# Appendix (An Extended Version of ELsEA)

Qian Zhou Wei Chen Li Zhang Pengpeng Zhao Jiajie Xu Lei Zhao
*School of Computer Science and Technology, Soochow University*
qzhou1@stu.suda.edu.cn {robertchen,zhangliml,ppzhao,xujj,zhaol}@suda.edu.cn

### A. Removed Parameter Analysis

To investigate the details of ELsEA, we evaluate the impact of training seed ratio using RREA and the 100K datasets: IDS100K and DWY100K. From Figure 1, we have the following observations.

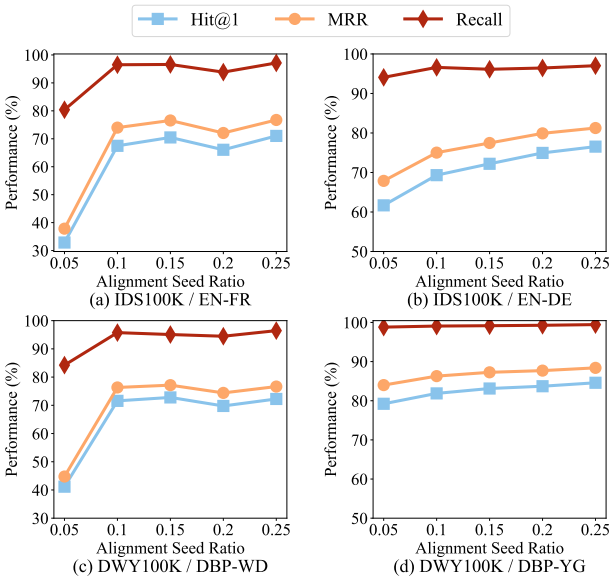


Fig. 1. Influence of training seed ratio on two cross-lingual datasets

In Fig. 1, an increased training seed ratio generally enhances the effectiveness of ELsEA. This highlights that training seeds are crucial for the LsEA performance, as they facilitate both the recall of ground-truth counterparts and effective EA. In addition, we find that when the training seed ratio is below 0.1, the overall LsEA performance declines sharply. This emphasizes the need for a sufficient quantity of training seeds to achieve optimal performance.

### B. Time Complexity of Algorithm 1

The time consumption of ELsEA primarily arises from several modules: a Metis-based weighted partitioner, a counterpart candidate generator, the VBIE module, the NE module, the MIPG strategy, and the neural EA approach.

Specifically, firstly, the time complexity of the Metis-based weighted partitioner and the neural EA approach is $\mathcal{O}(l+p+N\log(N))$ and $\mathcal{O}(\Psi)$, respectively, where $N$ is the sub-task number (with $N \leq 200$), $\mathcal{O}(\Psi)$ depends on the selected neural EA approach, $p = \max(|\mathcal{T}^s|, |\mathcal{T}^t|)$ and $l = \max(|\mathcal{E}^s|, |\mathcal{E}^t|)$. Notably, $|\mathcal{T}^s| \approx |\mathcal{T}^t|$ and $|\mathcal{E}^s| \approx |\mathcal{E}^t|$, hence we have $p = |\mathcal{T}^s|$ or $|\mathcal{T}^t|$ and $l = |\mathcal{E}^s|$ or $|\mathcal{E}^t|$.

Secondly, the time complexity of the counterpart candidate generator, the VBIE module, and the NE module is $\mathcal{O}(N \cdot q \cdot maxhop)$, $\mathcal{O}(N \cdot q)$ and $\mathcal{O}(N \cdot q^2)$, respectively, where $q = |\mathcal{N}^h|$, $maxhop \ll q$ and $q \approx \frac{|\mathcal{T}^s|}{N}$.

Thirdly, the main contributors to the time complexity of MIPG include Helsinki-NLP, MinHash-LSH, TransE, and Faiss. Their time complexities are $\mathcal{O}(n)$, $\mathcal{O}(l \cdot Top\text{-}K)$, $\mathcal{O}(p)$, and $\mathcal{O}(l \cdot \log(l))$, respectively, where $n = |\mathcal{E}^s| + |\mathcal{E}^t| \approx 2 \cdot l$ and $Top\text{-}K \ll l$. Therefore, the time complexity of MIPG is $\mathcal{O}(p + l \cdot (log(l) + 2))$.

Finally, the overall time complexity of the proposed model is $\mathcal{O}(p \cdot q + l \cdot (log(l) + Top\text{-}K) + \Psi)$.

### C. Details of MinHash-LSH Implementation

In this subsection, we briefly describe the details of MinHash-LSH used to filter out dissimilar pairs. Specifically, we first standardize each entity name by removing prefixes and punctuation, generating token sequences. Then, these token sequences are sorted to produce sorted token lists. Next, MinHash-LSH is employed using these sorted token lists to efficiently filter out dissimilar pairs, reducing the computational cost. Specifically, we use num_perm=128 hash functions to generate compact signatures, which are partitioned into b=8 bands and r=16 rows. The LSH index facilitates the retrieval of candidate pairs whose Jaccard similarity is above a predefined threshold (i.e., the lower bound of the string difference). Finally, we calculate the Levenshtein distance for each candidate to assess the similarity, generating accurate pseudo-seeds. For more details, please refer to LargeEA [2].

### D. Details of Baseline Implementation

The exact configuration of each baseline method is followed by their raw setting. Their raw links are added to the code repository, and the details are shown as follows:

- LargeEA [2]: https://github.com/ZJU-DAILY/LargeEA;
- ClusterEA [1]: https://github.com/xz-liu/ClusterEA;
- LargeGNN [7]: https://github.com/JadeXIN/LargeGNN;
- LIME [8]: https://github.com/DexterZeng/LIME;
- DivEA [3]: https://github.com/uqbingliu/DivEA.

The raw links of $\Psi$ are added to the code repository, and the details are presented as follows:

- GCN-Align [6]: https://github.com/1049451037/GCN-Align;
- RREA [5]: https://github.com/MaoXinn/RREA;
- Dual-AMN [4]: https://github.com/MaoXinn/Dual-AMN.

TABLE I
ABLATION RESULTS ON 1M (HITS@$k$, MRR, AND RECALL IN %; TIME IN SECONDS)

| EA approach | | IDS15K / EN-FR | | | | | | IDS15K / EN-DE | | | | | | IDS100K / EN-FR | | | | | | IDS100K / EN-DE | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Variants | Hit@1 | Hit@5 | MRR | Recall | Time | Mem. | Hit@1 | Hit@5 | MRR | Recall | Time | Mem. | Hit@1 | Hit@5 | MRR | Recall | Time | Mem. | Hit@1 | Hit@5 | MRR | Recall | Time | Mem. |
| GCN-Align | *w Random* | 4.83 | 11.90 | 8.23 | 26.03 | 1032.29 | 0.93 | 6.29 | 12.74 | 9.27 | 27.15 | 337.49 | 0.93 | 0.00 | 0.00 | 0.01 | 10.84 | 11570.70 | 1.54 | 2.20 | 4.21 | 3.18 | 11.56 | 13088.81 | 1.04 |
| | *w/o* $\mathcal{G}^{s,*}$ | 0.08 | 0.24 | 0.27 | 88.53 | 200.77 | 1.22 | 0.03 | 0.21 | 0.22 | 80.23 | 201.15 | 1.22 | 0.00 | 0.02 | 0.04 | 38.09 | 946.14 | 2.33 | 0.02 | 0.06 | 0.09 | 76.07 | 970.86 | 1.52 |
| | *w/o* $\mathcal{G}^{t,*}$ | 38.52 | 69.31 | 52.07 | 100.00 | 301.38 | 0.87 | 51.62 | 68.22 | 59.27 | 100.00 | 298.35 | 0.87 | 37.28 | 59.80 | 43.73 | 87.63 | 347.09 | 1.14 | 52.41 | 66.58 | 58.83 | 88.57 | 12299.74 | 2.57 |
| | *w neighbor* | 0.03 | 0.21 | 0.31 | 95.67 | 400.16 | 0.73 | 0.04 | 0.30 | 0.35 | 96.38 | 392.97 | 0.73 | 0.02 | 0.09 | 0.13 | 86.04 | 7321.60 | 1.22 | 0.02 | 0.09 | 0.11 | 89.34 | 4129.26 | 0.85 |
| | ELsEA | 52.07 | 85.01 | 65.98 | 96.98 | 324.95 | 1.17 | 78.02 | 89.40 | 83.02 | 97.97 | 317.93 | 1.37 | 39.02 | 59.42 | 48.17 | 86.63 | 2332.35 | 2.72 | 53.32 | 67.48 | 59.74 | 89.60 | 2219.90 | 2.72 |
| RREA | *w Random* | 20.57 | 25.39 | 22.92 | 37.24 | 905.42 | 0.71 | 25.98 | 30.11 | 28.06 | 42.82 | 642.34 | 0.71 | 0.01 | 0.02 | 0.02 | 12.08 | 3835.81 | 1.90 | 13.09 | 14.58 | 13.86 | 21.59 | 6856.80 | 1.90 |
| | *w/o* $\mathcal{G}^{s,*}$ | 0.21 | 0.65 | 0.56 | 90.25 | 192.42 | 0.97 | 0.18 | 0.45 | 0.48 | 88.61 | 193.37 | 0.97 | 0.02 | 0.10 | 0.13 | 45.40 | 3404.32 | 3.50 | 68.77 | 79.21 | 73.50 | 94.49 | 2743.40 | 3.51 |
| | *w/o* $\mathcal{G}^{t,*}$ | 64.62 | 79.98 | 71.36 | 100.00 | 340.17 | 0.73 | 68.02 | 82.05 | 74.43 | 100.00 | 347.41 | 0.71 | 0.04 | 0.14 | 0.16 | 40.23 | 3806.74 | 3.50 | 68.61 | 78.94 | 73.30 | 94.19 | 2822.61 | 3.51 |
| | *w neighbor* | 62.39 | 78.75 | 69.78 | 97.64 | 291.80 | 0.68 | 65.75 | 79.74 | 72.24 | 98.78 | 282.33 | 0.68 | 0.06 | 0.25 | 0.29 | 93.11 | 4883.19 | 1.91 | 62.91 | 75.54 | 68.78 | 94.28 | 4279.89 | 1.91 |
| | ELsEA | 79.85 | 92.91 | 85.49 | 98.94 | 349.98 | 1.08 | 83.09 | 89.83 | 86.18 | 98.14 | 348.68 | 1.08 | 64.74 | 78.34 | 70.82 | 93.86 | 4806.47 | 3.84 | 68.92 | 79.24 | 73.64 | 94.49 | 2556.76 | 3.85 |
| Dual-AMN | *w Random* | 33.12 | 36.36 | 34.99 | 53.10 | 1207.83 | 1.78 | 38.17 | 40.31 | 39.69 | 64.07 | 596.04 | 1.78 | 0.04 | 0.06 | 0.06 | 25.25 | 5975.12 | 4.41 | 19.72 | 20.49 | 20.32 | 35.33 | 6763.60 | 4.83 |
| | *w/o* $\mathcal{G}^{s,*}$ | 0.18 | 0.50 | 0.42 | 98.13 | 476.87 | 2.05 | 0.14 | 0.32 | 0.32 | 98.05 | 580.38 | 2.85 | 0.05 | 0.13 | 0.14 | 87.68 | 3536.89 | 6.44 | 74.63 | 81.15 | 77.67 | 97.02 | 4064.83 | 7.24 |
| | *w/o* $\mathcal{G}^{t,*}$ | 67.27 | 79.10 | 72.50 | 100.00 | 841.49 | 2.51 | 63.07 | 71.57 | 67.17 | 100.00 | 858.22 | 2.51 | 0.04 | 0.11 | 0.12 | 88.16 | 3136.34 | 5.64 | 74.34 | 80.84 | 77.41 | 96.95 | 4476.47 | 8.04 |
| | *w neighbor* | 64.88 | 75.66 | 69.83 | 99.08 | 806.19 | 2.76 | 70.46 | 80.37 | 75.01 | 99.17 | 805.70 | 2.76 | 0.05 | 0.16 | 0.18 | 97.28 | 20659.92 | 8.84 | 73.16 | 80.31 | 76.52 | 99.45 | 15322.87 | 8.04 |
| | ELsEA | 75.22 | 86.13 | 79.99 | 98.90 | 776.69 | 2.92 | 90.47 | 93.78 | 91.99 | 98.78 | 1064.52 | 2.76 | 68.00 | 78.01 | 72.58 | 97.10 | 4554.82 | 8.21 | 74.65 | 81.26 | 77.73 | 96.98 | 5919.38 | 7.42 |

TABLE II
ABLATION RESULTS ON CROSS-LINGUAL/MONOLINGUAL DATASETS WITH 15K AND 2M (HITS@$k$, MRR, AND RECALL IN %; TIME IN SECONDS FOR 15K, HOURS FOR 2M)

| EA approach | | DBP15K / ZH-EN | | | | | | DBP15K / FR-EN | | | | | | DBP15K / JA-EN | | | | | | FBDBP2M / FB-DBP | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Variants | Hit@1 | Hit@5 | MRR | Recall | Time | Mem. | Hit@1 | Hit@5 | MRR | Recall | Time | Mem. | Hit@1 | Hit@5 | MRR | Recall | Time | Mem. | Hit@1 | Hit@5 | MRR | Recall | Time | Mem. |
| GCN-Align | *w Random* | 8.72 | 18.99 | 13.30 | 32.81 | 1040.86 | 0.95 | 9.53 | 21.40 | 14.81 | 34.92 | 1165.75 | 0.95 | 7.43 | 17.84 | 12.06 | 32.58 | 1090.83 | 0.95 | 0.07 | 0.17 | 0.11 | 0.62 | 153163.89 | 1.76 |
| | *w/o* $\mathcal{G}^{s,*}$ | 0.04 | 0.13 | 0.19 | 88.96 | 197.04 | 1.23 | 0.03 | 0.17 | 0.19 | 76.13 | 197.88 | 1.24 | 0.06 | 0.19 | 0.22 | 74.33 | 217.52 | 1.24 | 0.00 | 0.01 | 0.02 | 25.59 | 148957.01 | 3.28 |
| | *w/o* $\mathcal{G}^{t,*}$ | 36.17 | 64.28 | 48.53 | 97.02 | 299.86 | 0.87 | 36.69 | 66.97 | 49.99 | 96.56 | 343.07 | 0.82 | 37.44 | 66.04 | 50.06 | 99.27 | 306.39 | 0.87 | 3.10 | 8.97 | 6.24 | 54.01 | 94225.96 | 4.88 |
| | *w neighbor* | 0.09 | 0.44 | 0.43 | 95.99 | 291.37 | 0.74 | 0.13 | 0.48 | 0.51 | 92.08 | 293.30 | 0.74 | 0.07 | 0.34 | 0.34 | 94.62 | 303.81 | 0.74 | 0.01 | 0.04 | 0.06 | 55.28 | 91330.22 | 0.71 |
| | ELsEA | 37.13 | 65.28 | 49.42 | 94.79 | 487.80 | 1.14 | 37.91 | 68.15 | 51.20 | 96.87 | 373.75 | 1.08 | 38.48 | 67.00 | 50.93 | 94.83 | 347.09 | 1.14 | 8.42 | 16.17 | 12.34 | 49.66 | 70335.54 | 4.47 |
| RREA | *w Random* | 32.87 | 37.76 | 35.14 | 46.05 | 831.54 | 1.10 | 35.41 | 41.52 | 38.17 | 49.69 | 695.95 | 1.10 | 26.99 | 32.32 | 29.54 | 42.61 | 365.00 | 1.10 | 0.31 | 0.36 | 0.34 | 0.81 | 162637.00 | 9.05 |
| | *w/o* $\mathcal{G}^{s,*}$ | 0.12 | 0.31 | 0.29 | 92.99 | 233.16 | 0.97 | 0.08 | 0.34 | 0.28 | 94.71 | 245.30 | 1.37 | 0.12 | 0.40 | 0.34 | 93.68 | 224.59 | 0.97 | 0.01 | 0.02 | 0.04 | 29.90 | 144140.47 | 7.91 |
| | *w/o* $\mathcal{G}^{t,*}$ | 66.73 | 83.27 | 74.03 | 97.95 | 421.34 | 0.00 | 69.53 | 86.94 | 77.11 | 99.00 | 398.53 | 0.83 | 67.47 | 84.82 | 75.01 | 100.00 | 388.39 | 0.83 | 9.11 | 15.82 | 12.63 | 68.06 | 94405.14 | 8.11 |
| | *w neighbor* | 65.78 | 83.68 | 73.73 | 97.07 | 319.35 | 0.68 | 66.89 | 85.38 | 75.07 | 99.08 | 331.27 | 1.08 | 65.73 | 84.26 | 73.87 | 97.07 | 322.92 | 0.68 | 9.15 | 16.27 | 12.89 | 68.09 | 94555.25 | 4.11 |
| | ELsEA | 67.04 | 83.42 | 74.19 | 97.29 | 396.32 | 1.40 | 70.21 | 87.28 | 77.58 | 98.32 | 427.87 | 1.39 | 67.61 | 84.14 | 74.82 | 97.83 | 406.48 | 1.40 | 22.20 | 31.76 | 26.96 | 66.06 | 80631.44 | 13.16 |
| Dual-AMN | *w Random* | 41.62 | 43.98 | 42.93 | 58.27 | 675.03 | 2.78 | 49.72 | 52.77 | 51.38 | 66.01 | 736.54 | 2.78 | 37.69 | 39.82 | 39.10 | 56.36 | 738.04 | 2.78 | 0.47 | 0.48 | 0.47 | 0.93 | 155829.87 | 5.59 |
| | *w/o* $\mathcal{G}^{s,*}$ | 0.18 | 0.40 | 0.36 | 98.45 | 551.89 | 2.65 | 0.14 | 0.33 | 0.32 | 98.84 | 626.15 | 2.85 | 0.12 | 0.34 | 0.30 | 98.58 | 576.33 | 2.65 | 0.01 | 0.02 | 0.04 | 29.90 | 144140.47 | 7.91 |
| | *w/o* $\mathcal{G}^{t,*}$ | 67.48 | 77.21 | 71.98 | 96.95 | 930.41 | 2.54 | 72.57 | 83.11 | 77.31 | 96.51 | 1254.89 | 2.54 | 67.43 | 77.98 | 72.26 | 99.28 | 1267.75 | 2.54 | 8.13 | 11.51 | 10.15 | 71.51 | 92187.27 | 8.64 |
| | *w neighbor* | 67.15 | 77.52 | 71.79 | 99.28 | 786.82 | 2.78 | 71.97 | 81.79 | 76.43 | 80.39 | 1193.17 | 2.78 | 67.43 | 77.86 | 72.10 | 98.81 | 868.90 | 2.78 | 8.13 | 11.51 | 10.15 | 70.51 | 91155.25 | 8.64 |
| | ELsEA | 69.75 | 78.39 | 73.77 | 98.67 | 830.16 | 2.94 | 73.92 | 83.97 | 78.39 | 99.15 | 1172.20 | 2.94 | 69.54 | 79.01 | 73.92 | 98.95 | 860.32 | 2.94 | 21.97 | 27.00 | 24.69 | 69.30 | 66651.44 | 10.04 |

TABLE III
ABLATION RESULTS ON CROSS-LINGUAL/MONOLINGUAL DATASETS WITH 100K AND 1M (HITS@$k$, MRR, AND RECALL IN %; TIME IN HOURS)

| EA approach | | DWY100K / DBP-WD | | | | | | DWY100K / DBP-YG | | | | | | DBP1M / EN-FR | | | | | | DBP1M / EN-DE | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Variants | Hit@1 | Hit@5 | MRR | Recall | Time | Mem. | Hit@1 | Hit@5 | MRR | Recall | Time | Mem. | Hit@1 | Hit@5 | MRR | Recall | Time | Mem. | Hit@1 | Hit@5 | MRR | Recall | Time | Mem. |
| GCN-Align | *w Random* | 1.16 | 2.90 | 2.01 | 11.03 | 11409.90 | 1.02 | 4.23 | 6.38 | 5.28 | 11.91 | 12946.56 | 1.04 | 0.25 | 0.64 | 0.43 | 2.06 | 60018.55 | 1.68 | 0.37 | 0.79 | 0.56 | 1.98 | 48005.38 | 1.59 |
| | *w/o* $\mathcal{G}^{s,*}$ | 0.01 | 0.06 | 0.08 | 77.83 | 850.47 | 1.52 | 0.03 | 0.07 | 0.10 | 88.40 | 877.98 | 1.48 | 7.47 | 15.39 | 11.46 | 54.13 | 55943.17 | 2.28 | 5.39 | 11.91 | 8.59 | 44.64 | 47378.42 | 1.63 |
| | *w/o* $\mathcal{G}^{t,*}$ | 44.14 | 66.28 | 54.10 | 100.00 | 1966.34 | 2.27 | 61.57 | 78.72 | 69.34 | 100.00 | 2115.60 | 3.70 | 8.61 | 18.76 | 13.79 | 64.19 | 40552.54 | 1.49 | 6.87 | 16.15 | 11.50 | 58.00 | 34251.47 | 1.00 |
| | *w neighbor* | 0.04 | 0.17 | 0.18 | 96.45 | 6220.12 | 0.85 | 0.03 | 0.19 | 0.18 | 100.00 | 1422.15 | 0.85 | 0.01 | 0.05 | 0.08 | 64.42 | 38494.16 | 0.58 | 6.47 | 14.39 | 10.46 | 53.13 | 55943.17 | 2.28 |
| | ELsEA | 51.32 | 73.24 | 61.02 | 93.14 | 3493.62 | 4.09 | 67.10 | 83.19 | 74.31 | 98.59 | 6249.83 | 4.09 | 8.52 | 17.30 | 12.94 | 60.24 | 66508.77 | 2.63 | 7.78 | 15.36 | 11.47 | 51.16 | 49979.40 | 3.01 |
| RREA | *w Random* | 8.00 | 9.16 | 8.60 | 15.89 | 2991.62 | 1.90 | 15.82 | 17.18 | 16.50 | 20.33 | 3216.15 | 1.90 | 1.61 | 1.73 | 1.67 | 3.11 | 58758.84 | 2.45 | 1.89 | 1.99 | 1.94 | 3.11 | 53677.53 | 2.34 |
| | *w/o* $\mathcal{G}^{s,*}$ | 67.34 | 79.44 | 72.79 | 94.10 | 3923.12 | 1.90 | 79.84 | 90.04 | 84.45 | 98.92 | 3534.15 | 3.51 | 0.05 | 0.12 | 0.11 | 51.39 | 43891.38 | 2.44 | 0.05 | 0.12 | 0.11 | 42.80 | 38741.17 | 1.55 |
| | *w/o* $\mathcal{G}^{t,*}$ | 67.42 | 79.50 | 72.86 | 94.01 | 2640.36 | 3.50 | 79.85 | 90.12 | 84.46 | 98.95 | 2562.41 | 3.51 | 20.03 | 29.42 | 24.95 | 70.32 | 44175.87 | 3.24 | 17.79 | 26.48 | 22.32 | 65.16 | 38306.07 | 1.63 |
| | *w neighbor* | 67.56 | 81.12 | 73.78 | 99.04 | 11186.62 | 1.11 | 78.09 | 88.96 | 83.06 | 100.00 | 10258.38 | 1.91 | 19.22 | 29.74 | 24.72 | 69.34 | 33687.24 | 2.06 | 17.35 | 26.33 | 21.54 | 64.45 | 27313.06 | 1.04 |
| | ELsEA | 74.95 | 86.05 | 79.89 | 96.40 | 4399.02 | 6.36 | 83.86 | 92.52 | 87.77 | 99.28 | 5476.33 | 6.36 | 20.58 | 28.78 | 24.80 | 65.45 | 61893.85 | 10.08 | 18.87 | 25.83 | 22.37 | 58.63 | 39683.71 | 3.91 |
| Dual-AMN | *w Random* | 11.57 | 12.01 | 11.92 | 24.02 | 5578.10 | 4.83 | 20.46 | 21.25 | 20.99 | 27.82 | 2704.93 | 4.83 | 1.88 | 1.91 | 1.90 | 3.94 | 59730.79 | 5.38 | 2.05 | 2.07 | 2.07 | 3.35 | 46721.99 | 4.82 |
| | *w/o* $\mathcal{G}^{s,*}$ | 70.29 | 79.59 | 74.52 | 96.56 | 7705.77 | 4.84 | 80.80 | 88.18 | 84.18 | 99.27 | 7546.28 | 4.84 | 19.97 | 22.76 | 21.77 | 68.54 | 60995.74 | 7.77 | 16.47 | 18.51 | 17.79 | 59.31 | 52221.48 | 5.26 |
| | *w/o* $\mathcal{G}^{t,*}$ | 70.33 | 79.72 | 74.55 | 96.51 | 8570.08 | 4.84 | 80.86 | 88.28 | 84.23 | 99.28 | 7941.33 | 4.84 | 18.28 | 21.85 | 20.61 | 71.35 | 47846.18 | 7.77 | 15.69 | 18.38 | 17.50 | 61.12 | 39609.31 | 4.57 |
| | *w neighbor* | 71.69 | 82.18 | 76.43 | 99.17 | 13963.57 | 4.84 | 81.79 | 89.43 | 85.30 | 99.28 | 11387.41 | 4.84 | 18.38 | 24.62 | 21.82 | 70.00 | 21556.92 | 6.57 | 15.85 | 20.10 | 18.37 | 67.00 | 23067.90 | 4.57 |
| | ELsEA | 77.57 | 86.09 | 81.34 | 97.59 | 9065.34 | 8.82 | 84.00 | 90.79 | 87.09 | 99.45 | 8604.09 | 8.82 | 21.20 | 24.00 | 23.01 | 70.35 | 54715.29 | 9.29 | 18.50 | 20.34 | 19.70 | 59.02 | 46829.67 | 9.24 |

### *E. Details of Quantitative Evidence*

Observed from the tables with overall results, we present the following quantitative evidence regarding the recall of counterparts and EA effectiveness. !) For the recall of counterparts, compared to other models, ELsEA presents at least a 0.28%, 0.27%, 0.21%, 0.81%, 3.77%, and 3.34% increase in Recall on IDS15K/EN-FR, DBP15K/ZH-EN, DBP15K/FR-EN, IDS100K/EN-FR, IDS100K/EN-DE, and DBP1M/EN-DE, respectively. Additionally, ELsEA obtains approaching recall performance on IDS15K/EN-DE, DBP15K/JA-EN, DWY100K/DBP-WD, DWY100K/DBP-YG, DBP1M/EN-FR, DBP1M/EN-DE, and FBDBP2M/FB-DBP. This is because

DivEA already achieves extremely high recall on these datasets (e.g., over 97% on DBP15K/ZH-EN using RREA), making further improvement challenging. However, ELsEA may still offer a subtle advantage in counterpart recall. 2) For EA effectiveness, compared to other models, ELsEA achieves at least a 7.5% and 4.1% increase in Hit@1 (6% and 4% in MRR) on IDS15K/EN-FR and IDS15K/EN-DE, respectively. On DBP15K/ZH-EN, DBP15K/FR-EN, and DBP15K/JA-EN, it presents a minimum increase of 1.8%, 0.83%, and 1.44% in Hit@1 (1.5%, 2%, and 1.46% in MRR), respectively. On IDS100K/EN-FR, IDS100K/EN-DE, DWY100K/DBP-WD, and DWY100K/DBP-YG, ELsEA shows the increase of 4.37%, 3.26%, 2.80%, and 0.28% in Hit@1 (4.23%, 3.91%, 2.74% and 0.24% in MRR), respectively. Finally, on DBP1M/EN-FR, DBP1M/EN-DE, and FBDBP2M/FB-DBP, ELsEA achieves at least a 1.30%, 0.95%, and 0.99% increase in Hit@1 (1.00%, 0.43%, and 1.06% in MRR), respectively. These results highlight the effectiveness of our sub-task partition strategy in leveraging critical structures when generating sub-tasks, as well as the importance of context graph generators (VBIE and NE) in integrating high-quality context.

### *F. Complete Ablation Study*

To get further insights into the proposed method ELsEA, we conduct ablation experiments on all datasets, as shown in tables I, II and III.

## REFERENCES


[1] Y. Gao, X. Liu, J. Wu, T. Li, P. Wang, and L. Chen, "ClusterEA: Scalable entity alignment with stochastic training and normalized mini-batch similarities," in *CIKM*, 2022, pp. 421–431.

[2] C. Ge, X. Liu, L. Chen, B. Zheng, and Y. Gao, "LargeEA: Aligning entities for large-scale knowledge graphs," *Proc. VLDB Endow.*, vol. 15, no. 2, pp. 237–245, 2021.

[3] B. Liu, W. Hua, G. Zuccon, G. Zhao, and X. Zhang, "High-quality task division for large-scale entity alignment," in *CIKM*, 2022, pp. 1258–1268.

[4] X. Mao, W. Wang, Y. Wu, and M. Lan, "Boosting the speed of entity alignment 10×: Dual attention matching network with normalized hard sample mining," in *WWW*, 2021, pp. 821–832.

[5] X. Mao, W. Wang, H. Xu, Y. Wu, and M. Lan, "Relational reflection entity alignment," in *CIKM*, 2020, pp. 1095–1104.

[6] Z. Wang, Q. Lv, X. Lan, and Y. Zhang, "Cross-lingual knowledge graph alignment via graph convolutional networks," in *EMNLP*, 2018, pp. 349–357.

[7] K. Xin, Z. Sun, W. Hua, W. Hu, J. Qu, and X. Zhou, "Large-scale entity alignment via knowledge graph merging, partitioning and embedding," in *CIKM*, 2022, pp. 2240–2249.

[8] W. Zeng, X. Zhao, X. Li, J. Tang, and W. Wang, "On entity alignment at scale," *The VLDB Journal*, vol. 31, no. 5, pp. 1009–1033, 2022.