# Transfer Learning on Edge Connecting Probability Estimation Under Graphon Model

**Yuyao Wang**[1], Yu-Hung Cheng[2],
Debarghya Mukherjee[1], Huimin Cheng[2]

[1]Department of Mathematics and Statistics, Boston University
[2]Department of Biostatistics, Boston University

**NeurIPS 2025**

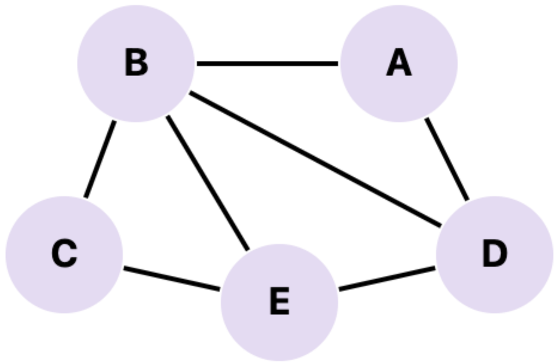
# Outline

1. Introduction
   - Background
   - Motivation
2. Method
3. Experiments
4. Conclusion

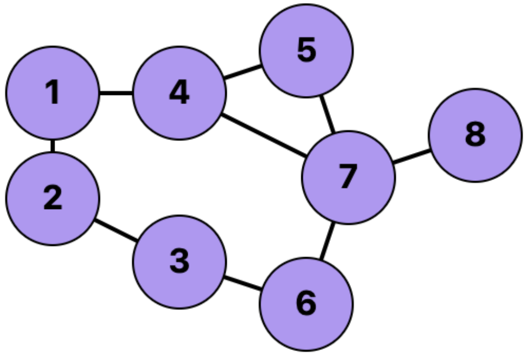
# Background & Motivation

- **Edge probabilities:** Core to network modeling (social, biological, etc.); basis for link prediction, denoising, augmentation.
- **Graphon framework:** Smooth function $W: [0,1]^2 \to [0,1]$, encoding connection probabilities
- **Challenge:** Accurate graphon estimation needs large graphs; small real datasets lead to high variance.
- **Transfer idea:** Leverage related large graphs for small-graph estimation.
- **Problem:** Prior methods assume node correspondences — unrealistic.
- **Goal:** Enable correspondence-free graphon transfer with reliable alignment and negative-transfer avoidance.

Small Network




Large Network



# Outline

# Problem Setup

## Problem Setup

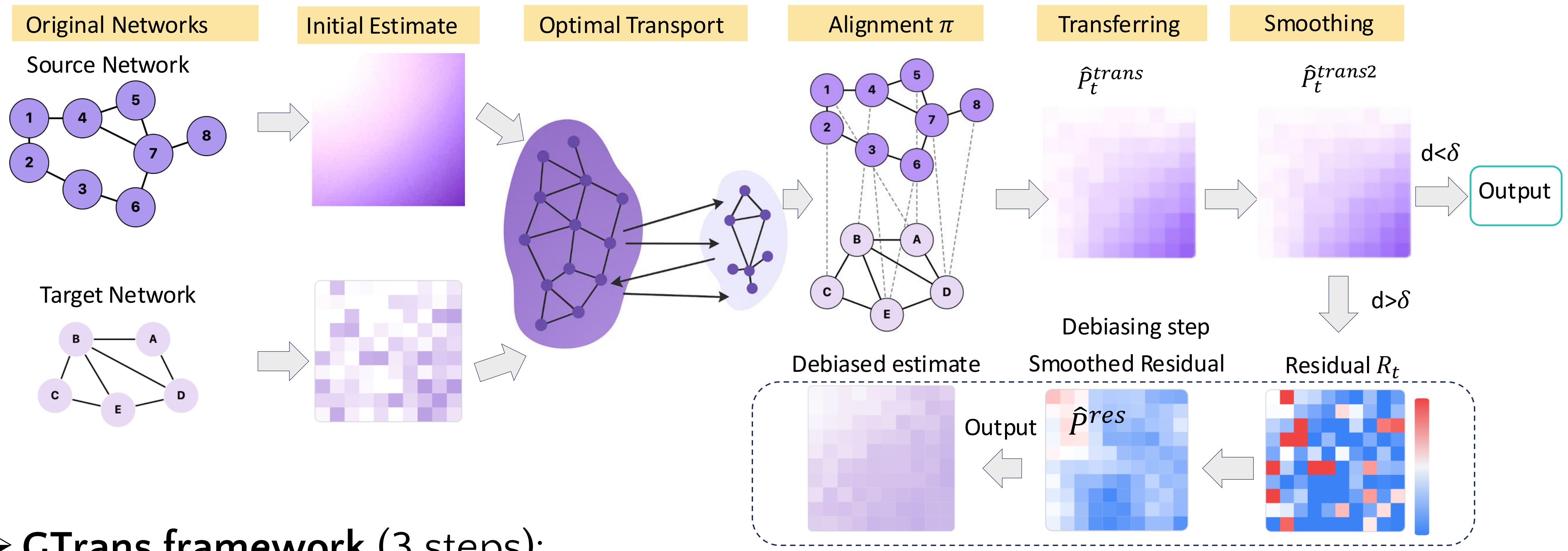
- Goal: estimate **target graphon** (latent probability matrix $P_t$ (for small graph
- **Source graph**: adjacency $A_s \in [0,1]^{n_s \times n_s}$, $n_s > n_t$.
- **Target graph**: adjacency $A_t \in [0,1]^{n_t \times n_t}$.
- Both generated from latent graphons $f_s, f_t$:

$$A_{s,ij} \sim Ber\left(f_s(u_{s,i}, u_{s,j})\right), A_{t,ij} \sim Ber\left(f_t(u_{t,i}, u_{t,j})\right).$$

## Objective

- Estimate $P_t$ using **own data + transferred knowledge from source**
- Adapt to **source–target differences** to avoid negative transfer

# Method-Overview: GTrans




➢ **GTrans framework** (3 steps):

- **Initial estimation** → Neighborhood smoothing on source & target
- **Transferring** → Align structural patterns using optimal transport,
- **Adaptive Debiasing** → Correct target-specific deviations, prevent negative transfer

# Initial Estimation Step

NEURAL INFORMATION PROCESSING SYSTEMS

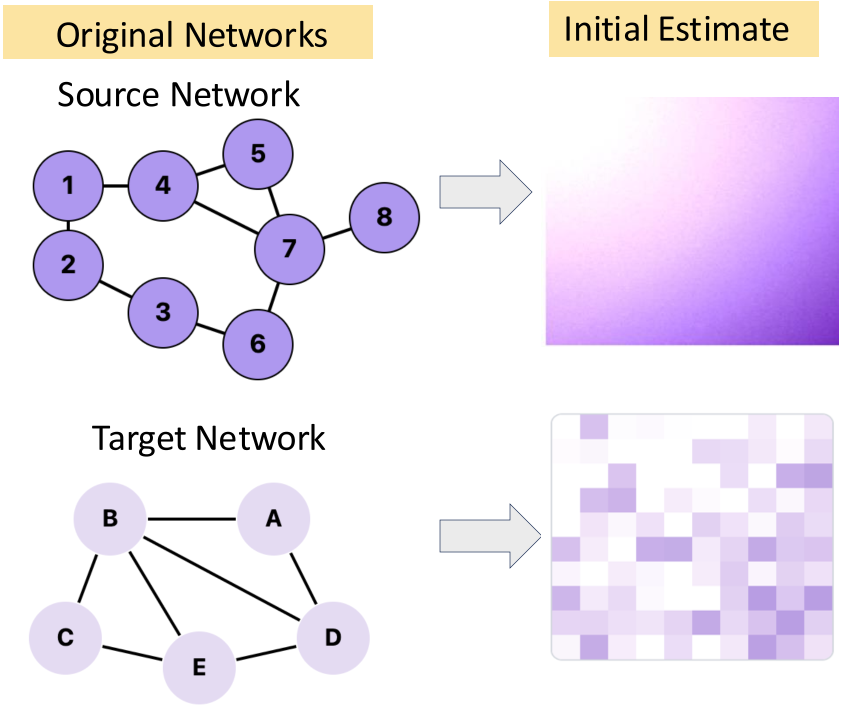
BOSTON UNIVERSITY

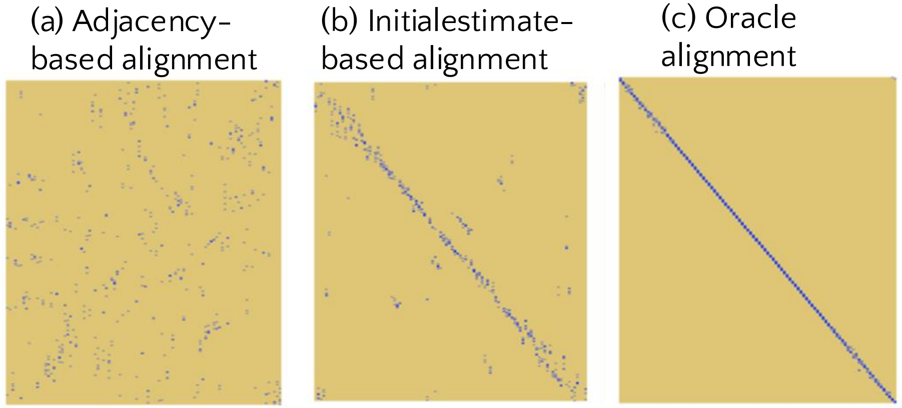
## Initial Graphon Estimation

- **Input:** source $A_s$ and target $A_t$
- **Neighborhood Smoothing (NS)** → denoise adjacency, capture coarse:

$$\hat{P}_s^{ini} \in [0,1]^{(n_s \times n_s)}, \hat{P}_t^{ini} \in [0,1]^{(n_t \times n_t)}$$

- **Advantage:** smoother, closer to ground truth than raw adjacency
- **Limitation:** target estimate less reliable (small $n_t$)

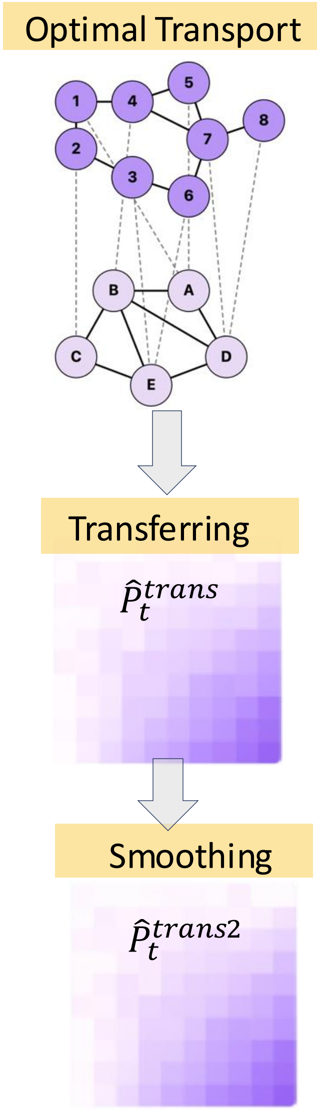


# Transferring Step

**Goal:** Align source–target graphs and transfer structure to target space

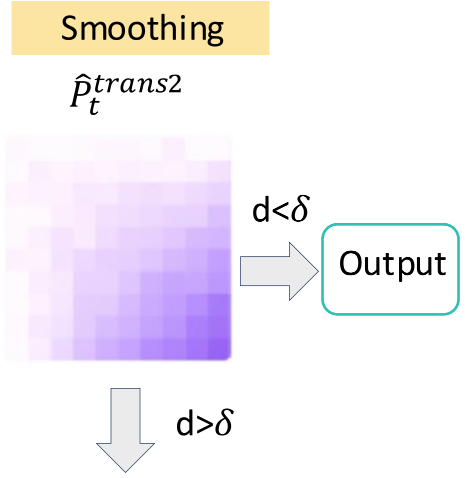
- Optimal Transport (GW / EGW):
- Compute alignment matrix $\hat{\pi} \in [0,1]^{n_s \times n_t}$ on $\widehat{P_s^{ini}}, \widehat{P_t^{ini}}$
    - GW → sharper alignment;
    - EGW → more efficient for large graphs
- **Normalization**:
    - Column-normalize $\hat{\pi} \rightarrow \tilde{\pi}$
    - $\widetilde{\pi_{ij}}$ = correspondence strength between nodes
- **Projection & Refinement**
    - Transfer: $\widehat{P_t^{trans}} = \widetilde{\pi}^T \widehat{P_s^{ini}} \widetilde{\pi}$
    - Apply smoothing → $\widehat{P_t^{trans2}}$
    - Final estimator is **aligned + smooth**



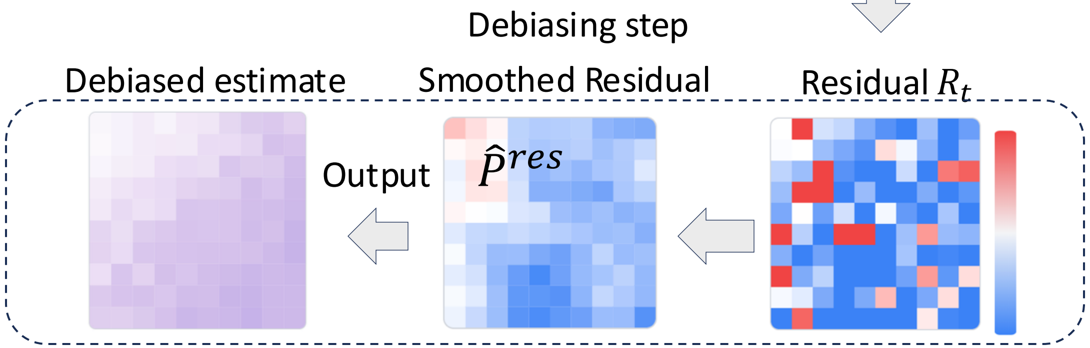
# Debiasing Step

Applied If GW distance $d > \delta \rightarrow$ large domain shift, risk of negative transfer

- **Residual Computation**
  - $R_t = \widehat{P_t^{ini}} - \widehat{P_t^{trans2}}$
  - captures target-specific patterns not covered by transfer
- **Residual Smoothing**
  - Apply NS on $R_t \rightarrow$ denoised $\widehat{P_t^{res}}$; keeps structure, reduces noise




- **Final Estimator**
  - $\widehat{P_t} = \widehat{P_t^{trans2}} + \widehat{P_t^{res}}$
  - If $d \leq \delta$: directly use $\widehat{P_t^{trans2}}$




- **Benefit:** Adaptive correction $\rightarrow$ preserve transfer gains, avoid negative transfer

# Outline

1. Introduction
2. Method
3. Experiments
   - Link Prediction
   - Graph Classification (via Data Augmentation)
4. Conclusion

# Link Prediction Experiments Setup



Link prediction AUC (%): Real Networks, with source wiki-vote ($n_s$ = 889):

| Dataset | GTrans-GW | GTrans-EGW | NS[1] | USVT[4] | SAS[2] | ICE[3] |
|---|---|---|---|---|---|---|
| Dolphins ($n_t$ = 62) | 75.96 ± 8.53 | 76.26 ± 8.54 | 70.60 ± 9.01 | 72.36 ± 10.16 | 50.66 ± 6.35 | 75.36 ± 8.40 |
| Firm ($n_t$ = 33) | 71.31 ± 12.18 | 71.26 ± 12.06 | 66.32 ± 12.27 | 65.56 ± 12.25 | 54.90 ± 7.59 | 64.26 ± 12.20 |
| Football ($n_t$ = 115) | 86.64 ± 3.66 | 86.74 ± 3.72 | 86.75 ± 3.49 | 85.32 ± 3.56 | 44.56 ± 7.38 | 82.46 ± 4.59 |
| Karate ($n_t$ = 34) | 82.47 ± 10.36 | 82.53 ± 10.46 | 76.74 ± 12.01 | 71.86 ± 15.20 | 63.88 ± 11.15 | 80.43 ± 11.22 |

- GTrans consistently outperforms baselines (NS, SAS, USVT, ICE)

[1] Yuan Zhang, Elizaveta Levina, and Ji Zhu. Estimating network edge probabilities by neighbourhood smoothing. Biometrika, 104(4):771–783, 2017.591
[2] Stanley Chan and Edoardo Airoldi. A consistent histogram estimator for exchangeable graph models. In Eric P. Xing and Tony Jebara, editors, Proceedings of the 31st International Conference on Machine Learning, volume 32 of Proceedings of Machine Learning Research,448 pages 208–216, 22–24 Jun 2014. PMLR.449
[3]. Yichen Qin, Linhan Yu, and Yang Li. Iterative Connecting Probability Estimation for Networks. In Advances in NeurIPS, volume 34, pages 1155–1166, 2021
[4] Sourav Chatterjee. Matrix estimation by Universal Singular Value Thresholding. The Annals of Statistics, 43(1), 2015

# Graph Classification Experiments Setup

BOSTON UNIVERSITY

- **Motivation：** Graph classification often suffers from **small graphs;** Graphon-based augmentation (G-Mixup) relies on accurate probability estimation → conventional estimators fail on small graphs.
- **Datasets**
  - **Targets**: IMDB-Binary, IMDB-Multi, PROTEINS-Full (all small graphs)
  - **Sources**: COLLAB (74 nodes), Reddit-Binary (429 nodes), D&D (284 nodes)
- **Baselines:**
  - **Non-Transfer baselines**: NS[1], SAS[2], ICE[3], USVT[4], GWB [5], IGNR [6], SIGL [7], Graphlet Kernel [8]
- **Evaluation Metrics:** average test accuracy over 10 runs with standard GCN classifier.


[1] Yuan Zhang, Elizaveta Levina, and Ji Zhu. Estimating network edge probabilities by neighbourhood smoothing. Biometrika, 104(4):771–783, 2017.591
[2] Stanley Chan and Edoardo Airoldi. A consistent histogram estimator for exchangeable graph models. In Eric P. Xing and Tony Jebara, editors, Proceedings of the 31st International Conference on Machine Learning, volume 32 of Proceedings of Machine Learning Research,448 pages 208–216, 22–24 Jun 2014. PMLR.449
[3]. Yichen Qin, Linhan Yu, and Yang Li. Iterative Connecting Probability Estimation for Networks. In Advances in NeurIPS, volume 34, pages 1155–1166, 2021
[4] Sourav Chatterjee. Matrix estimation by Universal Singular Value Thresholding. The Annals of Statistics, 43(1), 2015
[5] Hongteng Xu, Dixin Luo, Lawrence Carin, and Hongyuan Zha. Learning graphons via structured gromov-wasserstein barycenters. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, pages 10505–10513, 2021.
[6] Xinyue Xia, Gal Mishne, and Yusu Wang. Implicit graphon neural representation. In International Conference on Artificial Intelligence and Statistics, pages 10619–10634. PMLR, 2023
[7] Ali Azizpour, Nicolas Zilberstein, and Santiago Segarra. Scalable implicit graphon learning. arXiv preprint arXiv:2410.17464, 2024.
[8] Nino Shervashidze, SVN Vishwanathan, Tobias Petri, Kurt Mehlhorn, and Karsten Borgwardt. Efficient graphlet kernels for large graph comparison. In Artificial intelligence and statistics, pages 488–495. PMLR, 2009


NEURAL INFORMATION PROCESSING SYSTEMS

# Graph Classification Results



Graph classification accuracy (%) compared against baselines

| Source | Target | GTrans-GW | GTrans-EGW | NS | USVT | SAS | ICE | GWB | IGNR | SIGL | Graphlet |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reddit-B | IMDB-B | 76.30 | 76.80 | 72.90 | 73.85 | 74.25 | 74.30 | 75.30 | 74.35 | 73.50 | 61.10 |
| COLLAB | IMDB-B | 76.25 | 77.50 | 72.90 | 73.85 | 74.25 | 74.30 | 75.30 | 74.35 | 73.50 | 61.10 |
| Reddit-B | IMDB-M | 49.10 | 51.27 | 43.80 | 48.00 | 44.10 | 43.90 | 47.70 | 47.50 | 49.13 | 39.37 |
| COLLAB | IMDB-M | 50.47 | 50.23 | 43.80 | 48.00 | 44.10 | 43.90 | 47.70 | 47.50 | 49.13 | 39.37 |
| D&D | PROTEINS | 69.33 | 68.52 | 63.18 | 65.11 | 65.25 | 65.38 | 63.45 | 65.87 | 67.13 | 70.11 |

**GTrans-GW** and **GTrans-EGW**, consistently achieves the best or comparable performance in most cases.

# Outline

1. Introduction
2. Method
3. Experiments
4. Conclusion

# Conclusion

- **Understanding Gap**
  - Graphon estimation is unreliable on small, sparse networks.
  - Existing transfer methods assume known node correspondences, which is unrealistic.
  - Gap: transfer information without correspondences, while avoiding negative transfer.

- **GTRANS Framework**
  - Proposed GTRANS, with three core steps
  - First framework enabling robust, correspondence-free graphon transfer.

- **Theory**: consistency guarantees for alignment via smoothed estimators.

- **Experiments**
  - Synthetic: outperforms baselines, robust to domain shifts.
  - Real networks: Outperforms baselines in link prediction and graph classification.