Structural Bias in Knowledge Graphs for the Entity Alignment Task

Nikolaos Fanourakis[1], Vasilis Efthymiou[1], Vassilis Christophides[2], Dimitris Kotzinos[2], Evaggelia Pitoura[3], and Kostas Stefanidis[4]

[1] FORTH-ICS, Greece
{fanourakis,vefthym}@ics.forth.gr
[2] Lab. ETIS, CY Cergy Paris University, ENSEA, CNRS UMR 8051, France
Vassilis.Christophides@ensea.fr, Dimitrios.Kotzinos@cyu.fr
[3] University of Ioannina, Greece
pitoura@uoi.gr
[4] Tampere University, Finland
konstantinos.stefanidis@tuni.fi

Abstract. Knowledge Graphs (KGs) have recently gained attention for representing knowledge about a particular domain and play a central role in a multitude of AI tasks like recommendations and query answering. Recent works have revealed that KG embedding methods used to implement these tasks often exhibit direct forms of bias (e.g., related to gender, nationality, etc.) leading to discrimination. In this work, we are interested in the impact of indirect forms of bias related to the structural diversity of KGs in entity alignment (EA) tasks. In this respect, we propose an exploration-based sampling algorithm, SUSIE, that generates challenging benchmark data for EA methods, with respect to structural diversity. SUSIE requires setting the value of a single hyperparameter, which affects the connectivity of the generated KGs. The generated samples exhibit similar characteristics to some of the most challenging real-world KGs for EA tasks. Using our sampling, we demonstrate that state-of-the-art EA methods, like RREA, RDGCN, MultiKE and PARIS, exhibit different robustness to structurally diverse input KGs.

Keywords: Knowledge Graphs · Entity Alignment · Structural Bias.

1 Introduction

Knowledge Graphs (KGs) provide interlinked descriptions of real-world entities (e.g., persons, places, etc.) that play a central role in a multitude of AI tasks like recommendations [25] and query answering [37]. Recently, graph representation learning techniques have been used to automate several KG construction tasks, such as link prediction [46,42], node classification [41,52], and entity alignment (EA) [51,22]. The key idea of these methods is to embed the nodes (entities) and the edges (relations or attributes) of a KG into a low-dimensional vector space in such a way that similar entities in the original KG are close to each other in the embedding space, while dissimilar entities lie far from each other [38,57,56].

However, recent studies have shown that KG embeddings may reflect or even amplify biases that exist in the original KGs, for example biases related to gender, nationality, or popularity [45,8,9]. In this paper, we focus on a different type of bias. Specifically, we focus on whether structural characteristics of the original KGs introduce biases in the KG entity alignment (EA) task used to find pairs of nodes in two input KGs that refer to the same real-world entity [51,16].

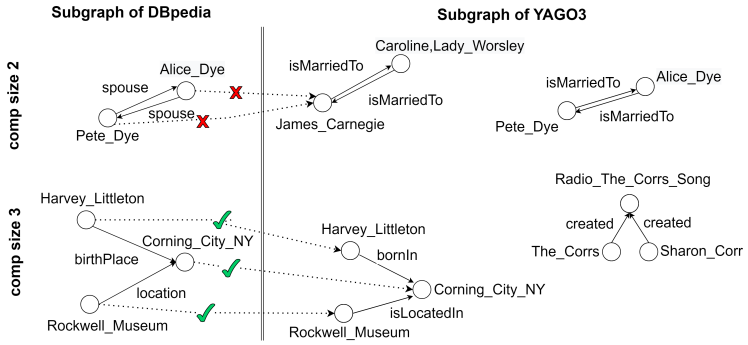


Fig. 1: Correct (check) and incorrect (X) matches suggested by RREA [38] for nodes belonging to connected components of different sizes on the D-Y dataset.

Several methods have been proposed for the EA task that rely on factual (attribute based) or structural (relation based) information of the entities in the KGs (e.g, [38,53,13]). Experimental studies have shown that their performance depends on the factual and structural heterogeneity of the input KGs [51,22]. Since most state-of-the-art EA methods exploit the structural characteristics of entities [22], we focus on structural diversity. Embedding-based EA methods seem to favor the alignment of entities from *rich structural neighborhoods in the two KGs [22,51].*

Example 1. Consider the two KGs (DBpedia and YAGO3) of Figure 1, where it is seemingly trivial to align entities based on their labels (e.g., "Pete Dye" and "Alice Dye" in DBpedia should be aligned to "Pete Dye" and "Alice Dye", respectively, in YAGO). RREA [38], a state-of-the-art KG embedding-based EA method, fails to find some of these trivial matches and instead, maps Alice_Dye to James_Carnegie and also Pete_Dye to James_Carnegie, as shown by the dotted lines (RREA mappings), as these entities exhibit the same structural cues. In contrast, RREA, as a relation-based EA method that exploits the structural similarity of entities, matches correctly all the entities in a component of size three (Harvey Littleton, Rockwell Museum and Corning City NY), where the structural information is richer [52]. The point is that EA systems face difficulties to correctly match entities when they belong to small graph components, since there are many more entities that have similar structures, unless we consider additional similarity evidence sources than just the entity structure. Figure 1 illustrates this problem over entity subgraphs of two real KGs.

In many cases, structural bias in a KG can be seen as an instance of indirect bias against protected groups defined over sensitive attributes (e.g., gender, race), since structural bias often reflects sampling and representation bias [7], where members of protected groups are incompletely described. This is because missing relations and values in KGs are frequent manifestations of several latent causes: protected group members are more reluctant to provide information that could be used against them, sensitive information may be erased by human curators, or data acquisition may be less complete for protected groups [39].

Following previous work, we quantify structural diversity relying on the number and size of connected components and on node degrees [27,26,47,59]. To generate KGs with adjustable levels of structural diversity for the EA task, we propose an exploration-based sampling method (SUSIE). Our sampling method directly controls the number of connected components, while component sizes and node degrees are affected indirectly. Our evaluation shows that state-of-the-art KG embedding-based EA methods exhibit indirect bias, due to structural diversity, against smaller, less connected regions of the benchmark datasets.

Unlike existing benchmarking (e.g., [15,18,58]) and sampling methods (e.g., Fairwalk [45], IDS [51], div2vec [28]), our exploration-based sampling produces EA datasets (i.e., two KGs and their alignment) of varying structural diversity. As highlighted by previous empirical studies [14,55], and also confirmed experimentally in the current study, real-life KGs are characterized by power-law distributions with respect to the number of connected components and node degrees. Our KG sampling aims to assess to what extent EA methods leave the long tail of entities of KGs under-represented in the correct matches (true positives). In summary, the contributions of this work are the following:

- We introduce the problem of structural-based indirect bias in the EA task.
- We propose SUSIE, an exploration-based sampling method to generate benchmark datasets with varying structural diversity, resembling the characteristics of real-world KGs that are typically left out of EA evaluations. Our method can be used to evaluate the trade-off between the matching accuracy and fairness of existing EA methods, covering the lack of publicly available related benchmarks as pointed out by recent surveys [15,18] .
- We show experimentally that state-of-the-art KG embedding-based EA methods exhibit structural bias against smaller, less connected regions of the KGs.

The source code used in this work is publicly accessible[5].

Outline. The rest of the paper is organized as follows. In Section 2, we introduce the basic notation used throughout the paper. In Section 3, we describe the sampling strategy followed to generate EA datasets of varying structural diversity. In Section 4, we report the experimental results that showcase the benefits of our sampling method. In Section 5, we position our study with respect to existing works, and we conclude the paper in Section 6.

[5] https://github.com/fanourakis/Sampling_for_Entity_Alignment.git

2 Preliminaries

Let $KG = (E, R, T)$ be a knowledge graph, consisting of a set of entities E (i.e., nodes), a set of relation types R (i.e., edge labels), and a set of triples T (i.e., edges). The problem of entity alignment (EA) is defined as follows: Given two knowledge Graphs $KG_1 = (E_1, R_1, T_1)$ and $KG_2 = (E_2, R_2, T_2)$, identify the set of node pairs $M \subseteq E_1 \times E_2$ that refer to the same entity. The following assumptions are common in the EA literature:

- One-to-one assumption (bijection): every node $e \in E_1$ should be mapped to exactly one node $e' \in E_2$ and vice-versa.
- Seed alignment: for training purposes, a subset $S \subseteq M$ of truly matching pairs is known in advance, commonly called the *seed alignment*.

A key notion in our sampling strategy, as well as its evaluation, is that of *weakly connected components*. Given a knowledge graph KG, a weakly connected component is a subgraph of KG where all nodes are connected to each other by some path, ignoring the direction of edges[6]. From now on, we may simply refer to weakly connected components as *connected components*, or just *components*.

Then, we adapt the component-based definition of structural diversity from [26,54], as follows:

Definition 1 (Structural diversity). *The structural diversity of a knowledge graph $KG = (E, R, T)$ is the number of connected components in KG whose size, measured by the number of vertices, is larger than or equal to an integer t, where $1 \leqslant t \leqslant |E|$.*

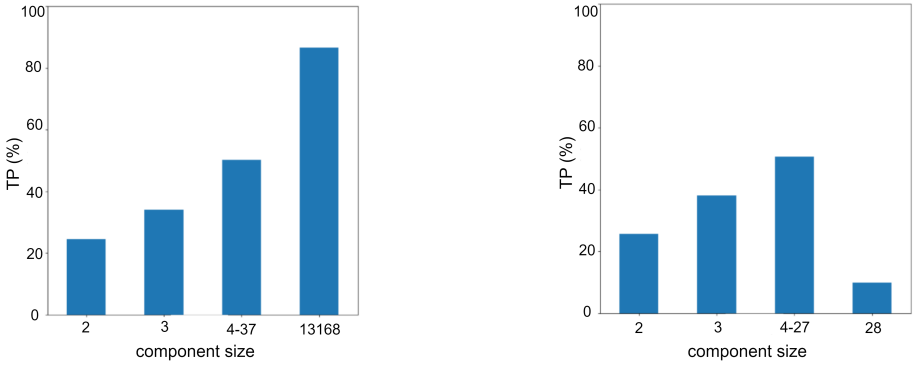
To measure structured diversity of KGs, we rely on the following graph-based metrics most of which (or unnormalized variations) have been previously investigated in [59,27] for social networks analysis.

Ratio of weakly connected components ($wccR$). The number of weakly connected components ($wcc(KG)$) indicates the connectivity of a KG. For a fixed number of nodes, the higher the number of weakly connected components, the less this graph is connected, i.e., there are many, small components in this graph. On the contrary, the fewer the connected components, in a graph of fixed size, the bigger those components are, i.e., bigger regions of the graph are connected. Intuitively, big components are easier to align than small ones, since the big components carry more relational information. To have a normalized score, we report here the ratio of the number of weakly connected components divided by the number of nodes in the knowledge graph (KG): $wccR(KG) = |wcc(KG)|/|E|$.

The number of weakly connected components has been previously used in [59,27] for measuring the structural diversity. According to [27], large number of weakly connected components, corresponds to high structural diversity.

Max component size ($maxCS$). This measure is inspired by our early findings (Figure 2 (a)), indicating that a large portion of the existing benchmark data belonged to a single connected component. Thus, the effectiveness

[6] As opposed to strongly connected components, where edge directions matter.



(a) input ($maxCS = 0.88$) (b) sample p=1 ($maxCS = 0.03$)

Fig. 2: Percentage of true matching pairs (TP) found by RREA for different sizes of connected components on KG_1 of the D-Y dataset. D-Y (a) consists of 15k nodes; its sample (b) consists of 1k nodes.

of existing methods on the biggest connected component largely determines the effectiveness of the method for the entire dataset. The effectiveness of the same method for smaller connected components is typically lower. To normalize this measure, we divide the size of the largest connected component to the number of nodes in the KG: $maxCS(KG) = \left(\max_{CC \in wcc(KG)}(|CC|)\right)/|E|$. In Figure 2 (a) the largest component has 13,168 nodes and the entire dataset has 15k nodes ($maxCS = 0.88$), while in a sample of the same dataset with 1k nodes (Figure 2 (b)), the largest component has 28 nodes ($maxCS = 0.028$).

Average node degree ($\overline{deg}$). The average node degree of a KG is defined as the ratio of the total number of incoming and outgoing edges ($deg(e)$) of each node e, divided by the number of nodes: $\overline{deg}(KG) = \frac{1}{|E|}\sum_{e_i \in E} deg\,(e_i)$. A low average node degree corresponds to high structural diversity [27].

To measure the effectiveness of each EA method, we report the values of the following standard measures.

Hits@k (H@k) measures the proportion of correctly aligned entities ranked in the top k candidates r: $Hits@k = \frac{|\{r \in \mathcal{I} | r \leqslant k\}|}{|\mathcal{I}|}$, where $\mathcal{I}$ is an individual ranked list of candidates, generated for each entity of the test set. $H@k \in [0, 1]$. This measure is easy to interpret, but it considers only top-k ranks.

Mean Reciprocal Rank (MRR) is the inverse of the harmonic mean rank: $MRR = \frac{1}{|\mathcal{I}|}\sum_{r \in \mathcal{I}} \frac{1}{r}$, where $MRR \in (0, 1]$. This metric is affected more by the top-ranked values rather than the bottom ones. Thus, MRR is less sensitive to outliers, while it considers all ranks in $\mathcal{I}$.

3 Exploration-based Sampling Algorithm

In this section, we introduce an exploration-based [7] sampling algorithm (SUSIE) for generating benchmark EA datasets out of two KGs given as input. SUSIE

[7] According to the sampling taxonomy proposed in [34].

samples parts of both input graphs by performing small random walks on each graph, before jumping/switching to the other graph. This way, it allows exploring diverse areas of both knowledge graphs, with respect to the size of the connected components that it samples. In this respect, it requires to define a desired output sample size s, measured in number of nodes in each KG, the desired minimum component size t to consider, from Definition 1, as well as a hyper-parameter p, controlling the jump probability.

SUSIE is described in Algorithm 1. In summary, it first computes the weakly connected components (Lines 3-4) and groups the nodes by the size of the components they belong to (Lines 5-6). Then, starting from KG_1 (Line 7), it performs uniform sampling on the component sizes limited by t (Line 8) and another uniform sampling to select a random node belonging to a component of this size (Line 9). While the desired sample size s has not been reached (Line 10), the algorithm adds the currently selected node, as well as one of its in- or out-neighbors (Lines 11-12) randomly, and adds them to the sampled nodes (Line 13), while also adding their aligned nodes, as given by the ground truth M (Line 14), also updating the generated seed alignment (Lines 15-16). Based on the given jump probability p, the algorithm then proceeds with a jump (Lines 19-26), also switching KGs (Line 20), or continues a random walk on the current KG (Lines 27-28). In the case of a jump, the algorithm prefers to jump to one of the nodes that may have been left disconnected (i.e., without neighbors) so far (Lines 24-26). Finally, the algorithm copies all edges between the selected sampled nodes from the original KGs, while also checking for any newly added disconnected nodes (Lines 29-34).

Complexity. The time complexity of Algorithm 1 is $O\left(2\left|E\right|\ \left(\left|E\right|+\left|R\right|\right)\right)$, which is the time complexity of generating the weakly connected components.

4 Experiments

In this section, we present our results of using our sampling algorithm to evaluate indirect fairness of state-of-the-art EA methods. A key takeaway is that we can use our sampling algorithm to measure how much existing EA methods are robust to structural diversity of the input KGs.

Experimental Setup. In our experiments, we have set the sample size s to 1,000, and the minimum component size t to 1 (Definition 1), and tested values 0, 0.15, 0.5, 0.85, and 1 for the jump probability p, with 0 corresponding to no jumps. For the experiments of EA methods, we used the code provided by RREA[8], OpenEA[9] and entity-matchers [10]. Particularly, there are two versions for MultiKE and PARIS; one that uses both relational and factual information (i.e., literal values) and another one that uses only relational information. In our work, we use the relation-based, since our sampling algorithm considers

[8] https://github.com/MaoXinn/RREA
[9] https://github.com/nju-websoft/OpenEA
[10] https://github.com/epfl-dlab/entity-matchers

Algorithm 1: SUSIE algorithm.

Input: $KG_1 = (E_1, R_1, T_1)$, $KG_2 = (E_2, R_2, T_2)$, ground truth M, jump probability p, sample size s, min component size t
Output: $KG'_1 = (E'_1, R'_1, T'_1)$, $KG'_2 = (E'_2, R'_2, T'_2)$, ground truth M'

$E'_1, E'_2, R'_1, R'_2, T'_1, T'_2, M' \leftarrow \varnothing$
$DE_1, DE_2 \leftarrow \varnothing$ // disconnected nodes
$wcc_1 \leftarrow KG_1.getWeaklyConnectedComponents()$
$wcc_2 \leftarrow KG_2.getWeaklyConnectedComponents()$
$cbs_1 \leftarrow groupByComponentSize(wcc_1)$
$cbs_2 \leftarrow groupByComponentSize(wcc_2)$
$i \leftarrow 1$ // start from KG_1
$compSize \leftarrow uniSampl(t, cbs_i.keys())$ // $t \leqslant$ random size $\leqslant |cbs_i.keys|$
$e \leftarrow uniSampl(cbs_i.nodes(compSize))$ // a random node
while $(|M'| < s)$ **do**
  $candNeighbs \leftarrow KG_i.get1HopInOutNeighbors(v)$
  $neigh \leftarrow uniSampl(candNeighbs)$
  $E'_i \leftarrow E'_i \cup e \cup neigh$
  $E'_j \leftarrow E'_j \cup matchOf(e, M) \cup matchOf(neigh, M)$ // $j = (i\%2) + 1$
  $M' \leftarrow M' \cup \{(e, matchOf(e, M))\}$ // reversed, if i=2
  $M' \leftarrow M' \cup \{(neigh, matchOf(neigh, M))\}$
  $wcc_i \leftarrow wcc_i \backslash (e \cup neigh)$
  $jump \leftarrow Binomial(p, 1-p)$ // Prob(jump) is p
  if $jump$ **then** // jump case
    $i \leftarrow (i\%2) + 1$ // switch KG
    if $DE_i = \varnothing$ **then**
      $compSize \leftarrow uniSampl(cbs_i.keys())$
      $e \leftarrow uniSampl(cbs_i.nodes(compSize))$
    else
      $e \leftarrow uniSampl(DE_i)$
      $DE_i \leftarrow DE_i \backslash e$
  else // random walk case
    $e \leftarrow neigh$
  // get all edges between sampled nodes
  for $i \in \{1, 2\}$ **do**
    foreach $(h, r, t) \in T_i$, *where* $h, t \in E'_i$ **do**
      $T'_i \leftarrow T'_i \cup \{(h, r, t)\}$
      $R'_i \leftarrow R'_i \cup \{r\}$
    $update(DE_i)$ // update the disconnected nodes
return KG'_1, KG'_2, M'

only relational information for structural diversity, while the structural diversity based on attribute information is left for future work.

Datasets. The benchmark datasets from the EA literature that we employ in our experiments, are summarized in Table 1 and briefly described next. **D-Y** [51] was constructed from DBpedia and YAGO3 KGs, describing actors, musicians, writers, films, songs, cities, football players and football teams. **D-W** [51] was constructed from DBpedia and Wikidata KGs, describing the same entity types as D-Y. **BBC-D** [20] was constructed from BBCmusic and DBpedia[11], describing various music-related entity types, such as bands, musicians, and their birth places. **MEM-E** [3] was constructed by Memory Alpha [4] and Star Trek Expanded Universe [6], describing TV series related to Star Trek.

Entity Alignment Methods. In our experiments, we employ the top-performing (according to recent experimental reviews [22,35,11,51]) embedding-

[11] https://www.csd.uoc.gr/~vefthym/minoanER/datasets.html

Table 1: Datasets Characteristics.

	Entities ($\|E_1\|$ / $\|E_2\|$)	**Relations** ($\|R_1\|$ / $\|R_2\|$)	**Triples** ($\|T_1\|$ / $\|T_2\|$)
D-Y	15,000 /15,000	165 / 28	30,291 / 26,638
D-W	15,000 /15,000	248 / 169	38,265 / 42,746
BBC-D	9,396 / 9,396	9 / 98	15,478 / 45,561
MEM-E	69,444 / 32,311	173 / 121	1,617,357 / 323,400

based relational methods (i.e., using the structural information of KGs) for EA, namely RREA [38], RDGCN [56] and MultiKE [57].

RREA [38] integrates *Graph Convolutional Networks* (GCNs) and *Graph Attention Networks* (GATs) with a *relational reflection transformation* operation in order to obtain relation-specific embeddings for KG entities. More precisely, RREA stacks multiple layers and uses weight coefficients (similar to GATs), for capturing and aggregating useful multihop-neighborhood information for the entity embeddings. The final entity embeddings come from the concatenation of the embedding of each layer and then, they are refined by minimizing the aligned (from seed alignment) entities' distance in the embedding space. RREA is a semi-supervised method, since, in each iteration, it enriches the training data with entity pairs that are mutually nearest aligned.

RDGCN [56] also utilizes GCNs and GATs. Differently to RREA, RDGCN uses dual relation graphs (i.e., graphs whose vertices are the relations of the input, primal KGs) for incorporating relational information in the entity embeddings and a GAT to encourage the interactions between the dual and the primal KGs. The generated embeddings are fed to a GCN in order to collect relation-aware entity embeddings from the primal graph and then, they are refined by learning a matrix (which is used as a linear transformation from the entities of KG_1), aiming to minimize the distance of the linearly transformed entity and its aligned entity (from seed alignment) in the embedding space. Unlike RREA, RDGCN is a supervised method, and it uses the pre-trained word embeddings of the entity names for the initialization of entity embeddings, instead of the random initialization in RREA.

MultiKE [57] learns the entity embeddings by adopting a translation-based method, TransE [57]. Particularly, given a relation triple $<h, r, t>$ in a KG, where h is a head entity, t is a tail entity, and r is a relation between the entities, it interprets the relation as a translation vector from h to t, aiming to minimize the distance $\mathbf{h}+\mathbf{r}-\mathbf{t}$ of the entity embeddings $\mathbf{h}$ and $\mathbf{t}$, plus the relation embedding $\mathbf{r}$. MultiKE, unlike RREA and RDGCN, considers only the one-hop neighbors. In parallel with learning the entity embeddings, it also aligns the relations using the relation embeddings. As for the refinement of the generated entity embeddings, it follows a variation of the method that RREA uses, enriching the training data with new triples that come from the replacement of either h or t of existing triples with their aligned entities from the seed alignment.

PARIS [50] is a probabilistic, holistic approach, i.e., it aligns both instances (entities) and schema, by estimating probabilities of equivalence (for matching), without learning KG embeddings, as previous methods do. More precisely, the

Table 2: The impact of jump probability (p) values on the sampled datasets, using graph connectivity measures.

p			input	0	0.15	0.5	0.85	1
D-Y	wccR	KG_1	0.03	0.01	0.15	0.24	0.28	0.28
		KG_2	0.04	0.05	0.18	0.24	0.28	0.29
	maxCS	KG_1	0.87	0.90	0.13	0.03	0.02	0.02
		KG_2	0.83	0.70	0.06	0.03	0.02	0.02
	$\overline{deg}$	KG_1	4.03	3.65	3.41	2.94	2.71	2.78
		KG_2	3.55	2.31	2.59	2.35	2.16	2.17
D-W	wccR	KG_1	0.01	0.01	0.14	0.24	0.28	0.29
		KG_2	0.02	0.03	0.11	0.20	0.24	0.24
	maxCS	KG_1	0.95	0.91	0.47	0.18	0.11	0.10
		KG_2	0.93	0.85	0.57	0.34	0.24	0.26
	$\overline{deg}$	KG_1	5.10	3.68	2.79	2.50	2.47	2.44
		KG_2	5.69	3.31	2.94	2.37	2.28	2.19
BBC-D	wccR	KG_1	0.18	0.16	0.23	0.29	0.34	0.36
		KG_2	0.07	0.01	0.16	0.24	0.28	0.31
	maxCS	KG_1	0.31	0.26	0.02	0.01	0.01	0.02
		KG_2	0.78	0.92	0.27	0.03	0.04	0.02
	$\overline{deg}$	KG_1	3.29	3.44	3.09	2.73	2.47	2.43
		KG_2	9.69	11.69	6.52	6.07	5.43	5.11
MEM-E	wccR	KG_1	0.00004	0.009	0.003	0.003	0.007	0.006
		KG_2	0.00009	0.001	0.003	0.005	0.011	0.008
	maxCS	KG_1	0.99	0.99	0.78	0.76	0.74	0.77
		KG_2	0.99	1	0.80	0.78	0.78	0.79
	$\overline{deg}$	KG_1	46.58	27.64	24.46	18.76	13.50	14.70
		KG_2	20.01	24.76	14.53	11.25	9.24	8.74

estimation of the probabilities, relies on quasi-functional relations, i.e., relations that for a given head entity, the expected number of tail entities is close to 1.

Finally, it is worth mentioning that the evaluation of EA methods commonly relies on the one-to-one assumption and a complete ground truth of matches, i.e., the size of seed alignment $|S|$ is the same as $|E_1|$ and $|E_2|$. Yet, recent efforts [58] have shown ways of dropping this assumption (e.g., by removing a percentage of entities from each KG). We have included a dataset (MEM-E) that does not conform to these assumptions. However, RDGCN and MultiKE could not by-pass this assumption and are thus, not evaluated on this dataset.

4.1 Experimental Results

In this section, we report our experimental findings, divided into three parts. First, we discuss how our sampling method affects the graph-related measures of the given datasets, for different values of the hyperparameter p. Then, we see how the choice of p also affects the effectiveness of the state-of-the-art evaluated EA methods. Finally, we check if there are statistically significant correlations between those graph measures and the methods' effectiveness, which in essence, reveals whether and how much an evaluated method is robust (no correlations) or not (correlations exist) to changes in the structural diversity of the input KGs.

Effects of sampling on dataset-related measures. We first report the structure-based results of the generated (sampled) datasets, when varying the hyperparameter value p (jump probability). Due to space limitations, we report in Table 2 only results for t=1. Note that samples generated by SUSIE with t=5 from D-Y for p=0.5, yield, for KG_1 and KG_2, a $wccR$ of 0.17 and 0.18, $maxCS$

Table 3: $wccR$ and $maxCS$ scores of some real-world KGs.

	LOCAH [2]	Restaurants [5]	Airlines [1]	IMDb [43]	TMDb [43]	TVDb [43]
$wccR$	0.34	0.33	0.14	0.06	0.06	0.02
$maxCS$	0.001	0.008	0.07	0.27	0.24	0.28

0.03 and 0.03, and $\overline{deg}$ 3.27 and 2.62, respectively. As expected, larger t values make the samples easier for EA methods, since larger components are sampled something that justifies the default choice of t=1 in our experiments.

Note that the denominator $|E|$ in all three graph-related measures ($wccR$, $maxCS$, $\overline{deg}$) is fixed in the generated samples, determined by the value of s. This means that, in this study, the scores of those measures are only determined by their nominators. Intuitively, we expect that more jumps (i.e., higher p) imply more connected components, of smaller size, and lower average node degree.

We first observe that in all cases, the connectivity of the input KGs to our sampling algorithm, is closer to the case of $p = 0$ (no jumps). As expected, while p is increasing, the number of weakly connected components (and $wccR$, since $|E|$ remains the same, as it is controlled by the parameter s) is also increasing, i.e., with more frequent jumps, we get more weakly connected components. We observe that $wccR$ scores are almost identical between the KGs in D-Y, D-W and MEM-E. Interestingly, this is not the case for small values of p in BBC-D, but the $wccR$ values of KG_1 and KG_2 start to converge to similar values as p increases, mostly affecting (increasing) the $wccR$ value of KG_2. Unlike other datasets, the $wccR$ of MEM-E it is much smaller, due to the limited number of weakly connected components, even in the KGs given as input to SUSIE.

Since the size of the KG remains the same, the more the weakly connected components the smaller they are. This is confirmed by the size of the largest connected component, which decreases as more jumps are performed. We further observe that the largest component sizes across the two KGs are very similar in D-Y and D-W ($maxCS$ >0.8 of the KG sizes in the input graphs and the sampled ones with p=0), while the values of $maxCS$ for the two KGs of BBC-D are very different (even by an order of magnitude in the case of p=0.15); another indicator about the heterogeneity of this dataset, explaining the lower effectiveness of EA methods. As for MEM-E, the largest component sizes across the two KGs are very similar, while $maxCS$ is much larger (even in the highest values of p), indicating that this dataset contains huge components.

Finally, the average node degree decreases as p increases, which implies that, as expected, the generated knowledge graph samples become sparser as we perform more random jumps. Again the average node degrees in D-Y and D-W are very similar between KG_1 and KG_2, which is not the case for BBC-D and MEM-E. It worth mentioning that the KGs in MEM-E are much denser than the ones of the other datasets.

As Table 3 demonstrates, there are many real-world KGs that exhibit similar $maxCS$ and/or $wccR$ to our sampled KGs. For instance, LOCAH exhibits $wccR$=0.34 and $maxCS$=0.001, while the sample generated by SUSIE for D-Y and p=0.5 exhibits very similar characteristics ($wccR$=0.24 and $maxCS$=0.03).

Table 4: The impact of sampling on the effectiveness of RREA, as the jump probability p increases.

p		input	0	0.15	0.5	0.85	1
D-Y	H@1	.807	.804	.500	.454	.367	.384
	H@10	.928	.931	.792	.717	.652	.682
	MRR	.855	.844	.605	.541	.467	.486
D-W	H@1	.697	.730	.465	.372	.354	.421
	H@10	.898	.918	.725	.640	.604	.672
	MRR	.772	.802	.554	.454	.435	.499
BBC-D	H@1	.389	.466	.404	.347	.315	.271
	H@10	.611	.707	.570	.477	.401	.392
	MRR	.472	.556	.473	.399	.350	.317
MEM-E	H@1	.249	.154	.134	.079	.064	.131
	H@10	.616	.591	.463	.333	.320	.416
	MRR	.367	.277	.237	.175	.152	.223

Table 5: The impact of sampling on the effectiveness of RDGCN.

p		input	0	0.15	0.5	0.85	1
D-Y	H@1	.924	.928	.908	.847	.908	.865
	H@10	.967	.973	.974	.947	.967	.948
	MRR	.940	.946	.934	.887	.934	.900
D-W	H@1	.526	.631	.500	.450	.438	.437
	H@10	.730	.820	.727	.642	.638	.640
	MRR	.591	.699	.586	.527	.518	.514
BBC-D	H@1	.067	.071	.080	.084	.102	.102
	H@10	.114	.146	.138	.140	.164	.154
	MRR	.080	.101	.106	.108	.126	.127

Table 6: The impact of sampling on the effectiveness of MultiKE.

p		input	0	0.15	0.5	0.85	1
D-Y	H@1	.554	.431	.264	.261	.247	.218
	H@10	.802	.763	.602	.570	.510	.511
	MRR	.636	.544	.382	.370	.340	.316
D-W	H@1	.286	.367	.235	.200	.225	.214
	H@10	.579	.727	.548	.400	.428	.418
	MRR	.377	.484	.347	.274	.296	.286
BBC-D	H@1	.247	.292	.270	.252	.255	.208
	H@10	.531	.674	.540	.452	.408	.387
	MRR	.342	.426	.377	.332	.314	.280

Effects of sampling on the effectiveness of EA methods. Next, we report the effectiveness results of the employed EA methods on the generated (sampled) datasets, when varying the hyperparameter value p (jump probability). Those results are summarized in Tables 4 (for RREA), 5 (for RDGCN), 6 (for MultiKE) and 7 (for PARIS).

Again, we observe that in all cases, the behavior of the EA algorithms in the input KGs (i.e., before sampling) is closer to the case of $p = 0$ (no jumps). In Table 4, for all datasets, the effectiveness of RREA are dropping while p is incrementally increasing from 0 to 0.85. The biggest effect is when comparing the effectiveness on samples of p=0 to those of p=0.15. In both BBC-D and MEM-E, unlike D-Y and D-W, the impact of changing p from 0.15 to 0.5 is also large. In all cases, the impact of changing p from 0.5 to 0.85 is much smaller, while changing p from 0.85 to 1 seems to have a negligible effect on the effectiveness of RREA. Overall, the effects of increasing p on the effectiveness of RREA are larger (e.g., there is a 44% drop in H@1 for D-Y) for the datasets (D-Y and D-W) in which RREA was having good results on the original input graphs, and smaller ($\leqslant$ 18% and 22% for any measure) for BBC-D and MEM-E, in which RREA was struggling. As mentioned in Section 4.1, we further include some indicative experiments, setting t=5 with the corresponding H@1, H@10 and MRR for p=0.5 for RREA: 0.460, 0.808, 0.592, showcasing that higher values of t make the EA datasets less challenging (since small components are excluded).

In Table 5, we see that the effectiveness RDGCN is not affected so much by changing the jump probability values, compared to RREA. There is no change

Table 7: The impact of sampling on the effectiveness (H@1) of PARIS.

p	input	0	0.15	0.5	0.85	1
D-Y	.979	.454	.267	.265	.271	.221
D-W	.841	.460	.242	.184	.213	.209
BBC-D	.387	.325	.302	.267	.288	.242
MEM-E	.082	.060	.047	.019	.011	.031

larger than 9.2% (H@10 in D-W) for any measure. This is probably due to the impact of initializing the embeddings with entity names in RDGCN, as opposed to random initialization in RREA. Having an additional source of alignment information helps RDGCN to get better results when the relational information become poorer (i.e., when p increases), as compared to RREA that relies entirely on relational information. RDGCN, was unable to run on MEM-E, due to one-to-one assumption, so we excluded this dataset from this experiment.

In Table 6, for all datasets, we see that the effectiveness of MultiKE, similarly to RREA, is dropping while p is incrementally increasing from 0 to 1. The largest impact in the effectiveness of the method due to our sampling algorithm, is observed when we increase p from 0 to 0.15 for D-Y and D-W. When changing p from 0.15 to 0.50, from 0.5 to 0.85 and from 0.85 to 1, the effect on the scores is much smaller compared to RREA and in some cases negligible (e.g., H@1 on D-Y when changing p from 0.15 to 0.5). This is probably due to the already bad results on the original input graphs, since unlike RREA and RDGCN, MultiKE considers only the one-hop neighbors and unlike RDGCN, it does not consider entity names as matching evidence. MultiKE could not run on MEM-E, either. Finally, we further extend our experiments by investigating the robustness of EA systems, like the version of MultiKE that accounts also textual facts, to structural diversity. The results showcase that it remains unaffected (Hits@1 scores in D-Y dataset, which are 0.90, 0.85, 0.92, 0.92, for input, p = 0, 0.15, 1, respectively) to structural diversity.

In Table 7, we consider F1-score (that PARIS reports) equal to H@1, since in the test phase, each source entity gets a list of candidates [51]. We observe that the effectiveness of PARIS is dropping while p is increasing. The largest impact on the effectiveness of PARIS, due to our sampling algorithm, is observed when p goes from 0 to 0.15 for D-Y and D-W. When changing p from 0.15 to 0.50, from 0.50 to 0.85 and from 0.85 to 1, the effect on the effectiveness of PARIS is negligible. This happens because PARIS uses only the functional relations, in contrast to the other methods that use all the relations for the embeddings.

Correlations between graph measures and EA effectiveness. Table 8 reports the statistically significant Spearman's correlations between data measures ($wccR$, $maxCS$, $\overline{deg}$) and effectiveness measures (H@k, MRR), while cells with a dash ('-') are those without statistically significant correlations. In this correlation analysis, we used only D-Y, D-W and BBC-D, since RDGCN and MultiKE were unable to run with MEM-E.

This table shows that in RREA, MultiKE and PARIS, $wccR$ is negatively correlated with effectiveness (i.e., bigger $wccR$ comes with worse results) of the methods, while $maxCS$ and $\overline{deg}$ are positively correlated (i.e., higher average node degree and bigger size of the largest component, are both associated with

Table 8: Spearman's correlations between the connectivity of sampled datasets and effectiveness results $Hits@k$ ($H@k$) and MRR. Dashed cells denote statistically not significant correlation, with p-values > 0.05.

		wccR		maxCS		$\overline{deg}$	
		KG_1	KG_2	KG_1	KG_2	KG_1	KG_2
RREA	H@1	-0.90	-0.81	0.81	0.71	0.86	-
	H@10	-0.85	-0.65	0.79	0.56	0.77	-
	MRR	-0.88	-0.75	0.79	0.65	0.84	-
RDGCN	H@1	-	-	-	-	-	-0.66
	H@10	-	-	-	-	-	-0.65
	MRR	-	-	-	-	-	-0.65
MultiKE	H@1	-0.70	-0.70	0.46	0.55	0.84	-
	H@10	-0.87	-0.77	0.68	0.61	0.93	-
	MRR	-0.80	-0.72	0.56	0.54	0.90	-
PARIS	H@1	-0.65	-0.68	0.49	0.60	0.81	0.51

better results). In RDGCN, $\overline{deg}$ in KG_2 is negatively correlated with the method effectiveness (i.e., smaller average node degree comes with better results).

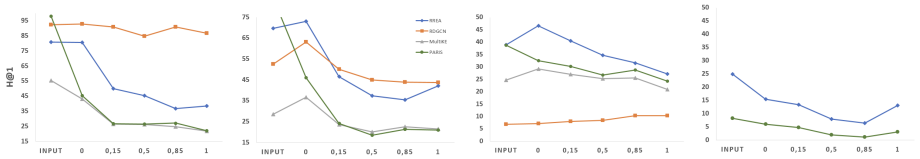
Discussion. We observe that probabilistic, non-embedding-based methods like PARIS, assuming that the input graphs are isomorphic and few relations are more important than others (i.e., quasi-functional), outperform embedding-based methods, but they are not robust even to the slightest structural variations (i.e., even with small values of p).

On the other hand, the robustness of embedding-based methods against the structural diversity of input graphs depends on the factual information (e.g., attributes values, entity names) that they exploit to align entities. For example, RDGCN is the most robust method to structural diversity, RREA is the least robust, while MultiKE lies somewhere in the middle. Unlike RREA, which only relies on the relational structure, RDGCN heavily relies on the naming of entities, which is not affected by our sampling. The small difference on the impact of structural diversity between RREA and MultiKE is due to the fact that RREA considers multi-hop neighbors, while MultiKE stops at hop-1 neighbors.

Finally, we observe that our sampling significantly impacts the evaluation of EA methods, as it changes the entity ranking of the top-performing methods, even in the same input KGs. This behavior is illustrated in Figure 3, where the performance of EA methods is measured using Hits@1 (similar observations hold for the other measures). For example, we observe that in D-Y, PARIS is the top-performing method for the input KGs, but it is the worst-performing method for our samples, for all values of p. Moreover, for D-W, PARIS has the same behavior as in D-Y, but also RREA and RDGCN outperform each other for different values of p; RDGCN starts outperforming RREA for $p \geqslant 0.15$. Overall, SUSIE reveals the performance differences of EA methods that enable to assess their robustness against increasing structural diversity of input graphs.

5 Related Work

In this section, we position our contributions w.r.t. previous works related to bias in KG-embedding-based prediction tasks (summarized in Table 9), as well as, to existing KG sampling algorithms.



(a) D-Y (b) D-W (c) BBC-D (d) MEM-E

Fig. 3: Robustness of EA methods' effectiveness (H@1) to structural diversity brought by SUSIE sampling.

Bias in KG embeddings. The majority of works on KG embeddings focus on direct forms of bias[12] (group fairness). Among them, only few are related to EA tasks, aiming to mitigate bias by satisfying fairness constraints [21], or to identify name matching bias using string similarity measures [31]. Most of the bias related work focuses on link prediction [32,23,24,49], node classification [33] and recommendation [36,8,45,10] tasks.

Fairwalk [45] is an embedding method based on modified random walks, that weights the edges between the node and their one-hop neighbors, in order the latter to be chosen equiprobably and independently to the sensitive group they belong to. Crosswalk [33] is also a random walk-based embedding method, but, unlike Fairwalk, extends the range of the weighting including multi-hop neighbors. Few works recently consider indirect forms of bias, such as graph modularity/homophily in node classification [19], link prediction [40] or complex networks [48]. We should also mention [17,29] that investigate individual fairness on GNNs from a ranking perspective. The main difference to these tasks is that EA involves two, instead of one, KGs.

To the best of our knowledge, no previous work addressed indirect forms of bias related to the structural diversity w.r.t. connected components of KGs in EA. Only the impact of node degree in the message passing protocol of GNNs has been investigated so far in node classification [52,12,30], link prediction or recommendation [28,47] tasks. Additionally, Div2vec [28] proposes a random walk-based method for generating diverse node embeddings, with the probability of a node to be selected being inversely proportional to its degree. Finally, some works investigate connectivity-related fairness in social networks [59,27]. Two of our evaluation measures ($wccR$ and $\overline{deg}$) are adaptations of the measures proposed in [59,27], while other measures like k-core and k-brace decomposition are not relevant to our sampling method that was not designed to affect them. Our work essentially covers the lack of publicly available benchmark data for assessing fairness of EA tasks, as pointed out by recent surveys [15,18].

Graph sampling algorithms. In order to reduce the size of the initial graphs while preserving their structural properties, different sampling methods have been proposed that fall under three categories [34]: random node, edge selection, and sampling by exploration.

Random Node selection by uniform sampling does not retain the power-law node degree distribution, while non-uniform sampling (Random PageRank and

[12] [18,44] investigate direct forms of bias on "flat", tabular data.

Table 9: Categorization of works related to bias in Node Classification (NC), Recommendation (REC), Link Prediction (LP) and Entity Alignment (EA) tasks.

		NC	REC	LP	EA
Direct	Group	[33]	[36,8,45,10]	[32,23,24,36,8,45,33,10,49]	[21,31]
Indirect	Individual	[17,19]	-	[17,29,48,40]	-
	Degree-related	[52,12,30]	[28]	[28,47]	[11]
	Connectivity-related	-	[27]	[59,27]	-

Random Degree Node), produces very dense KG samples, with too many high-degree nodes. Random Edge sampling aims to retain as much as possible the degree distribution, but by keeping only a subset of edges, we end up with sampled KGs with possibly different neighborhoods compared to the initial graphs (some edges in the input graphs may not exist in the sampled graphs).

On the other hand sampling by exploration, the category to which SUSIE belongs, selects a node uniformly at random and explores the nodes in its vicinity using random walks [34]. This allows us to control the probability with which the output sample will include entities of diverse connected component sizes. Motivated by the aforementioned works, Iterative Degree-based Sampling [51] preserves the degree distribution of the initial graphs, by removing nodes with probability proportional to the nodes PageRank scores. In addition, the sampling of [11] aims to reduce name bias of KGs while preserving the structural properties (e.g., degree distribution) of the input KGs. Unlike graph sampling methods used in embeddings (e.g., Fairwalk, Crosswalk, div2vec), or in node classification and link prediction tasks, our sampling method is the first exploration-based sampling that allows to control (directly) the number and (indirectly) the size of connected components of the two input KGs for the task of EA.

6 Conclusions

In this work, we have shown that the structural diversity of EA benchmark data, which has not been evaluated before, is a factor that affects the performance of state-of-the-art EA methods. To do that, we have introduced an exploration-based sampling algorithm (SUSIE) that detects challenging subgraphs of a given EA benchmark dataset, with respect to structural diversity. We have further shown that methods like RDGCN, that do not rely exclusively on relational data, but also consider other sources of alignment information (e.g., entity names) are more robust to such diversity, than other EA methods like RREA, relying exclusively on the graph structure of the input KGs.

Assessing diversity in EA is only the first step in our ongoing work for a method that exploits spectral GNNs for structural matching, as well as attributes and entity names. Thus, we plan to extend our sampling method to consider not only structural diversity, but also diversity in factual information (i.e., literal values), as well as to examine additional diversity measures and to determine the number and size distribution of the sampled connected components.

Acknowledgement. The work of N. Fanourakis and V. Efthymiou was funded from the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under GA No 969.

References

1. Airlines dataset. https://archive.org/download/kasabi, accessed: 2023-03-02
2. Locah dataset. http://data.archiveshub.ac.uk/, accessed: 2023-03-02
3. Mem-e dataset from oaei 2022. http://oaei.ontologymatching.org/2022/knowledgegraph/index.html, accessed: 2023-03-02
4. Memory-alpha dataset. http://memory-alpha.wikia.com/, accessed: 2023-03-02
5. Restaurants dataset from oaei 2010. http://oaei.ontologymatching.org/2010/im/, accessed: 2023-03-02
6. Star trek expanded universe dataset. http://stexpanded.wikia.com/, accessed: 2023-03-02
7. Biemer, P.P., de Leeuw, E.D., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L.E., Tucker, N.C., West, B.T.: Total survey error in practice. John Wiley & Sons (2017)
8. Bose, A.J., Hamilton, W.L.: Compositional fairness constraints for graph embeddings. In: ICML. vol. 97, pp. 715–724 (2019)
9. Bourli, S., Pitoura, E.: Bias in knowledge graph embeddings. In: ASONAM. pp. 6–10 (2020)
10. Buyl, M., Bie, T.D.: Debayes: a bayesian method for debiasing network embeddings. CoRR **abs/2002.11442** (2020)
11. Chaurasiya, D., Surisetty, A., Kumar, N., Singh, A., Dey, V., Malhotra, A., Dhama, G., Arora, A.: Entity alignment for knowledge graphs: Progress, challenges, and empirical studies. CoRR **abs/2205.08777** (2022)
12. Chen, M., Wei, Z., Huang, Z., Ding, B., Li, Y.: Simple and deep graph convolutional networks. CoRR **abs/2007.02133** (2020)
13. Chen, M., Tian, Y., Chang, K., Skiena, S., Zaniolo, C.: Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment. In: IJCAI. pp. 3998–4004 (2018)
14. Cheng, W., Wang, C., Xiao, B., Qian, W., Zhou, A.: On statistical characteristics of real-life knowledge graphs. In: BPOE. vol. 9495, pp. 37–49 (2015)
15. Choudhary, M., Laclau, C., Largeron, C.: A survey on fairness for machine learning on graphs. CoRR **abs/2205.05396** (2022)
16. Christophides, V., Efthymiou, V., Stefanidis, K.: Entity Resolution in the Web of Data. Morgan & Claypool Publishers (2015)
17. Dong, Y., Kang, J., Tong, H., Li, J.: Individual fairness for graph neural networks: A ranking based approach. In: KDD. pp. 300–310 (2021)
18. Dong, Y., Ma, J., Chen, C., Li, J.: Fairness in graph mining: A survey. CoRR **abs/2204.09888** (2022)
19. Dwork, C., Hardt, M., Pitassi, T., Reingold, O., Zemel, R.S.: Fairness through awareness. In: Innovations in Theoretical Computer Science. pp. 214–226 (2012)
20. Efthymiou, V., Stefanidis, K., Christophides, V.: Big data entity resolution: From highly to somehow similar entity descriptions in the web. In: IEEE BigData. pp. 401–410 (2015)
21. Efthymiou, V., Stefanidis, K., Pitoura, E., Christophides, V.: Fairer: Entity resolution with fairness constraints. In: CIKM. pp. 3004–3008 (2021)

22. Fanourakis, N., Efthymiou, V., Kotzinos, D., Christophides, V.: Knowledge graph embedding methods for entity alignment: An experimental review. CoRR **abs/2203.09280** (2022)
23. Fisher, J.: Measuring social bias in knowledge graph embeddings. CoRR **abs/1912.02761** (2019)
24. Fisher, J., Mittal, A., Palfrey, D., Christodoulopoulos, C.: Debiasing knowledge graph embeddings. In: EMNLP. pp. 7332–7345 (2020)
25. Guo, Q., Zhuang, F., Qin, C., Zhu, H., Xie, X., Xiong, H., He, Q.: A survey on knowledge graph-based recommender systems. IEEE Trans. Knowl. Data Eng. **34**(8), 3549–3568 (2022)
26. Huang, X., Cheng, H., Li, R., Qin, L., Yu, J.X.: Top-k structural diversity search in large networks. VLDB J. pp. 319–343 (2015)
27. Huang, X.L., Tiwari, M., Shah, S.: Structural diversity in social recommender systems. In: RecSys (2013)
28. Jeong, J., Yun, J., Keam, H., Park, Y., Park, Z., Cho, J.: div2vec: Diversity-emphasized node embedding. In: RecSys (2020)
29. Kang, J., He, J., Maciejewski, R., Tong, H.: Inform: Individual fairness on graph mining. In: KDD. pp. 379–389 (2020)
30. Kang, J., Zhu, Y., Xia, Y., Luo, J., Tong, H.: Rawlsgcn: Towards rawlsian difference principle on graph convolutional network. CoRR **abs/2202.13547** (2022)
31. Karakasidis, A., Pitoura, E.: Identifying bias in name matching tasks. In: EDBT. pp. 626–629 (2019)
32. Keidar, D., Zhong, M., Zhang, C., Shrestha, Y.R., Paudel, B.: Towards automatic bias detection in knowledge graphs. In: EMNLP. pp. 3804–3811 (2021)
33. Khajehnejad, A., Khajehnejad, M., Babaei, M., Gummadi, K.P., Weller, A., Mirzasoleiman, B.: Crosswalk: Fairness-enhanced node representation learning. In: AAAI. pp. 11963–11970 (2022)
34. Leskovec, J., Faloutsos, C.: Sampling from large graphs. In: SIGKDD. pp. 631–636 (2006)
35. Li, J., Song, D.: Uncertainty-aware pseudo label refinery for entity alignment. In: WWW. pp. 829–837 (2022)
36. Li, P., Wang, Y., Zhao, H., Hong, P., Liu, H.: On dyadic fairness: Exploring and mitigating bias in graph connections. In: ICLR (2021)
37. Luo, Y., Yang, B., Xu, D., Tian, L.: A survey: Complex knowledge base question answering. In: ICICSE. pp. 46–52 (2022)
38. Mao, X., Wang, W., Xu, H., Wu, Y., Lan, M.: Relational reflection entity alignment. In: CIKM. pp. 1095–1104 (2020)
39. Martínez-Plumed, F., Ferri, C., Nieves, D., Hernández-Orallo, J.: Missing the missing values: The ugly duckling of fairness in machine learning. Int. J. Intell. Syst. pp. 3217–3258 (2021)
40. Masrour, F., Wilson, T., Yan, H., Tan, P., Esfahanian, A.: Bursting the filter bubble: Fairness-aware network link prediction. In: AAAI. pp. 841–848 (2020)
41. Molokwu, B.C., Shuvo, S.B., Kar, N.C., Kobti, Z.: Node classification in complex social graphs via knowledge-graph embeddings and convolutional neural network. In: ICCS. vol. 12142, pp. 183–198 (2020)
42. Nathani, D., Chauhan, J., Sharma, C., Kaul, M.: Learning attention-based embeddings for relation prediction in knowledge graphs. In: ACL. pp. 4710–4723 (2019)
43. Obraczka, D., Schuchart, J., Rahm, E.: EAGER: embedding-assisted entity resolution for knowledge graphs. CoRR **abs/2101.06126** (2021)
44. Quy, T.L., Roy, A., Iosifidis, V., Zhang, W., Ntoutsi, E.: A survey on datasets for fairness-aware machine learning. WIREs Data Mining Knowl. Discov. **12**(3) (2022)

45. Rahman, T.A., Surma, B., Backes, M., Zhang, Y.: Fairwalk: Towards fair graph embedding. In: IJCAI. pp. 3289–3295 (2019)
46. Rossi, A., Barbosa, D., Firmani, D., Matinata, A., Merialdo, P.: Knowledge graph embedding for link prediction: A comparative analysis. ACM Trans. Knowl. Discov. Data **15**(2), 14:1–14:49 (2021)
47. Sanz-Cruzado, J., Pepa, S.M., Castells, P.: Structural novelty and diversity in link prediction. In: WWW. pp. 1347–1351 (2018)
48. Saxena, A., Fletcher, G., Pechenizkiy, M.: Hm-eiict: Fairness-aware link prediction in complex networks using community information. Journal of Combinatorial Optimization pp. 1–18 (2021)
49. Sinha, A., Cazabet, R., Vaudaine, R.: Systematic biases in link prediction: Comparing heuristic and graph embedding based methods. In: COMPLEX NETWORKS. Studies in Computational Intelligence, vol. 812, pp. 81–93 (2018)
50. Suchanek, F.M., Abiteboul, S., Senellart, P.: PARIS: probabilistic alignment of relations, instances, and schema. PVLDB **5**(3), 157–168 (2011)
51. Sun, Z., Zhang, Q., Hu, W., Wang, C., Chen, M., Akrami, F., Li, C.: A benchmarking study of embedding-based entity alignment for knowledge graphs. Proc. VLDB Endow. **13**(11), 2326–2340 (2020)
52. Tang, X., Yao, H., Sun, Y., Wang, Y., Tang, J., Aggarwal, C.C., Mitra, P., Wang, S.: Investigating and mitigating degree-related biases in graph convoltuional networks. In: CIKM. pp. 1435–1444 (2020)
53. Trisedya, B.D., Qi, J., Zhang, R.: Entity alignment between knowledge graphs using attribute embeddings. In: AAAI. pp. 297–304 (2019)
54. Ugander, J., Backstrom, L., Marlow, C., Kleinberg, J.M.: Structural diversity in social contagion. Proc. Natl. Acad. Sci. USA **109**(16), 5962–5966 (2012)
55. Wang, S.: On the analysis of large integrated knowledge graphs for economics, banking and finance. In: EDBT/ICDT Workshops. vol. 3135 (2022)
56. Wu, Y., Liu, X., Feng, Y., Wang, Z., Yan, R., Zhao, D.: Relation-aware entity alignment for heterogeneous knowledge graphs. In: IJCAI. pp. 5278–5284 (2019)
57. Zhang, Q., Sun, Z., Hu, W., Chen, M., Guo, L., Qu, Y.: Multi-view knowledge graph embedding for entity alignment. In: IJCAI 2019 (2019)
58. Zhang, R., Trisedya, B.D., Li, M., Jiang, Y., Qi, J.: A benchmark and comprehensive survey on knowledge graph entity alignment via representation learning. VLDB J. **31**(5), 1143–1168 (2022)
59. Zhang, Y., Wang, L., Zhu, J.J.H., Wang, X., Pentland, A.S.: The strength of structural diversity in online social networks. CoRR **abs/1906.00756** (2019)