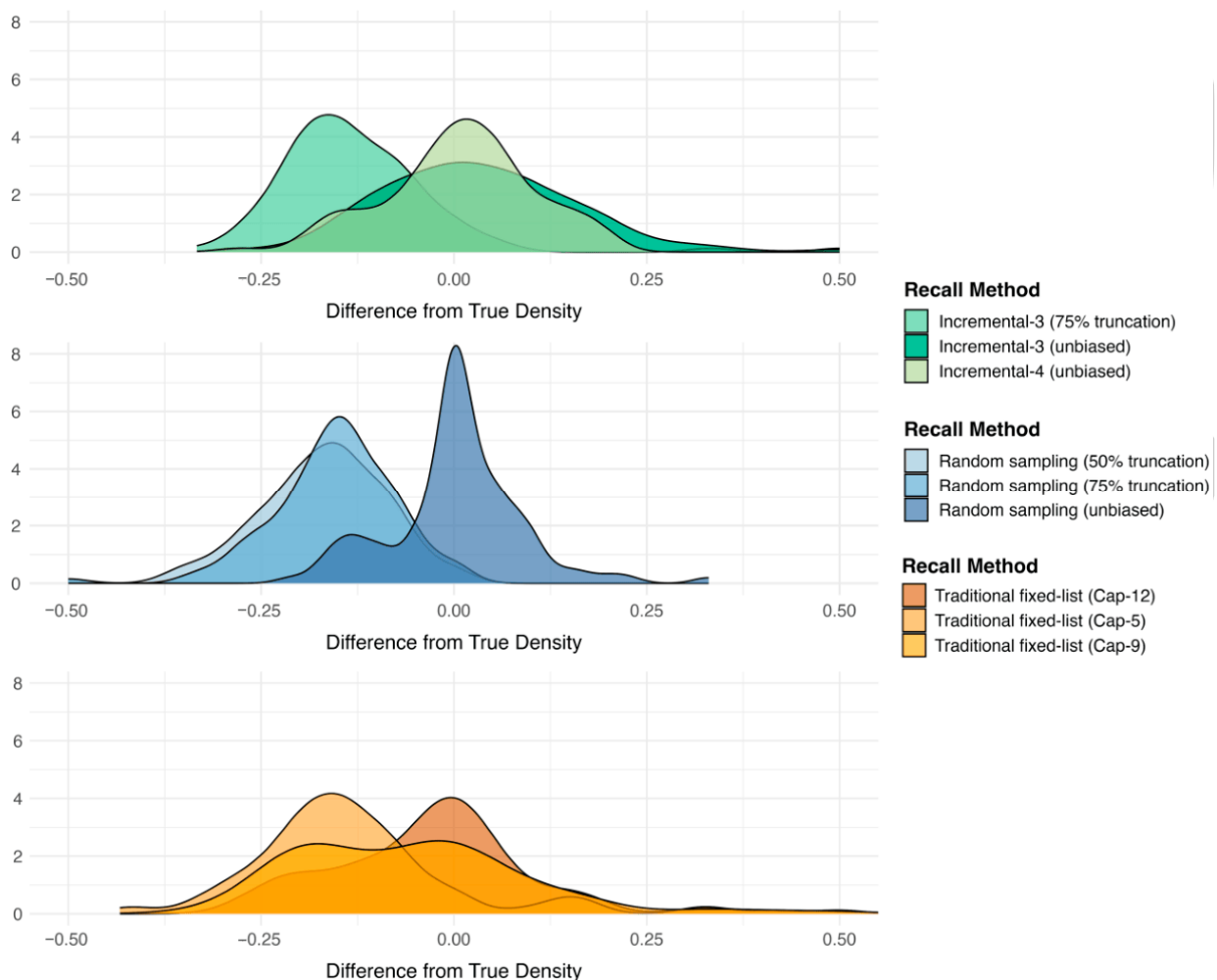


Figure 3. Accuracy of Network Density Estimates by Recall Method. Note. Density plots of mean and standard deviation of raw bias (estimated density—true density) by method. Only the top three Incremental Recall Methods are shown for ease of viewing.

Table 4. Comparison of Sampling Methods with Empirical Networks.

Method	MSE	Median Relative Bias	Mean Relative Bias	SD _{Mean Relative Bias}
Random sampling (unbiased)	0.007	0.00	0.00	0.11
Incremental-4 (unbiased)	0.009	−0.01	0.00	0.18
Incremental-3 (unbiased)	0.017	−0.02	−0.04	0.18
Traditional fixed-list (Cap-12)	0.024	0.02	0.05	0.39
Incremental-3 (75% truncation)	0.028	0.17	0.16	0.18
Incremental-4 (75% truncation)	0.028	0.17	0.17	0.21
Incremental 4 (preferential recall)	0.029	0.17	0.22	0.45
Random Sampling (75% truncation)	0.030	0.18	0.21	0.16
Traditional fixed-list (Cap-9)	0.031	0.06	0.09	0.40
Incremental 3 (preferential recall)	0.033	0.18	0.24	0.45
Incremental-4 (50% truncation)	0.034	0.19	0.20	0.31
Incremental-3 (50% truncation)	0.036	0.19	0.22	0.32
Random sampling (50% truncation)	0.037	0.20	0.24	0.18
Traditional fixed-list (Cap-5)	0.043	0.17	0.20	0.68

Note. Methods ordered by ascending MSE.

4. Discussion

As a core structural property of egocentric networks, density has been widely linked to social support, information flow, and well-being across the social sciences. Yet measuring density reliably has long required researchers to balance accuracy against participant burden and systematic recall bias. The present findings address this challenge by providing evidence that incremental recall offers a robust approach to density estimation. Specifically, Incremental-4, which samples four alters from each relationship category, showed the strongest performance under unbiased recall conditions. Importantly, its advantages persisted under biased recall scenarios, including restricted recall (i.e., omission of peripheral ties) and preferential recall (i.e., prioritization of highly interconnected alters). Meanwhile, Incremental-3 frequently performed comparably to random sampling, which involves first eliciting a complete free list of alters and then randomly selecting 10 for detailed reporting. Across both simulations and empirical validation, the incremental recall methods consistently outperformed traditional fixed-list recall and performed at least as well as, and often better than, random sampling.

Complete and unbiased recall is unlikely in practice; therefore, rather than assuming perfect reporting, our simulations explicitly modeled realistic recall biases and evaluated how different methods perform under these conditions. Under these biased conditions, Incremental-4 showed superior performance, while Incremental-3 reliably outperformed traditional fixed-list recall and performed comparably to random sampling. In the case of empirical validation, under 75% truncation, Incremental-3 actually performed better than random sampling. This pattern suggests that even with fewer total sampled alters, the Incremental-3 method can yield competitive density estimates to random sampling methods.

Another particularly important finding is the robustness of the incremental recall method when participants were unable to provide the full number of alters specified by design (e.g., being unable to sample the full 12 alters for Incremental-4). This occurred both when recall was incomplete (e.g., under 75% truncation) and when networks exhibited more extreme compositions, such as the absence of close friends or acquaintances. Despite these structural disadvantages, the incremental methods consistently outperformed traditional fixed-list recall and performed at least as well as random sampling, with Incremental-4 often yielding the most accurate estimates. This robustness was further confirmed by the empirical validation. Although the empirical sample included highly diverse network compositions (including networks with as few as two alters) the incremental recall methods

continued to perform well, even under restricted recall and preferential recall bias. For applied researchers, this finding is consequential. It indicates that the method remains effective even when some categories are sparsely populated or left unfilled, alleviating concerns that incomplete category-level nominations will invalidate density estimates.

The only condition under which incremental recall performed poorly was when we examined the most extreme recall constraint: the 50% truncation. Importantly, this condition represents a highly conservative stress test rather than a realistic expectation. Prior research indicates that recall omissions are typically closer to 20% and disproportionately affect weaker ties [10,11]. Under 50% truncation, more than 90% of simulated networks failed to contain enough recalled alters to meet the sampling requirements of the incremental method. Although method performance seemed to be the best for Incremental-3 and -4 (50% truncation) when only considering the cases where the requirements were met, no reliable conclusions about method performance under this bias condition should be drawn as it is likely only a statistical artifact of aggregating over such a small subset. However, in the empirical validation under the same 50% truncation assumption, random sampling's performance suffered equally to Incremental-3 and Incremental-4. Together, these findings suggest that extreme truncation undermines performance across methods rather than uniquely disadvantaging incremental recall. For most populations and research contexts in which recall bias is less severe, the proposed methods are expected to perform well and reliably.

4.1. Practical Implications

The incremental recall method also offers a substantial practical advantage by reducing cognitive burden. Unlike random sampling approaches, which require participants to first generate an exhaustive free list of alters, incremental recall avoids full network enumeration. Although Incremental-4 requires reporting two additional alters relative to the common 10-alter random sampling benchmark, this modest increase is likely far less burdensome than first listing an entire social network. Importantly, Incremental-3, which requires only nine alters, already achieves performance comparable to random sampling under similar bias assumptions, while Incremental-4 leverages minimal additional effort to achieve meaningfully better accuracy.

4.2. Theoretical Implications

Beyond methodological considerations, these findings have implications for theory development in egocentric social network research. Weak ties have long been theorized to provide access to novel information, opportunities, and social mobility [32], yet they are systematically underrepresented in egocentric network surveys. Traditional name generators disproportionately capture strong, salient relationships while neglecting weaker, more peripheral connections, which may matter most for outcomes like information diffusion and structural advantage.

This measurement bias has theoretical consequences. Much of the existing literature linking network density to outcomes such as social support and belonging may reflect only the density of close social circles rather than the full network. By improving representation of ties across relationship categories, incremental recall enables more comprehensive estimation of egocentric density and allows for more precise tests of foundational network theories.

4.3. Limitations and Future Directions

Although the present study combines large-scale simulations with empirical validation, several limitations warrant consideration. Firstly, while simulation parameters were chosen to reflect theoretically plausible ranges in network size, composition, and

interconnectedness, real-world networks may exhibit structural features not fully captured in our design (e.g., temporal dynamics). Accordingly, the simulations should be interpreted as systematic approximations rather than complete representations of social complexity.

Secondly, egocentric network data are inherently based on self-report, where all alter–alter ties are reported by the ego. Respondents’ perceptions of network structure may diverge from objective reality. Specifically, the effectiveness of the incremental recall approach may depend in part on how respondents classify alters into relationship categories. Individual differences or cultural variation in category classification could influence density estimates in applied settings. Although these limitations reflect broader constraints of self-reported egocentric network research, future research should incorporate externally validated network data to assess the absolute accuracy of the proposed method. Other potential directions may include testing how people’s subjective definition of each relationship category may affect their classification of alters and how this then influences density estimation.

Lastly, the empirical validation relied on proxy to true network density. Although the empirical study was designed to maximize recall completeness, these networks cannot be considered objective ground truth in a strict sense. The sample’s modest size and relative homogeneity in age and social context of the undergraduate population further limits the findings’ generalizability. Yet, the empirical component was not intended to provide population-level density estimates, but rather to test whether performance patterns observed in simulation would replicate under real-world reporting conditions. Future research should evaluate incremental recall in more diverse populations (such as older adults and non-Western cultural contexts) where network size and composition may differ substantially from undergraduate settings. Additional evaluations of group differences, such as gender, may also provide more insight into the effectiveness of the proposed methods across contexts.

5. Conclusions

Taken together, these findings suggest that while no egocentric network method is immune to recall bias, the incremental recall approach offers a particularly favorable trade-off between feasibility and accuracy. For researchers seeking reliable density estimates in large-scale or multi-measure studies, the method provides a practical alternative that reduces participant burden while remaining robust to realistic recall limitations.

Author Contributions: Conceptualization, C.L.; Methodology, C.A.D. and C.L.; Software, C.A.D.; Formal analysis, C.A.D.; Data curation, C.L.; Writing—original draft, C.A.D. and C.L.; Writing—review & editing, C.A.D. and C.L.; Visualization, C.A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study’s protocol was reviewed and approved by the Institutional Review Board at the University of Virginia (IRB Protocol #4638, expiration on 10 September 2026).

Informed Consent Statement: Informed consent was obtained for all participants, prior to their participation in the study.

Data Availability Statement: All data and additionally materials are available at <https://github.com/Chad-A-Davis/Incremental-Recall>, accessed on 31 January 2025.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Egocentric Social Network Alter Nomination Prompts

1. Who are the people you turn to for personal or emotional support (e.g., discussing important matters, seeking advice, sharing personal news)?
2. Who are the people you interact with on a regular basis for any purpose (e.g., family, classmates, co-workers, club members)?
3. Who are the people you occasionally chat with, greet, or see socially (e.g., neighbors, ‘friends of friends’), even if you don’t share personal matters with them?
4. Who do you see around campus/work/residence often enough to know by name but rarely spend extended time with?
5. Who are the people you interact with on social media, gaming platforms, or online forums regularly or occasionally?
6. Are there any old friends, former classmates, or ex-colleagues you still keep in touch with occasionally?

References

1. Marsden, P.V. Network Data and Measurement. *Annu. Rev. Sociol.* **1990**, *16*, 435–463. [\[CrossRef\]](#)
2. Borgatti, S.P.; Mehra, A.; Brass, D.J.; Labianca, G. Network Analysis in the Social Sciences. *Science* **2009**, *323*, 892–895. [\[CrossRef\]](#)
3. Burt, R.S.; Kilduff, M.; Tasselli, S. Social Network Analysis: Foundations and Frontiers on Advantage. *Annu. Rev. Psychol.* **2013**, *64*, 527–547. [\[CrossRef\]](#)
4. Montgomery, S.C.; Donnelly, M.; Bhatnagar, P.; Carlin, A.; Kee, F.; Hunter, R.F. Peer social network processes and adolescent health behaviors: A systematic review. *Prev. Med.* **2020**, *130*, 105900. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Smith, E.B.; Brands, R.A.; Brashears, M.E.; Kleinbaum, A.M. Social Networks and Cognition. *Annu. Rev. Sociol.* **2020**, *46*, 159–174. [\[CrossRef\]](#)
6. Burt, R.S. Network Disadvantaged Entrepreneurs: Density, Hierarchy, and Success in China and the West. *Entrep. Theory Pract.* **2019**, *43*, 19–50.
7. Coleman, J.S. Social Capital in the Creation of Human Capital. *Am. J. Sociol.* **1988**, *94*, S95–S120. [\[CrossRef\]](#)
8. Redhead, D.; McElreath, R.; Ross, C.T. Reliable network inference from unreliable data: A tutorial on latent network modeling using STRAND. *Psychol. Methods* **2024**, *29*, 1100–1122. [\[CrossRef\]](#)
9. Brewer, D.; Rinaldi, G.; Mogoutov, A.; Valente, T. A Quantitative Review of Associative Patterns in the Recall of Persons. *J. Soc. Struct.* **2005**, *6*. Available online: https://www.cmu.edu/joss/content/articles/volume6/Brewer/index_new.html (accessed on 1 February 2026).
10. Brewer, D.D. Forgetting in the recall-based elicitation of personal and social networks. *Soc. Netw.* **2000**, *22*, 29–43. [\[CrossRef\]](#)
11. Marin, A. Are respondents more likely to list alters with certain characteristics? *Soc. Netw.* **2004**, *26*, 289–307. [\[CrossRef\]](#)
12. Brewer, D.D. Supplementary Interviewing Techniques to Maximize Output in Free Listing Tasks. *Field Methods* **2002**, *14*, 108–118. [\[CrossRef\]](#)
13. Gerich, J.; Lehner, R. Collection of Ego-Centered Network Data with Computer-Assisted Interviews. *Methodology* **2006**, *2*, 7–15.
14. Hsieh, Y.P. Check the phone book: Testing information and communication technology (ICT) recall aids for personal network surveys. *Soc. Netw.* **2015**, *41*, 101–112. [\[CrossRef\]](#)
15. Stark, T.H.; Krosnick, J.A. GENSI: A new graphical tool to collect ego-centered network data. *Soc. Netw.* **2017**, *48*, 36–45. [\[CrossRef\]](#)
16. Peng, S.; Roth, A.R.; Perry, B.L. Random sampling of alters from networks: A promising direction in egocentric network research. *Soc. Netw.* **2023**, *72*, 52–58. [\[CrossRef\]](#)
17. Stadel, M.; Stulp, G. Balancing bias and burden in personal network studies. *Soc. Netw.* **2022**, *70*, 16–24. [\[CrossRef\]](#)
18. Giner-Sorolla, R. Changing Practices and Priorities in Social Psychological Research Methods and Reporting. In *Handbook of Social Psychology*, 6th ed.; Gilbert, D., Fiske, S., Finkel, E., Mendes, W., Eds.; Situational Press: London, UK, 2025.
19. Merluzzi, J.; Burt, R.S. How many names are enough? Identifying network effects with the least set of listed contacts. *Soc. Netw.* **2013**, *35*, 331–337. [\[CrossRef\]](#)
20. Hall, J.A. How many hours does it take to make a friend? *J. Soc. Pers. Relatsh.* **2019**, *36*, 1278–1296. [\[CrossRef\]](#)
21. Hays, R.B. The Day-to-Day Functioning of Close versus Casual Friendships. *J. Soc. Pers. Relatsh.* **1989**, *6*, 21–37. [\[CrossRef\]](#)
22. Mendelson, M.J.; Aboud, F.E. Measuring friendship quality in late adolescents and young adults: McGill Friendship Questionnaires. *Can. J. Behav. Sci.* **1999**, *31*, 130–132. [\[CrossRef\]](#)
23. Bryant, E.M.; Marmo, J. The rules of Facebook friendship: A two-stage examination of interaction rules in close, casual, and acquaintance friendships. *J. Soc. Pers. Relatsh.* **2012**, *29*, 1013–1035. [\[CrossRef\]](#)

24. Shetty, N.P.; Muniyal, B.; Maben, L.M.; Jayaraj, R.; Saxena, S. Dynamic Twitter friend grouping based on similarity, interaction, and trust to account for ever-evolving relationships. *IET Commun.* **2024**, *18*, 1018–1048.
25. Apostolou, M.; Sullman, M.; Ayers, J.D.; Błachnio, A.; Choubisa, R.; Gadelrab, H.F.; Hill, T.; Kamble, S.; Lisun, Y.; Manrique-Millones, D.; et al. Preferences for Closeness and Endurance in Friendship: A Cross-Cultural Investigation. *Cross-Cult. Res.* **2025**, *59*, 455–480.
26. Cornwell, E.Y.; Waite, L.J. Social disconnectedness, perceived isolation, and health among older adults. *J. Health Soc. Behav.* **2009**, *50*, 31–48.
27. Brands, R.A. Cognitive social structures in social network research: A review. *J. Organ. Behav.* **2013**, *34*, S82–S103. [[CrossRef](#)]
28. Burt, R.S. Network items and the General Social Survey. *Soc. Netw.* **1984**, *6*, 293–339. [[CrossRef](#)]
29. Suzman, R. The National Social Life, Health, and Aging Project: An introduction. *J. Gerontol. B Psychol. Sci. Soc. Sci.* **2009**, *64*, i5–i11.
30. Golinelli, D.; Ryan, G.; Green, H.D.; Kennedy, D.P.; Tucker, J.S.; Wenzel, S.L. Sampling to reduce respondent burden in personal network studies and its effect on estimates of structural measures. *Field Methods* **2010**, *22*, 217–230. [[CrossRef](#)]
31. McCarty, C.; Killworth, P.D.; Rennell, J. Impact of methods for reducing respondent burden on personal network structural measures. *Soc. Netw.* **2007**, *29*, 300–315.
32. Granovetter, M.S. The Strength of Weak Ties. *Am. J. Sociol.* **1973**, *78*, 1360–1380. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.