Supplementary material

Anonymous submission

In this supplementary material, we provide:

- the algorithm of **BaLu** (e.i., Algorithm 1);
- the formal analysis of time and space complexity of Algorithm 1;
- the reproducibility checklist.

Technical Appendix

A. Algorithm and Computing Complexity

The algorithm of **BaLu** is described in **Algorithm** 1.
Time Complexity Analysis. Let $n = |V_C|$ be the number of units, $m = |P_C^{\text{ctx}}|$ the number of contextual attributes, $\bar{m}$ the average number of observed attributes per unit, $\bar{n}$ the average relational degree (i.e., neighbors per unit in $\mathcal{E}_{\text{rel}}$), h the embedding dimension for nodes and messages, and d the embedding dimension for edge attributes. Note that L and K represent the number of *L-layer* and *K-layer*, respectively.

Data Imputation Component (per layer). Each of the *L-layer* performs the following operations:

- *Unit–Attribute Message Passing:* For each edge $(u, p) \in \mathcal{E}_{\text{dat}}$, the message is computed and aggregated. The overall cost is

$$O(|\mathcal{E}_{\text{dat}}| \cdot (h + d)) = O(n \cdot \bar{m} \cdot (h + d)).$$

- *Relational Message Passing:* For each unit u, relational information from neighbors is aggregated, with total cost

$$O(|\mathcal{E}_{\text{rel}}| \cdot h) = O(n \cdot \bar{n} \cdot h).$$

- *Edge Embedding Update:* Each edge in $\mathcal{E}_{\text{dat}}$ updates its embedding using connected node representations. The total cost matches the message passing cost:

$$O(|\mathcal{E}_{\text{dat}}| \cdot (h + d)) = O(n \cdot \bar{m} \cdot (h + d)).$$

The total cost of the data imputation component across *L-layer* is:

$$O\left(L \cdot n \cdot (\bar{m} \cdot (h + d) + \bar{n} \cdot h)\right).$$

Causal Estimation Component.

- *Contextual Vector Construction:* For each unit u, a context vector is constructed by concatenating observed and imputed attributes. This costs:

$$O(n \cdot m \cdot h).$$

- *Interference Modeling (K-layer):* Each unit propagates treatment representations across its neighbors. The cost is:

$$O(K \cdot n \cdot \bar{n} \cdot h).$$

- *Treatment Effect Estimation:* The final ITE for each unit is computed using an MLP applied to contextual and interference embeddings:

$$O(n \cdot h).$$

Total Complexity per Training Epoch. Combining both components, the overall time complexity per epoch is:

$$O\left(L \cdot n \cdot (\bar{m} \cdot (h + d) + \bar{n} \cdot h) + n \cdot m \cdot h + K \cdot n \cdot \bar{n} \cdot h\right).$$

Simplified Complexity under Typical Settings. Assuming $L = K$ and $d = h$, the complexity becomes:

$$O\left(n \cdot L \cdot h \cdot (\bar{m} + \bar{n}) + n \cdot m \cdot h\right).$$

Best-case Scenario (Sparse KGs). If the KG is sparse—as is common in real-world graphs (Pujara, Augustine, and Getoor 2017)—with $\bar{m} \ll m$ and $\bar{n} \ll n$, the complexity reduces to:

$$O(n \cdot \bar{m} \cdot L \cdot h + n \cdot \bar{n} \cdot L \cdot h + n \cdot m \cdot h),$$

which is linear in both n and m.

Worst-case Scenario (Dense KGs). In the worst case, where $\bar{m} = m$ and $\bar{n} = n$, the time complexity becomes:

$$O(n \cdot m \cdot L \cdot h + n^2 \cdot L \cdot h),$$

which is linear in the number of attributes m, but quadratic in the number of units n due to dense relational connections.

Space Complexity Analysis. The space complexity of **BaLu** is determined by the storage requirements for node and edge embeddings, model parameters, and intermediate activations used during forward and backward passes.

- **Node Embeddings:** Each unit $u \in V_C$ and each attribute $p \in P_C^{\text{ctx}}$ is assigned a node embedding of dimension h per layer. For *L-layer*, the storage for node embeddings is:

$$O(L \cdot (n+m) \cdot h).$$

- **Edge Embeddings:** Each observed edge $(u,p) \in \mathcal{E}_{\text{dat}}$ is associated with an embedding of dimension d per layer, resulting in:

$$O(L \cdot n \cdot \bar{m} \cdot d).$$

- **Context and Imputation Vectors:** For each unit, the contextual vector $\mathbf{d}_u \in \mathbb{R}^m$, the combined representation $\mathbf{x}_u \in \mathbb{R}^{m+h}$, and imputed values $\hat{e}_{u,p}$ are stored. The space cost is:

$$O(n \cdot (m+h)).$$

- **Interference Embeddings:** Each unit maintains a interference representation across *K-layer*. The space required is:

$$O(K \cdot n \cdot h).$$

- **Model Parameters:** The model includes learnable parameters for GNNs, MLPs ($f_e, f_x, f_y^{(0)}, f_y^{(1)}$), and aggregation functions. Assuming each has constant number of layers and moderate width, the total parameter size is:

$$O((L+K) \cdot h^2 + m \cdot h^2).$$

  This is negligible compared to data-dependent terms when $n \gg h$.

Total Space Complexity. Combining the terms above and assuming $L = K$ and $d = h$, the total space complexity becomes:

$$\begin{aligned}
&O(L \cdot h \cdot (n+m) \quad \text{(node embeddings)}\\
&\quad + L \cdot n \cdot \bar{m} \cdot h \quad \text{(edge embeddings)}\\
&\quad + n \cdot (m+h) \quad \text{(context vectors \& imputation)}\\
&\quad + L \cdot n \cdot h \quad \text{(interference embeddings)}\\
&\quad + L \cdot h^2 + m \cdot h^2 \quad \text{(model parameters)})
\end{aligned}$$

This expression can be further factorized into:

$$O\big(L \cdot h \cdot (n \cdot (1+\bar{m}) + m) + n \cdot (m+h) + L \cdot n \cdot h + (L+m) \cdot h^2\big)$$

This representation makes the interaction between dataset size $(n, m, \bar{m})$, model depth (L), and embedding dimensionality (h) explicit, and highlights which terms dominate under different sparsity regimes.

Simplified Bounds.

- **Best Case (Sparse KGs):** When $\bar{m} \ll m$ and both $\bar{m}$ and $\bar{n}$ are constants (i.e., the KG is sparse), the dominant terms reduce to:

$$O(L \cdot h \cdot (n+m) + n \cdot (m+h) + (L+m) \cdot h^2)$$

  The total space is linear in both n and m.

- **Worst Case (Dense KGs):** When $\bar{m} = m$, the dominant term becomes:

$$O(L \cdot h \cdot n \cdot m)$$

  which grows linearly with m and n.

Reproducibility Checklist

1. General Paper Structure

1.1. Includes a conceptual outline and/or pseudocode description of AI methods introduced (yes/partial/no/NA) **yes** (Algorithm 1 in Technical Appendix and detailed descriptions in Sections 4-5)

1.2. Clearly delineates statements that are opinions, hypothesis, and speculation from objective facts and results (yes/no) **yes**

1.3. Provides well-marked pedagogical references for less-familiar readers to gain background necessary to replicate the paper (yes/no) **yes** (Section 2 and references throughout)

2. Theoretical Contributions

2.1. Does this paper make theoretical contributions? (yes/no) **yes** (Check the section **Technical Appendix. A**.)

If yes, please address the following points:

2.2. All assumptions and restrictions are stated clearly and formally (yes/partial/no) **yes**

2.3. All novel claims are stated formally (e.g., in theorem statements) (yes/partial/no) **yes**

2.4. Proofs of all novel claims are included (yes/partial/no) **yes**

2.5. Proof sketches or intuitions are given for complex and/or novel results (yes/partial/no) **yes**

2.6. Appropriate citations to theoretical tools used are given (yes/partial/no) **yes**

2.7. All theoretical claims are demonstrated empirically to hold (yes/partial/no/NA) **partial**

2.8. All experimental code used to eliminate or disprove claims is included (yes/no/NA) **NA**

3. Dataset Usage

3.1. Does this paper rely on one or more datasets? (yes/no) **yes** (we demonstrate the effectiveness of our methods on four benchmark datasets, but our proposed solution is generalized for other datasets)

If yes, please address the following points:

3.2. A motivation is given for why the experiments are conducted on the selected datasets (yes/partial/no/NA) **yes** (Section 5 explains the need for synthetic datasets and the choice of semi-synthetic benchmarks)

3.3. All novel datasets introduced in this paper are included in a data appendix (yes/partial/no/NA) **yes** (Dataset generation code in `https://anonymous.4open.science/r/BaLu-E932/Data-BaLu/synetic_simulation/syn_instagram.py`)

3.4. All novel datasets introduced in this paper will be made publicly available upon publication of the paper with a license that allows free usage for research purposes (yes/partial/no/NA) **yes**

3.5. All datasets drawn from the existing literature (potentially including authors' own previously published work) are accompanied by appropriate citations (yes/no/NA) **yes** (BlogCatalog, Flickr (Guo, Li, and Liu 2020; Huang et al. 2023), YouTube (Lin et al. 2023))

3.6. All datasets drawn from the existing literature (potentially including authors' own previously published work) are publicly available (yes/partial/no/NA) **yes**

3.7. All datasets that are not publicly available are described in detail, with explanation why publicly available alternatives are not scientifically satisficing (yes/partial/no/NA) **NA**

4. Computational Experiments

4.1. Does this paper include computational experiments? (yes/no) **yes**

If yes, please address the following points:

4.2. This paper states the number and range of values tried per (hyper-) parameter during development of the paper, along with the criterion used for selecting the final parameter setting (yes/partial/no/NA) **yes** (stated in Section 5)

4.3. Any code required for pre-processing data is included in the appendix (yes/partial/no) **yes** (Data preprocessing and generation code in `https://anonymous.4open.science/r/BaLu-E932/Data-BaLu` folder)

4.4. All source code required for conducting and analyzing the experiments is included in a code appendix (yes/partial/no) **yes** (Complete implementation available at https://anonymous.4open.science/r/BaLu-E932/)

4.5. All source code required for conducting and analyzing the experiments will be made publicly available upon publication of the paper with a license that allows free usage for research purposes (yes/partial/no) **yes**

4.6. All source code implementing new methods have comments detailing the implementation, with references to the paper where each step comes from (yes/partial/no) **partial** (BaLu implementation in `https://anonymous.4open.science/r/BaLu-E932/BaLu_Plus` folder. The code has some comments to describe the the implementation, but it is well-structured for understanding.)

4.7. If an algorithm depends on randomness, then the method used for setting seeds is described in a way sufficient to allow replication of results (yes/partial/no/NA) **yes** (10 independent runs with different random seeds)

4.8. This paper specifies the computing infrastructure used for running experiments (hardware and software), including GPU/CPU models; amount of memory; operating system; names and versions of relevant software libraries and frameworks (yes/partial/no) **yes** (NVIDIA A100 GPU 80GB, PyTorch 2.6, EconML, Hyperimpute)

4.9. This paper formally describes evaluation metrics used and explains the motivation for choosing these metrics (yes/partial/no) **yes** (Section 5, Equation 8 - $\sqrt{\epsilon_{\text{PEHE}}}$ and ϵ_{MAE})

4.10. This paper states the number of algorithm runs used to compute each reported result (yes/no) **yes** (10 independent runs)

4.11. Analysis of experiments goes beyond single-dimensional summaries of performance (e.g., average; median) to include measures of variation, confidence, or other distributional information (yes/no) **yes** (mean ± standard deviation reported in Table 2 and Figure 3)

4.12. The significance of any improvement or decrease in performance is judged using appropriate statistical tests (e.g., Wilcoxon signed-rank) (yes/partial/no) **yes** (see Table 3 in the main paper)

4.13. This paper lists all final (hyper-)parameters used for each model/algorithm in the paper's experiments (yes/partial/no/NA) **yes** (Section 5, Experimental Settings: $h = 64$, $L = 2$, $K = 1$, $d = 16$, learning rate=0.01, dropout=0.1, $\alpha = 1$, $\beta = 10^{-4}$, $\gamma = 10^{-4}$, $\eta = 10^{-4}$)

References

Guo, R.; Li, J.; and Liu, H. 2020. Learning individual causal effects from networked observational data. In *Proceedings of the 13th international conference on web search and data mining*, 232–240.

Huang, Q.; Ma, J.; Li, J.; Guo, R.; Sun, H.; and Chang, Y. 2023. Modeling interference for individual treatment effect estimation from networked observational data. *ACM Transactions on Knowledge Discovery from Data*, 18(3): 1–21.

Lin, X.; Zhang, G.; Lu, X.; Bao, H.; Takeuchi, K.; and Kashima, H. 2023. Estimating treatment effects under heterogeneous interference. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 576–592. Springer.

Pujara, J.; Augustine, E.; and Getoor, L. 2017. Sparsity and noise: Where knowledge graph embeddings fall short. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, 1751–1756.

Algorithm 1 BaLu: Causal Inference over Incomplete Knowledge Graphs under Interference

Input: Knowledge Graph $G = (V, L, E)$, target class C, treatment p_T, outcome p_Y, hyperparameters $\alpha, \beta, \gamma, \eta$, network layers L and K (*L-layer* and *K-layer*, respectively)
Output: Estimated ITEs $\{\hat{\tau}_u\}_{u \in V_C}$ and ATE $\hat{\tau}_{\text{ATE}}$

// Graph Representation
$V_C \leftarrow \{e \in V \mid (e, \texttt{type}, C) \in E\}$ ▷ Units
$P_C^{\text{ctx}} \leftarrow P_C^{\text{dat}} \setminus \{p_T, p_Y\}$ ▷ Contextual attributes
$\mathcal{E}_{\text{rel}} \leftarrow \{(u, r, v) \mid u, v \in V_C, r \in P_C^{\text{rel}}, (u, r, v) \in E\}$
$\mathcal{E}_{\text{dat}} \leftarrow \{(u, p) \mid u \in V_C, p \in P_C^{\text{ctx}}, \exists (u, p, o) \in V\}$
$\mathcal{B} \leftarrow (V_C \cup P_C^{\text{ctx}},\, \mathcal{E}_{\text{dat}})$ ▷ Bipartite graph

// Initialization of Embeddings
$\forall u \in V_C, \mathbf{h}_u^{(0)} \leftarrow \mathbf{1} \in \mathbb{R}^h$ ▷ Units
$\forall p \in P_C^{\text{ctx}}, \mathbf{h}_p^{(0)} \leftarrow \textsc{OneHot}(p)$ ▷ Attributes
$\forall (u, p) \in \mathcal{E}_{\text{dat}}, \mathbf{e}_{u,p}^{(0)} \leftarrow \textsc{Embed}(\phi(u, p))$ ▷ Edges

// Data Imputation Component (*L-layer*)
for $l = 1$ **to** L **do**
    // Unit–Attribute Message Passing
    for each node $v \in V_C \cup P_C^{\text{ctx}}$ **do**
        $\mathbf{m}_v^{(l)} \leftarrow \text{AGG}\big\{\sigma\big(\mathbf{Q}_{\text{dat}}^{(l)} \cdot (\mathbf{h}_v^{(l-1)} \oplus \mathbf{e}_{u,v}^{(l-1)})\big) \mid (u, v) \in \mathcal{E}_{\text{dat}}\big\}$
        $\mathbf{h}_v^{(l)} \leftarrow \sigma\big(\mathbf{W}_{\text{dat}}^{(l)} \cdot (\mathbf{h}_v^{(l-1)} \oplus \mathbf{m}_v^{(l)})\big)$
    // Relational Message Passing
    for each unit $u \in V_C$ **do**
        $\mathbf{h}_u^{(l)} \leftarrow \text{Conv}\big(\mathbf{h}_u^{(l-1)}, \{\mathbf{h}_v^{(l-1)} \mid (u, r, v) \in \mathcal{E}_{\text{rel}}\}\big)$
    // Edge Update
    for each $(u, v) \in \mathcal{E}_{\text{dat}}$ **do**
        $\mathbf{e}_{u,v}^{(l)} \leftarrow \sigma\big(\mathbf{Q}_{\text{edge}}^{(l)} \cdot (\mathbf{e}_{u,v}^{(l-1)} \oplus \mathbf{h}_u^{(l)} \oplus \mathbf{h}_v^{(l)})\big)$

// Attribute Imputation
for each $(u, p) \notin \mathcal{E}_{\text{dat}}$ **do**
    $\hat{e}_{u,p} \leftarrow f_e(\mathbf{h}_u^{(L)} \oplus \mathbf{h}_p^{(L)})$

// Causal Estimation Component
for each unit $u \in V_C$ **do**
    $\mathbf{d}_u \leftarrow \bigoplus_{p \in P_C^{\text{ctx}}} \big[(1-\nu(u, p))\, \hat{e}_{u,p} + \nu(u, p)\, \phi(u, p)\big]$
    $\mathbf{x}_u \leftarrow \mathbf{d}_u \oplus \mathbf{h}_u^{(L)}$

// Interference Modeling (*K-layer*)
Initialize: $\forall u \in V_C, \mathbf{g}_u^{(0)} \leftarrow f_x(\mathbf{x}_u)$
for $k = 1$ **to** K **do**
    for each unit $u \in V_C$ **do**
        $\mathbf{g}_u^{(k)} \leftarrow \text{Conv}\big(\mathbf{g}_u^{(k-1)}, \{\mathbf{g}_v^{(k-1)} \mid (u, r, v) \in \mathcal{E}_{\text{rel}}\}\big)$

// Treatment Effect Estimation
for each unit $u \in V_C$ **do**
    $\hat{y}_u^{(0)} \leftarrow f_y^{(0)}(\mathbf{x}_u \oplus \mathbf{g}_u^{(K)})$
    $\hat{y}_u^{(1)} \leftarrow f_y^{(1)}(\mathbf{x}_u \oplus \mathbf{g}_u^{(K)})$
    $\hat{\tau}_u \leftarrow \hat{y}_u^{(1)} - \hat{y}_u^{(0)}$
$\hat{\tau}_{\text{ATE}} \leftarrow \frac{1}{|V_C|} \sum_{u \in V_C} \hat{\tau}_u$
return $\{\hat{\tau}_u\}_{u \in V_C},\ \hat{\tau}_{\text{ATE}}$