

Evaluating Group-Level Recovery of Contemporaneous and Lagged Connectivity in GIMME: A Simulation Study

Gates and Molenaar, 2012

Yijia Xue and Linlin Liu
University of California, Riverside
STAT 293

December 4, 2025

Table of Contents

- 1 Motivation and Introduction
- 2 The GIMME Solution
- 3 Original Results
- 4 Our Replication and Results
- 5 Conclusion and Discussion
- 6 References

Table of Contents

- 1 Motivation and Introduction
- 2 The GIMME Solution
- 3 Original Results
- 4 Our Replication and Results
- 5 Conclusion and Discussion
- 6 References

The Problem: Brain Connectivity Analysis

Two Critical Issues: ¹

1. Poor Reliability

- Most methods: 50-70% detection of true connections
- Direction accuracy: barely above chance (50%)
- Lagged methods: only 20% detection

2. Ignored Heterogeneity

- Traditional methods assume all individuals are identical
- Group models may not describe any individual
- Individual differences treated as noise

¹Smith et al. (2011); Molenaar (2004)

The Problem: Brain Connectivity Analysis

Two Critical Issues: ¹

1. Poor Reliability

- Most methods: 50-70% detection of true connections
- Direction accuracy: barely above chance (50%)
- Lagged methods: only 20% detection

2. Ignored Heterogeneity

- Traditional methods assume all individuals are identical
- Group models may not describe any individual
- Individual differences treated as noise

Need: Reliable individual-level connectivity maps

¹Smith et al. (2011); Molenaar (2004)

Types of Heterogeneity

Heterogeneity² can appear as different connection strengths, different structures, or even reversed directions between individuals.

²Molenaar (2004); Fair et al. (2010); Hillary et al. (2011)

Types of Heterogeneity

Heterogeneity² can appear as different connection strengths, different structures, or even reversed directions between individuals.

Key Point

Concatenation and averaging fail when heterogeneity exists.

²Molenaar (2004); Fair et al. (2010); Hillary et al. (2011)

Table of Contents

- 1 Motivation and Introduction
- 2 The GIMME Solution**
- 3 Original Results
- 4 Our Replication and Results
- 5 Conclusion and Discussion
- 6 References

The GIMME Solution: Two-Stage Approach

The GIMME Solution: Two-Stage Approach

Stage 1: Group Search

- Find connections common to most, not all, individuals
- Use modification indices (MIs) across subjects
- Select paths that improve $\geq 75\%$ of individual models
- Iteratively build group structure

The GIMME Solution: Two-Stage Approach

Stage 1: Group Search

- Find connections common to most, not all, individuals
- Use modification indices (MIs) across subjects
- Select paths that improve $\geq 75\%$ of individual models
- Iteratively build group structure

Stage 2: Individual Refinement

- Start with group structure as baseline
- Add person-specific connections for each individual
- Optimize individual model fit
- Check fit indices (RMSEA, CFI, SRMR)

The GIMME Solution: Two-Stage Approach

Stage 1: Group Search

- Find connections common to most, not all, individuals
- Use modification indices (MIs) across subjects
- Select paths that improve $\geq 75\%$ of individual models
- Iteratively build group structure

Stage 2: Individual Refinement

- Start with group structure as baseline
- Add person-specific connections for each individual
- Optimize individual model fit
- Check fit indices (RMSEA, CFI, SRMR)

Leverages shared information while allowing heterogeneity

Unified Structural Equation Model (uSEM): ³

$$X_i(t) = \underbrace{A_i X_i(t)}_{\text{Contemporaneous}} + \underbrace{\Phi_i X_i(t-1)}_{\text{Lagged}} + \text{noise}$$

- A_i : Same time-point influences (ROI $j \rightarrow$ ROI i)
- Φ_i : Time-delayed influences (autoregressive)

GIMME Extension: Each parameter has two components

- Group component (A^g, Φ^g): common to most individuals
- Individual component (A_i, Φ_i): person-specific deviations

³Kim et al. (2007); Gates et al. (2010, 2011)

Why Does GIMME Work?

Why Does GIMME Work?

1. Signal Accumulation

- True connections appear consistently across individuals
- Noise-driven connections don't replicate across people
- Similar to signal processing in genomics

Why Does GIMME Work?

1. Signal Accumulation

- True connections appear consistently across individuals
- Noise-driven connections don't replicate across people
- Similar to signal processing in genomics

2. No Forced Homogeneity

- Group search provides common backbone
- Individual search captures real differences
- Unlike concatenation which forces one model

Table of Contents

- 1 Motivation and Introduction
- 2 The GIMME Solution
- 3 Original Results**
- 4 Our Replication and Results
- 5 Conclusion and Discussion
- 6 References

Practical Considerations

Advantages

- Works with $N \geq 25$ subjects
- R package available (gimme)
- Enables subgroup discovery

Advantages

- Works with $N \geq 25$ subjects
- R package available (gimme)
- Enables subgroup discovery

Limitations

- Limit: ≈ 50 ROIs (memory)
- Computationally intensive
- Very short scans problematic

Practical Considerations

Advantages

- Works with $N \geq 25$ subjects
- R package available (`gimme`)
- Enables subgroup discovery

Limitations

- Limit: ≈ 50 ROIs (memory)
- Computationally intensive
- Very short scans problematic

Applications

- Clinical heterogeneity (ADHD, depression, TBI)
- Developmental and longitudinal studies
- Subgroup identification within diagnostic categories

Gates & Molenaar (2012) show:

- GIMME can recover group-level and individual-level connectivity well
- Works under heterogeneous structures and different designs
- Uses group information to stabilize individual estimates

Gates & Molenaar (2012) show:

- GIMME can recover group-level and individual-level connectivity well
- Works under heterogeneous structures and different designs
- Uses group information to stabilize individual estimates

Core Question

How does the number of observations per subject (T) affect GIMME's ability to recover the true group-level contemporaneous and lagged networks?

From the Paper to Our Question

Gates & Molenaar (2012) show:

- GIMME can recover group-level and individual-level connectivity well
- Works under heterogeneous structures and different designs
- Uses group information to stabilize individual estimates

Core Question

How does the number of observations per subject (T) affect GIMME's ability to recover the true group-level contemporaneous and lagged networks?

Why This Matters

- Time series length is expensive to increase in real studies
- Need guidance on "how much T is enough" for reliable group networks

Table of Contents

- 1 Motivation and Introduction
- 2 The GIMME Solution
- 3 Original Results
- 4 Our Replication and Results**
- 5 Conclusion and Discussion
- 6 References

Simulation Design: Overview

Goal: Controlled setting with known ground truth for both A and Φ

Design Summary

- $N = 100$ subjects, $p = 8$ variables (nodes)
- Observed time points per subject: $T_{\text{obs}} \in \{50, 150, 300\}$
- For each T_{obs} , $R = 20$ Monte Carlo replications
- Burn-in of 100 time points and keep last T_{obs} observations

Model: unified SEM (uSEM)

$$\eta_i(t) = A_i \eta_i(t) + \Phi_i \eta_i(t-1) + \xi_i(t),$$

$$\eta_i(t) = (I - A_i)^{-1} [\Phi_i \eta_i(t-1) + \xi_i(t)], \quad \xi_i(t) \sim N(0, \Psi_i).$$

The true group-level A_{common} and Φ_{common} are known and are compared to GIMME's estimated group networks.

True Network Structure and Heterogeneity

Group-Level Contemporaneous Structure (A_{common})

- Simple 8-node directed ring: $8 \rightarrow 7 \rightarrow \dots \rightarrow 1$
- No self-loops or bidirectional contemporaneous edges
- Moderate positive weights (0.2–0.35)

Group-Level Lagged Structure (Φ_{common})

- Autoregressive (AR) self-loops for all nodes
- Additional first-, second-, and third-lag edges to create a moderately dense lag structure

Individual Heterogeneity

- Each subject receives 0–2 random contemporaneous edges and 2–5 random lagged edges
- Ensures individuals share a common backbone but differ in finer details
- Structured noise covariance Ψ_i with small off-diagonal covariances (1–4 random edges)

Example: Group vs. Individual Matrices

Individual Heterogeneity

$$A_{\text{common}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \otimes & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \otimes & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \otimes & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \otimes & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \otimes & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \otimes & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \otimes & 0 \end{bmatrix}$$

$$A_i = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \otimes & 0 & 0 & 0 & 0 & \star & 0 & 0 \\ 0 & \otimes & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \star & \otimes & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \otimes & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \otimes & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \otimes & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \otimes & 0 \end{bmatrix}$$

$$\otimes \sim \text{runif}(1, 0.2, 0.35) \quad \star \sim \text{runif}(1, 0.03, 0.12)$$

Example: Group vs. Individual Matrices Cont.

$$\Phi_{\text{common}} = \begin{bmatrix} \otimes & 0 & 0 & 0 & \blacksquare & 0 & 0 & 0 \\ \diamond & \otimes & 0 & 0 & 0 & 0 & 0 & \blacksquare \\ \blacklozenge & \diamond & \otimes & 0 & 0 & 0 & 0 & 0 \\ 0 & \blacklozenge & \diamond & \otimes & 0 & 0 & 0 & 0 \\ 0 & \blacksquare & \blacklozenge & \diamond & \otimes & 0 & 0 & 0 \\ 0 & 0 & 0 & \blacklozenge & \diamond & \otimes & 0 & 0 \\ 0 & 0 & 0 & 0 & \blacklozenge & \diamond & \otimes & 0 \\ 0 & 0 & 0 & 0 & 0 & \blacklozenge & \diamond & \otimes \end{bmatrix}$$

$$\Phi_i = \begin{bmatrix} \otimes & 0 & 0 & 0 & \blacksquare & \star & 0 & 0 \\ \diamond & \otimes & 0 & 0 & 0 & 0 & 0 & \blacksquare \\ \blacklozenge & \diamond & \otimes & 0 & 0 & 0 & 0 & \star \\ 0 & \blacklozenge & \diamond & \otimes & 0 & 0 & 0 & 0 \\ 0 & \blacksquare & \blacklozenge & \diamond & \otimes & 0 & 0 & 0 \\ 0 & \star & 0 & \blacklozenge & \diamond & \otimes & \star & 0 \\ 0 & 0 & 0 & 0 & \blacklozenge & \diamond & \otimes & \star \\ 0 & 0 & 0 & 0 & 0 & \blacklozenge & \diamond & \otimes \end{bmatrix}$$

$$\otimes = 1.5 \quad \diamond \sim \text{runif}(1, 0.2, 0.3)$$

$$\blacklozenge \sim \text{runif}(1, 0.1, 0.2) \quad \blacksquare \sim \text{runif}(1, 0.2, 0.3) \quad \star \sim \text{runif}(1, 0.05, 0, 15)$$

Example: Group vs. Individual Matrices Cont.

$$A_i = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \otimes & 0 & 0 & 0 & 0 & \star & 0 & 0 \\ 0 & \otimes & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \star & \otimes & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \otimes & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \otimes & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \otimes & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \otimes & 0 \end{bmatrix} \quad \Psi_i = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & \bullet \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \bullet & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \bullet & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\bullet \sim \text{runif}(1, 0.01, 0.05)$$

Fitting GIMME in Our Study

For each simulated dataset:

- Apply `gimme()` to all 100 subjects' time series
- Use group-level cutoff = 0.75 and individual-level cutoff = 0.50
- Allow autoregressive (lag-1) effects

Group-Level Recovery

- Edges included in the group network if they improve at least 75% of subjects' models
- We obtain estimated A_{est} and Φ_{est} for each T

Evaluation

- Compare estimated vs. true group matrices using TP, FP, FN, TN
- Results averaged over the 20 Monte Carlo replications per T

For each simulation and each matrix (A and Φ), we compute:

- True Positives (TP), False Positives (FP)
- False Negatives (FN), True Negatives (TN)

Key Metrics

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}.$$

Summary Across Replications

- Metrics averaged over the 20 Monte Carlo runs for each T
- Separate evaluations for contemporaneous (A) and lagged (Φ) networks

Goal: Evaluate how time series length (T) affects recovery of

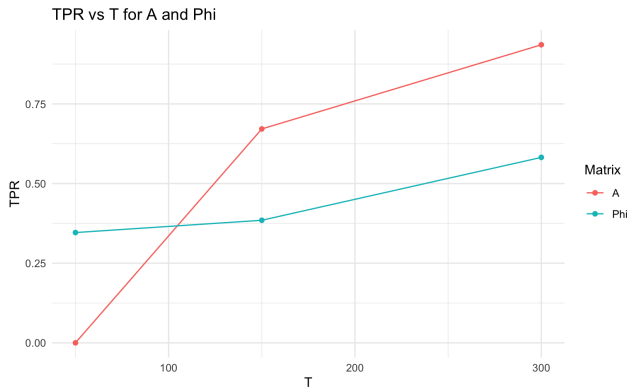
$$A_{\text{common}}, \quad \Phi_{\text{common}}.$$

Three complementary analyses:

- Quantitative recovery metrics (TPR, FPR)
- Summary tables across 20 replications
- Visual comparison of true vs. estimated networks

Key idea: Longer time series should improve estimation, but contemporaneous and lagged edges may behave differently.

Recovery Improves with Time Series Length



- Contemporaneous edges (A) show strong improvement as T increases.
- Lagged edges (Φ) improve more slowly and remain harder to recover.
- Indicates that group-level Φ requires substantially more data than A .

Group-Level Recovery of A Across Replications

T_{obs} <dbl>	mean_TPR <dbl>	sd_TPR <dbl>	mean_FPR <dbl>	sd_FPR <dbl>	mean_spec <dbl>	sd_spec <dbl>	mean_FP <dbl>	sd_FP <dbl>	mean_FN <dbl>	sd_FN <dbl>	mean_FNR <dbl>	sd_FNR <dbl>
50	0.0000000	0.00000000	0.00000000	0.00000000	1.0000000	0.00000000	0.00	0.0000000	7.00	0.0000000	1.00000000	0.00000000
150	0.6714286	0.17404026	0.000877193	0.003922926	0.9991228	0.003922926	0.05	0.2236068	2.30	1.2182818	0.32857143	0.17404026
300	0.9357143	0.09804753	0.002631579	0.006427150	0.9973684	0.006427150	0.15	0.3663475	0.45	0.6863327	0.06428571	0.09804753

- $T = 50$: almost no recovery of the ring structure.
- $T = 150$: moderate but incomplete recovery.
- $T = 300$: high TPR and low FPR—nearly full recovery of A .

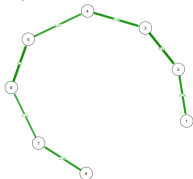
Group-Level Recovery of Φ Across Replications

T_obs <dbl>	mean_TPR <dbl>	sd_TPR <dbl>
50	0.3461321	0.01463161
150	0.3847414	0.03872115
300	0.5821311	0.02824990

- Lagged edges are consistently more difficult to detect.
- TPR increases with T , but remains far below contemporaneous recovery.
- Even with $T = 300$, many true lagged edges do not meet the group cutoff.

Example Networks: True vs. Estimated A

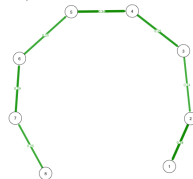
True A, $T = 50$



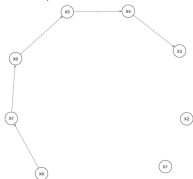
Estimated A, $T = 50$



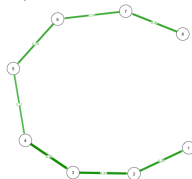
True A, $T = 150$



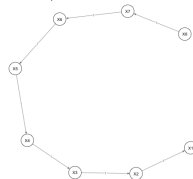
Estimated A, $T = 150$



True A, $T = 300$

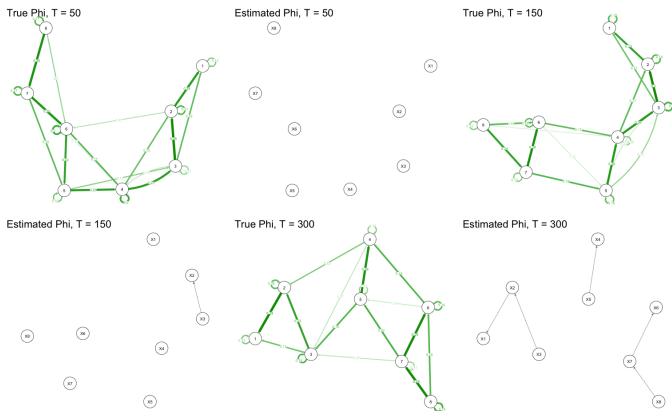


Estimated A, $T = 300$



- True A: an 8-node directed ring.
- $T = 50$: almost no group edges are detected.
- $T = 150$: partial recovery of the ring.
- $T = 300$: estimated network closely matches the true structure.

Example Networks: True vs. Estimated Φ



- True Φ : autoregressive loops + several off-diagonal effects.
- Estimated Φ remains sparse at all T .
- Many true lagged edges fail to reach the 75% group inclusion threshold.

Key Insights from the Results

Contemporaneous edges (A)

- Accurately recovered once T is moderately large.
- By $T = 300$, group-level recovery is excellent.

Lagged edges (Φ)

- Require much more data to reliably detect.
- Under-recovered even at $T = 300$ due to weaker temporal signatures.

Overall conclusion

- Longer time series substantially improve group-level connectivity estimation, especially for contemporaneous relationships.

Table of Contents

- 1 Motivation and Introduction
- 2 The GIMME Solution
- 3 Original Results
- 4 Our Replication and Results
- 5 Conclusion and Discussion**
- 6 References

The Power of Group Information

- GIMME leverages consistency across individuals to stabilize path selection
- With enough data, this yields highly accurate group-level contemporaneous networks

Heterogeneity and Lagged Effects

- Our simulation confirms that heterogeneity can be accommodated without forcing one model on everyone
- Lagged edges are more fragile, as they require more information and may compete with contemporaneous paths

Practical Takeaways

- For designs like ours, $T \approx 150$ is a lower bound for meaningful group-level A , but $T \approx 300$ is much better
- Better lagged dynamics estimation may need longer series, looser thresholds, or complementary methods

Thank you!

Questions?

Table of Contents

- 1 Motivation and Introduction
- 2 The GIMME Solution
- 3 Original Results
- 4 Our Replication and Results
- 5 Conclusion and Discussion
- 6 References**



Fair, D. A., Posner, J., Nagel, B. J., Bathula, D., Dias, T. G. C., Mills, K. L., Blythe, M. S., Giwa, A., Schmitt, C. F., & Nigg, J. T. (2010). Atypical default network connectivity in youth with attention-deficit/hyperactivity disorder. *Biological Psychiatry*, 68(12), 1084–1091.



Hillary, F. G., Medaglia, J. D., Gates, K. M., Molenaar, P. C. M., Slocumb, J., Peechatka, A., & Good, D. C. (2011). Examining working memory task acquisition in a disrupted neural network. *Brain: A Journal of Neurology*, 134, 1555-1570.



Gates, K. M., Molenaar, P. C. M., Hillary, F. G., Ram, N., & Rovine, M. J. (2010). Automatic search in fMRI connectivity mapping: an alternative to Granger causality using formal equivalences between SEM path modeling, VAR, and unified SEM. *NeuroImage*, 53, 1118-1125.



Gates, K. M., Molenaar, P. C. M., Hillary, F. G., & Slobounov, S. (2011). Extended unified SEM approach for modeling event-related fMRI data. *NeuroImage*, 54, 1151-11158.



Gates, K.M., & Molenaar, P.C.M. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, 63, 310–319.



Kim, J., Zhu, W., Chang, L., Bentler, P.M., & Ernst, T. (2007). Unified structural equation modeling for the analysis of multisubject, multivariate functional fMRI data. *Human Brain Mapping*, 28, 85–93.



Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: bringing the person back into scientific psychology, this time forever. *Measurement*, 2(4), 201–218.



Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., Ramsey, J. D., & Woolrich, M. W. (2011). Network modeling methods for FMRI. *NeuroImage*, 54, 875–891.