A Critical Re-Evaluation of Neural Methods for Entity Alignment

Manuel Leone,* Stefano Huber,* Akhil Arora,* Alberto García Durán,* and Robert West
Data Science Laboratory (DLAB), EPFL

Why is Entity alignment needed?

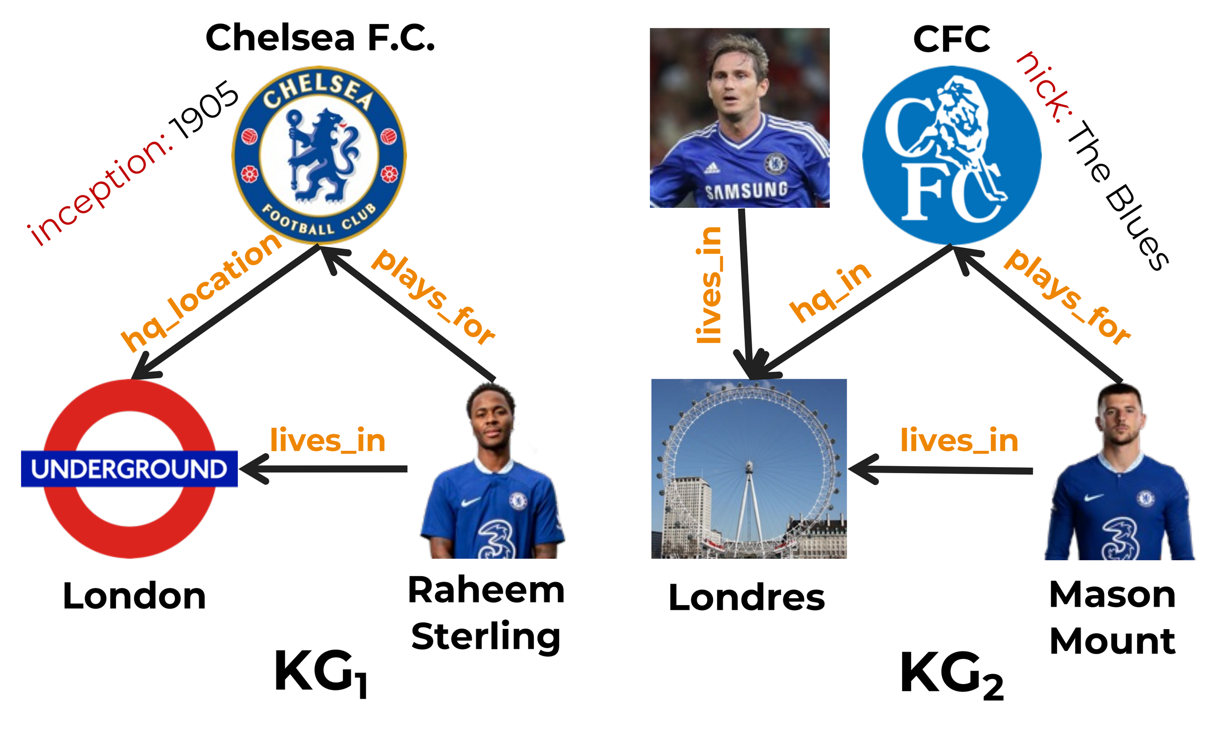
KGs are critical in today's world

- Search
- Recommender systems
- Question answering
- Conversational agents & digital assistants

KGs are seldom complete

- Contain overlapping as well as complementary information

Entity alignment in KGs



Input: KG_1 and KG_2

[Optional] **Seed** alignments

(KG_1:Chelsea F.C., same_as, KG_2:CFC)

Output: Aligned entity pairs

- ✓ **(KG_1:London, same_as, KG_2:Londres)**
- ✗ **(KG_1:Raheem Sterling, same_as, KG_2:Mason Mount)**

Why all the fuss?

- Hundreds of papers: semantic web, data management, NLP/IR/ML
- Endogamic comparisons!

1-to-1 Assumption: The key culprit

- Assumes partial knowledge of the ground truth
  - **KG_1:Raheem Sterling, same_as, KG_2:?**
- Unfair and Unrealistic

Key research questions

- Is the evaluation setup employed by neural EA methods meaningful?
- What is the true progress achieved on account of neural EA?
- What lies in the future for neural EA?

Towards a realistic evaluation setup

Type	Scope	Main Characteristics
OpenEA	Primary	1-to-1 assumption.
RealEA	Primary	no 1-to-1 assumption.
XRealEA	Primary	no 1-to-1 assumption, cross-lingual.
SupRealEA	Ablation	no 1-to-1 assumption, varying amount of supervision.
AttRealEA	Ablation	no 1-to-1 assumption, varying amount of attributes.
SpaRealEA	Ablation	no 1-to-1 assumption, sparser KG.
RealEA_NoObfs	Ablation	no 1-to-1 assumption, non-obfuscated URIs.
XRealEA_Pure	Ablation	no 1-to-1 assumption, purely cross-lingual.

- ✓ Approximate degree distribution of original KGs
- ✓ No 1-to-1 assumption
- ✓ Semantics free Entity URIs
- ✓ Appropriate metrics: Prec/Rec/F1 instead of Hits@k/MRR

Datasets, datasets, datasets …

	RealEA			
Dataset	DB-YG-15K	DB-WD-15K	DB-YG-100K	DB-WD-100K
#Entities	19,865 - 21,050	20,038 - 19,581	126,145 - 136,211	129,847 - 137,721
#Relations	290 - 32	306 - 214	386 - 32	456 - 329
#Attributes	247 - 34	307 - 490	366 - 38	478 - 785
#Rel. Triples	60,329 - 82,109	50,007 - 65,017	479,510 - 653,261	399,061 - 489,698
#Att. Triples	129,330 - 392,845	85,331 - 112,786	677,721 - 1,427,545	566,073 - 668,925
#Matchable Ent.	15,000	15,000	100,000	100,000

Results: RealEA

		DB-YG-15K (RealEA)			DB-WD-15K (RealEA)			DB-YG-100K (RealEA)			DB-WD-100K (RealEA)		
Category	**Method**	**Precision**	**Recall**	**F_1-score**	**Precision**	**Recall**	**F_1-score**	**Precision**	**Recall**	**F_1-score**	**Precision**	**Recall**	**F_1-score**
Neural (EA)	BootEA	0.459 ± 0.008	0.313 ± 0.009	0.372 ± 0.007	0.609 ± 0.007	0.280 ± 0.009	0.383 ± 0.008	0.671 ± 0.005	0.487 ± 0.004	0.565 ± 0.003	0.548 ± 0.008	0.272 ± 0.007	0.363 ± 0.006
Neural (EA)	RDGCN	0.822 ± 0.003	0.709 ± 0.004	0.761 ± 0.003	0.583 ± 0.012	0.242 ± 0.009	0.342 ± 0.011	0.846 ± 0.001	0.708 ± 0.002	0.771 ± 0.001	0.538 ± 0.003	0.203 ± 0.001	0.295 ± 0.001
Neural (EA)	BERT-INT	0.817 ± 0.001	0.827 ± 0.004	0.822 ± 0.002	0.604 ± 0.030	0.075 ± 0.006	0.134 ± 0.010	0.841 ± 0.001	0.865 ± 0.006	0.853 ± 0.003	0.698 ± 0.009	0.120 ± 0.002	0.206 ± 0.003
Neural (EA)	TransEdge	0.335 ± 0.025	0.203 ± 0.017	0.253 ± 0.020	0.589 ± 0.126	0.183 ± 0.034	0.279 ± 0.054	0.566 ± 0.011	0.438 ± 0.018	0.494 ± 0.016	0.339 ± 0.041	0.147 ± 0.012	0.205 ± 0.018
Neural (RL)	DMatch	0.851 ± 0.023	0.787 ± 0.014	0.821 ± 0.012	0.234 ± 0.009	0.162 ± 0.011	0.186 ± 0.013	0.878 ± 0.008	0.691 ± 0.007	0.773 ± 0.012	0.048 ± 0.021	0.344 ± 0.000	0.092 ± 0.014
Neural (RL)	Ditto	0.873 ± 0.012	0.821 ± 0.014	0.838 ± 0.003	0.339 ± 0.017	0.214 ± 0.004	0.262 ± 0.002	**0.916 ± 0.011**	0.682 ± 0.001	0.784 ± 0.011	0.757 ± 0.012	0.248 ± 0.009	0.376 ± 0.008
Non-neural (EA)	Paris+	**0.906 ± 0.000** †	**0.931 ± 0.001**†	**0.918 ± 0.001**	**0.928 ± 0.002**†	**0.551 ± 0.004**†	**0.691 ± 0.003**†	**0.923 ± 0.000**†	**0.939 ± 0.000**†	**0.931 ± 0.000**†	**0.927 ± 0.001**†	**0.615 ± 0.001**†	**0.740 ± 0.001**†

Takeaways

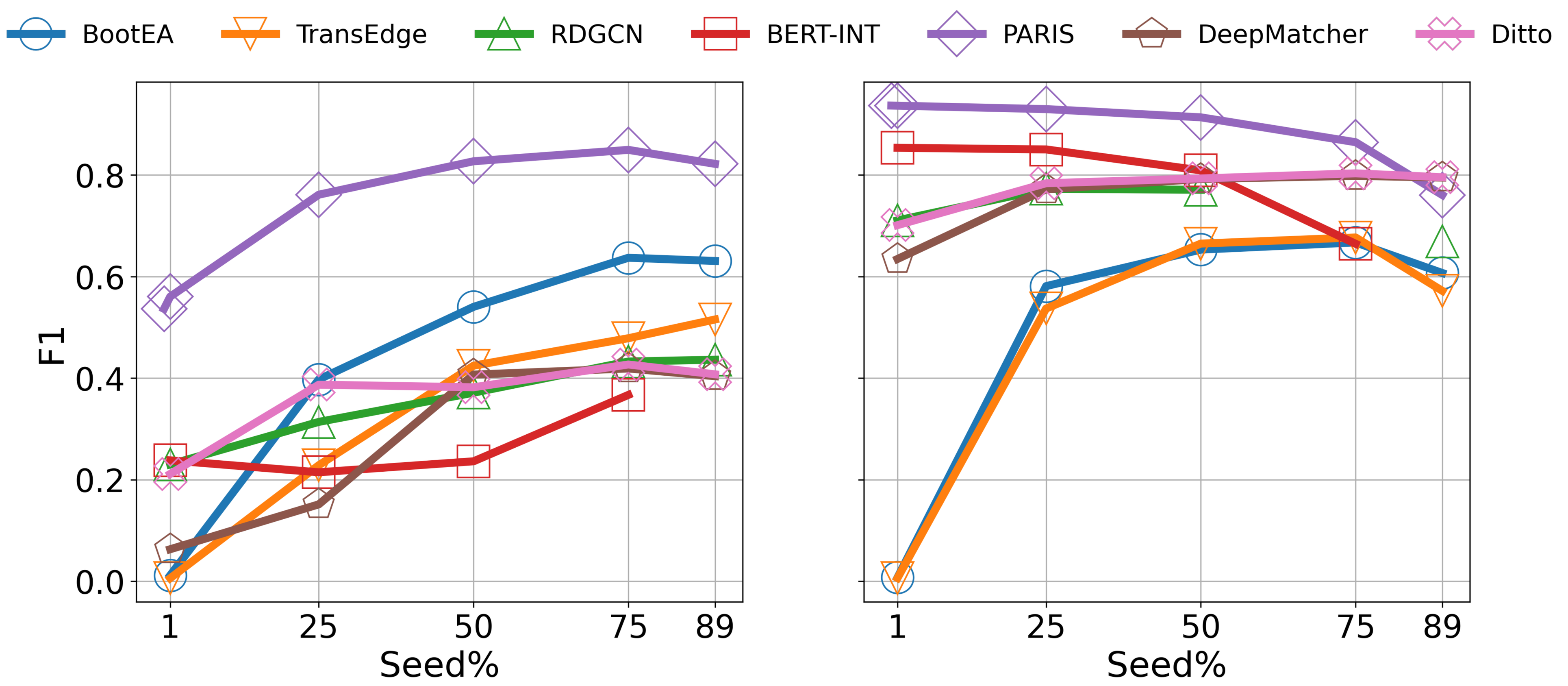
PARIS+ is the best EA method till date

- Statistically significantly better in quality
- Several orders of magnitude faster

ER/RL methods can perform EA reasonably well

- Not as good as PARIS+, but competitive to neural EA

Neural EA methods need to be repositioned to showcase their true potential



(a) DB-WD-100K (b) DB-YG-100K

Broader Impact

A nudge towards the end of endogamic comparisons

- Encouraging other communities to follow suit!

GitHub

This work was presented at the 48th International Conference on Very Large Databases (VLDB), September 5–9, 2022, Sydney, Australia.

VLDB 2022