

Matched Excess-Outranker Regularization for Candidate-Set Interference in Continual Knowledge Graph Embedding

Hao Ren
University of New South Wales
Sydney, NSW, Australia
hao.ren@unsw.edu.au

Junbin Gao
The University of Sydney
Sydney, NSW, Australia
junbin.gao@sydney.edu.au

Jiaojiao Jiang[✉]
University of New South Wales
Sydney, NSW, Australia
jiaojiao.jiang@unsw.edu.au

Abstract

Continual knowledge graph embedding updates entity and relation representations as a graph grows. Existing methods primarily address catastrophic forgetting, but entity admission also changes the candidate universe of every compatible query. A historical answer can therefore lose rank even when its score and its ordering among old entities are preserved. We formalize this effect as candidate-set interference and introduce Matched Excess-Outranker Regularization (MEOR), a host-level objective that compares smooth answer-relative newcomer pressure with score-blind, structurally matched old references. Its one-sided penalty acts only when newcomer competition exceeds the matched reference, preserving the host learner's signal for legitimate new entities. Across eight paired runs on ENTITY-ComplEx, MEOR improves historical current-universe mean reciprocal rank (MRR) by 0.0057 over replay and reduces candidate-set interference by 0.0055, with one-sided 95% lower bounds of 0.0052 and 0.0051, respectively. It satisfies the preservation criteria for old-universe ranking and newcomer acquisition and improves historical current-universe MRR over persistent calibration, matched maximum regularizer (MMR), and unmatched old regularizer (UOR). Direct ablations support each component of its reference construction and aggregation. Adding MEOR also improves historical ranking in all ten reported FBInc-S and FBInc-L host and backbone settings, with every paired 95% confidence interval excluding zero. These results establish candidate admission as a distinct source of continual rank loss and show that it can be controlled without replacing the underlying embedding architecture or continual learner.

CCS Concepts

• **Computing methodologies** → **Knowledge representation and reasoning**; • **Information systems** → *Information integration*; *Novelty in information retrieval*; Graph-based database models.

Keywords

Continual Knowledge Graph Embedding, Evolving Knowledge Graphs, Candidate-Set Interference, Ranking Regularization

ACM Reference Format:

Hao Ren, Junbin Gao, and Jiaojiao Jiang. XXXX. Matched Excess-Outranker Regularization for Candidate-Set Interference in Continual Knowledge Graph Embedding. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 11 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 Introduction

Knowledge graphs (KGs) in deployed systems expand as new entities and facts become available, while link prediction must remain reliable for knowledge already represented [2, 21]. Continual knowledge graph embedding (KGE) supports this process by updating a model without retraining it from the beginning. Existing continual KGE research has focused primarily on catastrophic forgetting, in which learning from new facts changes representations acquired earlier. This account is incomplete when graph growth expands the entity set. Every admitted entity joins the candidate universe of each compatible query and can displace an established answer even when the answer score and its ordering relative to every previously known entity remain unchanged [17]. The research problem addressed in this paper is how to control this admission-induced historical rank loss without materially weakening the model's ability to learn newly admitted entities when they are correct answers. We address this problem with Matched Excess-Outranker Regularization (MEOR), a host-level objective that suppresses only the newcomer pressure exceeding a score-blind, structurally matched old-candidate reference.

Figure 1 illustrates the cumulative entity admission that creates this additional competition. We study the resulting failure mode as candidate-set interference. Unlike catastrophic forgetting, it does not require a change in model parameters or in the relative ordering of old entities. It can be isolated at a fixed checkpoint by evaluating the same historical query over two candidate universes. The old universe measures ranking among previously known entities, whereas the current universe additionally contains the entities admitted through graph growth. Their difference identifies the rank loss attributable to newcomer competition. Current-universe evaluation exposes this effect, but it does not provide a training rule for controlling it. Replay, distillation, and parameter regularization protect previously learned representations, yet they do not directly regulate the query-specific competition created when new entities enter the candidate universe [3, 4, 9, 13]. Recent work identifies entity interference as an evaluation concern [17]. Our central contribution is to convert this diagnostic distinction into an explicit objective for continual learning.

MEOR operates on historical replay queries during continual refinement [19]. For each query, it measures the positive score

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Conference acronym 'XX, Woodstock, NY

© XXXX Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-XXXX-X/XXXX/XX
<https://doi.org/XXXXXXX.XXXXXXX>

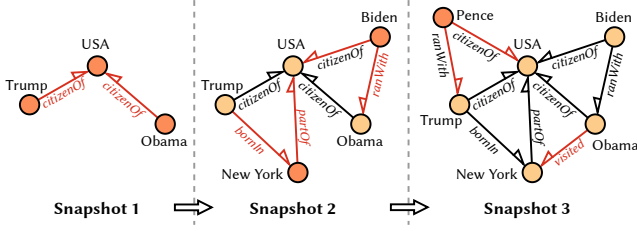


Figure 1: Cumulative evolution of a continual knowledge graph across three snapshots.

gaps between newly admitted candidates and the recorded answer, normalizes these gaps to the local score scale, and smoothly aggregates the above-answer tail. It then compares this newcomer pressure with pressure from old candidates matched by prediction role, degree, and incident relation-role structure. The matching procedure does not use model scores, so the reference cannot adapt to the outcome against which it is evaluated. A one-sided squared hinge penalizes only the pressure exceeding the matched reference. Newcomers whose aggregate pressure remains consistent with structurally comparable old candidates receive no penalty, while the original host objective continues to govern the learning of new facts. MEOR therefore targets the harmful component of admission-induced competition without replacing the embedding architecture or continual update rule of the host.

The experiments support both the ranking benefit and the proposed mechanism. Across eight paired runs on ENTITY-CompLex, MEOR improves historical current-universe mean reciprocal rank (MRR) by 0.0057 over replay, with a one-sided 95% lower bound of 0.0052, and reduces candidate-set interference by 0.0055, with a lower bound of 0.0051. Relative to replay, the mean changes are 0.0002 in old-universe MRR and 0.0015 in target-newcomer MRR, and both lower bounds satisfy the preservation rule. MEOR also improves historical ranking over persistent calibration, Matched Maximum Regularizer (MMR), and Unmatched Old Regularizer (UOR). Direct ablations show positive contributions from smooth tail aggregation, structural matching, query-specific assignment, and old-reference centering. On FBInc-S and FBInc-L, adding MEOR improves historical current-universe MRR in all ten reported combinations of stream, backbone, and continual host. Every paired 95% confidence interval excludes zero, and every newcomer-acquisition bound remains within the preservation margin.

This work makes three contributions. First, we formalize candidate-set interference as a continual KGE optimization problem distinct from catastrophic forgetting and provide a same-checkpoint rank decomposition that separates admission-induced competition from changes in the ordering of old entities. Second, we introduce MEOR, a differentiable intervention that combines answer-relative smooth tail aggregation, score-blind structural matching, and one-sided excess regularization while preserving the host learning objective. Third, we provide a paired, mechanism-focused empirical analysis that connects historical-ranking gains to reduced candidate-set interference, isolates the contribution of each design component, and demonstrates transfer across the reported graph streams, embedding backbones, and continual-learning hosts.

2 Related Work

2.1 Static KGE and Ranked Evaluation

Knowledge graph embedding represents entities and relations in a continuous space and assigns a plausibility score to each candidate triple. Translation-based models such as TransE interpret a relation as a displacement between entity embeddings [1]. Bilinear models score triples through multiplicative interactions, and ComplEx extends this family to complex-valued embeddings, allowing a single scoring form to represent both symmetric and antisymmetric relations [25]. More recent probabilistic approaches, such as Normalizing Flow Embedding, replace deterministic point embeddings with distributional representations and score triples through similarities between learned flows, providing a richer treatment of representation uncertainty [26]. These models differ in geometry and scoring function, but they share the same static link-prediction interface: given a query formed by masking one endpoint of a triple, the model scores candidate entities and ranks them by plausibility.

Although KGE models differ in representation geometry, they are commonly assessed through link prediction, where learned triple scores are converted into ranked candidate lists. A link-prediction query masks the head or tail entity of a known fact, scores the eligible replacement entities, and ranks its recorded answer among them. Filtered evaluation removes other known correct answers from the candidate list so that valid alternatives are not counted as errors [20]. The resulting rank therefore depends not only on the learned score function, but also on the filtering rule and the candidate universe. In static benchmarks, this universe is usually fixed. In a growing knowledge graph, however, newly admitted entities become additional competitors for historical queries. Candidate construction therefore becomes part of the continual learning problem, rather than a neutral evaluation detail.

2.2 Temporal and Continual KGE

Temporal KGE and continual KGE address different forms of change. Temporal KGE treats time as part of the fact being modeled: facts are associated with timestamps or validity intervals, and these temporal annotations directly condition the learned representations or scoring function [28]. HyTE represents each timestamp as a hyperplane on which entities and relations are projected [5], while diachronic embedding makes entity representations explicit functions of time [7]. In these models, time is semantic information used at prediction time.

Continual KGE addresses a different setting: the model is updated as facts, entities, or relations are admitted across successive graph states. The update order determines what information is available at each stage, but time need not be a semantic argument of the score function. A graph state may reflect data acquisition, benchmark construction, or incremental deployment rather than a factual validity interval. Thus, temporal KGE asks which fact holds at a specified time, whereas continual KGE asks how an existing representation can be preserved and extended as the graph [18].

Continual KGE treats graph evolution as an incremental representation learning problem and adapts continual-learning strategies to update knowledge graph embeddings without relearning all concepts from scratch [4]. Representative methods use selected

historical data and disentangled components [4, 9], masked reconstruction, embedding transfer, and distillation [3, 13], or efficient adaptation, dynamic regularization, coordinated masks, and task-conditioned transfer [14, 23, 29, 30]. Despite their architectural differences, these methods share the goal of retaining useful knowledge while incorporating each graph update.

2.3 Catastrophic Forgetting in Continual KGE

The central concern in continual KGE is catastrophic forgetting [15]. Training on newly available facts can move established representations and reduce performance on historical knowledge, creating a tension between stability and plasticity. The model must preserve useful structure from earlier graph states while remaining responsive to new entities, relations, and facts.

Existing methods address this problem through complementary strategies. Replay retains selected historical observations, regularization limits destructive parameter change, and distillation transfers information from an earlier model state [19]. Selective and modular updates restrict which representations are modified, while reconstruction objectives preserve information encoded in the earlier graph [3, 4, 9, 13]. FMR combines rotated representations with dynamic regularization [29], CMKGE coordinates stability and plasticity through dual masks [23], and Dynamic Temperature Distillation adjusts knowledge transfer according to fact volatility [22]. Other work improves continual updates through informed initialization [16], adaptive representation capacity [12], structure-aware representation learning [10], and Bayesian-guided evolution [11].

These methods primarily control how learned representations change between graph states. Their evaluation therefore emphasizes whether knowledge encoded before an update remains recoverable afterward. This perspective captures parameter and representation forgetting, but it does not fully describe the ranking consequences of enlarging the entity universe.

2.4 Candidate-Set Interference

Candidate-set interference is a distinct source of historical rank loss. When new entities enter the graph, they become eligible answers to earlier queries. A historical answer can therefore move down in the ranking even when its score and its ordering relative to every old entity remain unchanged. The loss arises because the query faces additional competitors, not because the model has forgotten how to rank established entities.

Pons et al. [17] identify this effect as entity interference and show that evaluation restricted to old candidates overstates continual performance. A same-checkpoint comparison makes the distinction precise. Evaluating the same historical query first over the old candidate universe and then over the expanded current universe isolates the rank loss attributable to admitted candidates. Catastrophic forgetting instead concerns changes caused by model updating. Candidate-set interference can occur without forgetting, and forgetting can occur while the candidate universe remains fixed. The two mechanisms are complementary explanations for historical degradation.

Current-universe evaluation exposes candidate-set interference, but the continual objectives reviewed above do not directly regulate the ranking pressure that admitted entities create for each historical

answer. This intervention gap defines the problem addressed here. We treat candidate competition as an explicit learning target within a fixed continual KGE process. The next section formalizes the resulting rank decomposition and evaluation objectives.

3 Continual KGE Problem Formulation

3.1 Graph Growth and Candidate Admission

Let $\mathcal{G}_0, \dots, \mathcal{G}_T$ denote a sequence of cumulative knowledge-graph snapshots. At snapshot u , the updated graph is represented as $\mathcal{G}_u = (\mathcal{E}_u, \mathcal{R}_u, \mathcal{F}_u)$, where $\mathcal{E}_u$ and $\mathcal{R}_u$ denote the sets of entities and relations observed up to snapshot u , respectively, and $\mathcal{F}_u \subseteq \mathcal{E}_u \times \mathcal{R}_u \times \mathcal{E}_u$ denotes the corresponding set of observed training triples. The continual setting assumes

$$\mathcal{E}_{u-1} \subseteq \mathcal{E}_u, \quad \mathcal{R}_{u-1} \subseteq \mathcal{R}_u, \quad \mathcal{F}_{u-1} \subseteq \mathcal{F}_u. \quad (1)$$

The entities and relations admitted at update u are

$$\Delta\mathcal{E}_u = \mathcal{E}_u \setminus \mathcal{E}_{u-1}, \quad \Delta\mathcal{R}_u = \mathcal{R}_u \setminus \mathcal{R}_{u-1}. \quad (2)$$

We write $\mathcal{U} \subseteq \{1, \dots, T\}$ for the evaluated updates. For a fact (h, r, t) , tail prediction uses the query $q = (h, r, \cdot)$ with answer $a = t$, whereas head prediction uses $q = (\cdot, r, t)$ with answer $a = h$. We therefore let $\mathcal{D} \subseteq \{\text{head}, \text{tail}\}$ denote the evaluated prediction modes and use $d \in \mathcal{D}$ to identify the missing endpoint. The notation $r(q)$ refers to the relation in q , and e denotes a candidate entity. The experimental choices of $\mathcal{U}$ and $\mathcal{D}$ appear in Section 5.

At update u , a KGE model with parameters $\theta_{m,k,u}$ assigns score $s_{m,k,u}(q, e)$ to candidate e for query q . The indices m and k identify the learning method and training realization. Higher scores indicate stronger preference. The formulation is independent of the embedding architecture.

3.2 Historical Rank Loss Under Candidate Growth

Consider a historical query occurrence $x = (q, a, u, d)$ whose source fact appeared before update u . Its answer a belongs to $\mathcal{E}_{u-1}$. Using the same model state, let C_x^{old} and C_x^{cur} be the filtered candidate sets drawn from $\mathcal{E}_{u-1}$ and $\mathcal{E}_u$. The same filtering rule is applied to both sets, and the answer is retained. Consequently, $C_x^{\text{old}} \subseteq C_x^{\text{cur}}$. We write $\Delta C_x = C_x^{\text{cur}} \setminus C_x^{\text{old}}$ for the candidates introduced by the expansion.

The score function together with a fixed tie rule induces a total order. We write $e \succ_{m,k,x} a$ when candidate e precedes answer a under model state (m, k, u) . For $v \in \{\text{old}, \text{cur}\}$, the corresponding rank is

$$r_{m,k,u}^v(x) = 1 + \sum_{e \in C_x^v \setminus \{a\}} \mathbb{I}[e \succ_{m,k,x} a]. \quad (3)$$

The two ranks satisfy

$$r_{m,k,u}^{\text{cur}}(x) = r_{m,k,u}^{\text{old}}(x) + \sum_{e \in \Delta C_x} \mathbb{I}[e \succ_{m,k,x} a]. \quad (4)$$

This identity isolates the rank loss attributable to newly admitted candidates. It holds even when the ordering among old entities is unchanged and makes no assumption about whether their representations also change. Evaluation in the current candidate universe therefore reveals a source of interference that evaluation over old candidates alone cannot measure [17].

3.3 Query Roles and Evaluation Objectives

We distinguish three query roles. A historical occurrence comes from a fact observed before the current update. A target newcomer occurrence has an answer in $\Delta\mathcal{E}_u$. A query newcomer occurrence has a known query entity in $\Delta\mathcal{E}_u$ and an answer in $\mathcal{E}_{u-1}$. The latter roles capture different learning requirements and are therefore evaluated separately.

For data split σ , let $C^{\sigma,H} \subseteq \mathcal{U} \times \mathcal{D}$ be the predefined historical evaluation cells. Let $Q_{u,d}^{\sigma,H}$ contain the historical occurrences in cell (u, d) . Historical MRR in the current candidate universe is

$$H_{\text{cur},m,k}^{\sigma} = \frac{1}{|C^{\sigma,H}|} \sum_{(u,d) \in C^{\sigma,H}} \frac{1}{|Q_{u,d}^{\sigma,H}|} \sum_{x \in Q_{u,d}^{\sigma,H}} \frac{1}{r_{m,k,u}^{\text{cur}}(x)}. \quad (5)$$

Replacing current ranks with old candidate ranks defines $H_{\text{old},m,k}^{\sigma}$. Their difference is

$$D_{\text{MCL},m,k}^{\sigma} = H_{\text{old},m,k}^{\sigma} - H_{\text{cur},m,k}^{\sigma}. \quad (6)$$

This quantity measures the loss associated with candidate expansion. A small value is not sufficient evidence of useful ranking because both terms can decrease together. We therefore treat current-universe historical MRR as the primary utility outcome and old-universe historical MRR as a preservation outcome.

Let $C^{\sigma,TN} \subseteq \mathcal{U} \times \mathcal{D}$ be the predefined target newcomer cells, and let $Q_{u,d}^{\sigma,TN}$ be their query sets. Target newcomer acquisition is measured by

$$A_{\text{TN},m,k}^{\sigma} = \frac{1}{|C^{\sigma,TN}|} \sum_{(u,d) \in C^{\sigma,TN}} \frac{1}{|Q_{u,d}^{\sigma,TN}|} \sum_{x \in Q_{u,d}^{\sigma,TN}} \frac{1}{r_{m,k,u}^{\text{cur}}(x)}. \quad (7)$$

Each included cell receives equal weight, and queries receive equal weight within a cell. An endpoint is defined only when every predefined query set is nonempty. An empty set is never assigned zero. Query newcomer performance is constructed in the same manner. When the split, method, and training realization are clear, we abbreviate the four endpoints as H_{cur} , H_{old} , D_{MCL} , and A_{TN} .

4 Matched Excess-Outranker Regularization

Equation (4) motivates an answer-relative training signal. Newly admitted entities matter when they move above a historical answer. A raw outranker count is discontinuous, while an unnormalized score gap is not comparable across queries. Penalizing newcomers against zero or an arbitrary old sample also confounds cohort identity with visible structural differences.

MEOR combines a smooth, normalized measure of newcomer competition relative to the answer with an old reference matched on observable structure. It penalizes only the portion of newcomer competition that exceeds this reference. Consequently, a large newcomer aggregate incurs no penalty when a structurally comparable old cohort produces a similar value.

Figure 2 illustrates how candidate-set interference arises and how MEOR addresses it.

4.1 Answer-Relative Newcomer Competition

The regularizer operates on historical occurrences drawn from replay. Throughout this section, we fix a method m and training realization k ; every score-derived quantity below inherits these

indices. For an occurrence $x = (q, a, u, d)$, let $P_u^{\text{tr}}(q, d)$ contain the answers visible in cumulative training facts through update u . The eligible newcomer cohort is

$$N_x = \text{sort}_{\text{id}}(\Delta\mathcal{E}_u \setminus (\{a\} \cup P_u^{\text{tr}}(q, d))). \quad (8)$$

The complete eligible old pool is

$$\mathcal{E}_{u-1}^{-,\text{tr}}(x) = \text{sort}_{\text{id}}(\mathcal{E}_{u-1} \setminus (\{a\} \cup P_u^{\text{tr}}(q, d))). \quad (9)$$

Both sets are defined only from the training prefix. Validation and test labels therefore cannot affect cohort membership. The answer is excluded explicitly, even when it is absent from $P_u^{\text{tr}}(q, d)$.

We draw a scale sample C_x^{scale} from the eligible old pool without using model scores, with

$$|C_x^{\text{scale}}| = \min \{K_{\text{scale}}, |\mathcal{E}_{u-1}^{-,\text{tr}}(x)|\}. \quad (10)$$

Its identifiers are selected independently of model scores and shared by paired methods. This prevents the observed scores from influencing normalization.

Let IQR_x be the interquartile range of the detached finite scores in C_x^{scale} [6]. Independently, let MAD_x^{fb} be the median absolute deviation from the first adequate training-prefix pool in the following hierarchy: the same relation and prediction mode, the same prediction mode, and the complete update. We set this term to zero when no pool is adequate and define

$$\tau_x = \max \{\text{IQR}_x, \eta_{\text{MAD}} \text{MAD}_x^{\text{fb}}, \varepsilon_{\tau}\}. \quad (11)$$

The scale is detached from the gradient. It makes local score differences dimensionless and limits the influence of queries with unusually large score ranges. The parameters $\eta_{\text{MAD}} > 0$ and $\varepsilon_{\tau} > 0$ are numerical safeguards rather than learned quantities. The pooled fallback is robust to isolated extreme scores but can reflect shifts in score location across occurrences. It is therefore not solely a measure of dispersion within individual queries.

For candidate e , the normalized gap relative to the answer is

$$z_x(e) = \frac{s_{m,k,u}(q, e) - s_{m,k,u}(q, a)}{\tau_x}. \quad (12)$$

For a nonempty ordered candidate multiset $C = (e_1, \dots, e_n)$, define

$$U_{\beta}(x, C) = \frac{1}{\beta} \log \left(\frac{1}{n} \sum_{i=1}^n \exp \{ \beta [z_x(e_i)]_+ \} \right), \quad \beta > 0. \quad (13)$$

The normalization by n prevents cohort size alone from increasing the aggregate and makes it invariant to exact replication of the complete multiset. A repeated entity remains a repeated score instance. The aggregate is zero exactly when no member scores above the answer, and it never exceeds the largest positive normalized gap in C . The parameter β determines how strongly the largest active gaps influence the smooth aggregate. The quantity is neither an outranker count nor a probability.

4.2 Score-Blind Matched References

Inspired by Jones and Love [8], we assume old reference should resemble the newcomer cohort in observable structural characteristics while remaining independent of the scores being compared. For entity e and prediction mode d , let $b_u(e, d)$ be its endpoint-degree bin and let $\rho_u(e, d)$ be its set of incident relation-role pairs, both computed from the cumulative training prefix. Matching uses three

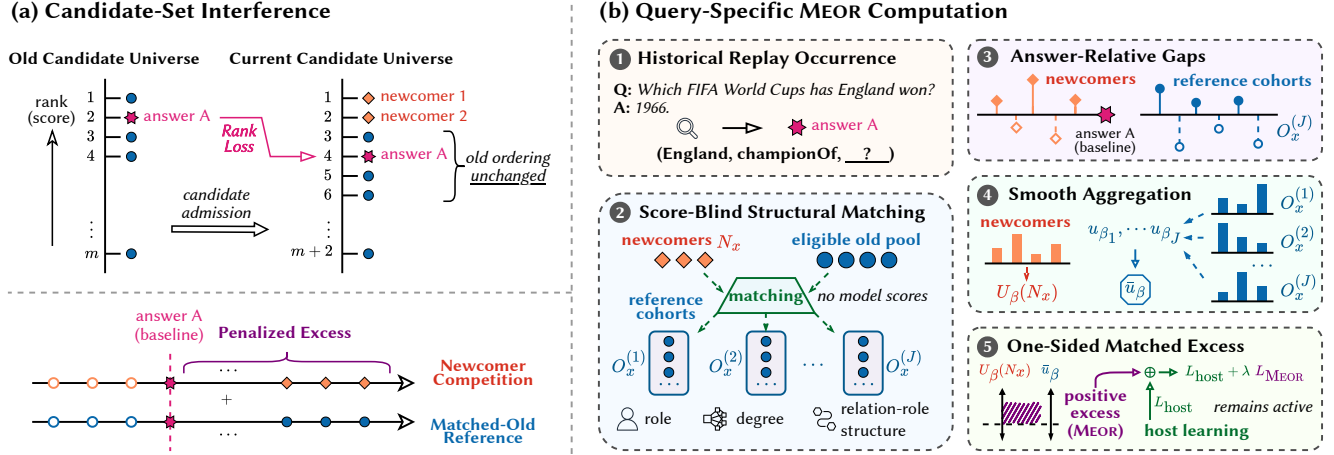


Figure 2: Candidate-set interference and the query-specific computation of MEOR.

nested, score-blind keys. The finest key retains $(d, b_u(e, d), \rho_u(e, d))$, the next drops the relation-role signature, and the terminal key retains only d . For each newcomer, we select the first key whose corresponding old-entity cell is nonempty. Each coarser key only merges cells formed at the preceding level.

Within the selected cell, sampling proceeds without replacement until the cell is exhausted and uses replacement only for any remaining positions. Concatenating the selected candidates across newcomer instances yields an ordered old multiset $O_x^{(j)}$ with $|O_x^{(j)}| = |N_x|$.

We construct J matched draws. Their candidate identifiers are fixed independently of scores and shared by paired methods. Each model then supplies its own scores for those candidates. This design reduces observable cohort imbalance without treating the matched references as causal counterfactuals.

4.3 Matched Excess Objective

Let $\mathcal{B}_u$ denote the current refinement batch. The ordered multiset B_M contains its eligible replay query occurrences in batch order. An occurrence enters B_M only when its newcomer cohort, scale sample, and terminal old matching pool are nonempty and every required score and scale input is finite. Membership is frozen before computing any method's aggregate and is shared by all paired controls. Repeated occurrences remain repeated instances. We write $\mathcal{L}_{\text{host}}(\theta_{m,k,u}; \mathcal{B}_u)$ for the differentiable refinement objective supplied by the underlying continual KGE learner before MEOR is added. This objective retains the learner's positive training signal for current facts and any replay facts selected by the host.

The mean old reference is

$$\bar{U}_O(x) = \frac{1}{J} \sum_{j=1}^J U_{\beta}(x, O_x^{(j)}). \quad (14)$$

For nonempty B_M , the MEOR loss is

$$\mathcal{L}_{\text{MEOR}} = \frac{1}{|B_M|} \sum_{x \in B_M} [U_{\beta}(x, N_x) - \bar{U}_O(x)]_+^2. \quad (15)$$

The mean in Equation (14) is computed before applying the hinge to positive excess. If the newcomer aggregate equals the mean reference, its contribution is zero even when it exceeds an individual draw. Applying a separate hinge to each draw before averaging would impose downward pressure in that case. The squared hinge has both value and gradient zero at the boundary, and its penalty increases with positive excess.

If B_M is empty, we define $\mathcal{L}_{\text{MEOR}} = 0$. With regularization coefficient $\lambda \geq 0$, the complete refinement objective is

$$\mathcal{L}(\theta_{m,k,u}; \mathcal{B}_u) = \mathcal{L}_{\text{host}}(\theta_{m,k,u}; \mathcal{B}_u) + \lambda \mathcal{L}_{\text{MEOR}}. \quad (16)$$

The host objective retains its positive learning signal for newly admitted entities. Thus MEOR controls excessive competition on historical queries rather than treating every newcomer as an error.

4.4 Controls for the Proposed Mechanism

The central contribution of MEOR is to combine smooth answer-relative aggregation with score-blind structural matching. Two controls isolate these components. Both control losses are defined as zero when B_M is empty; the displayed averages apply otherwise. The Matched Maximum Regularizer (MMR) replaces the smooth aggregate with

$$V(x, C) = \max_{1 \leq i \leq |C|} [z_x(e_i)]_+ \quad (17)$$

and uses the same matched reference cohorts. Its loss is

$$\mathcal{L}_{\text{MMR}} = \frac{1}{|B_M|} \sum_{x \in B_M} \left[V(x, N_x) - \frac{1}{J} \sum_{j=1}^J V(x, O_x^{(j)}) \right]_+^2. \quad (18)$$

This comparison isolates the effect of distributing pressure across the active score tail rather than controlling only the largest gap.

The Unmatched Old Regularizer (UOR) retains U_{β} but replaces each matched reference with an equal-cardinality multiset $R_x^{(j)}$ drawn from the terminal role-only old pool. It uses

$$\mathcal{L}_{\text{UOR}} = \frac{1}{|B_M|} \sum_{x \in B_M} \left[U_{\beta}(x, N_x) - \frac{1}{J} \sum_{j=1}^J U_{\beta}(x, R_x^{(j)}) \right]_+^2. \quad (19)$$

Algorithm 1: MEOR objective for a refinement batch

Input: training-prefix graph at update u , refinement batch $\mathcal{B}_u$, model scores $s_{m,k,u}$, host objective $\mathcal{L}_{\text{host}}$, and parameters (J, β, λ)

Output: regularized refinement objective $\mathcal{L}(\theta_{m,k,u}; \mathcal{B}_u)$

```

1  $B_M \leftarrow \emptyset$ ;
2 foreach historical replay occurrence  $x = (q, a, u, d) \in \mathcal{B}_u$  do
3   construct  $N_x$  using Equation (8);
4   draw the score-blind scale sample  $C_x^{\text{scale}}$ ;
5   construct  $J$  score-blind matched cohorts  $\{O_x^{(j)}\}_{j=1}^J$  with
       $|O_x^{(j)}| = |N_x|$ ;
6   if  $x$  has all required nonempty cohorts and finite score
      and scale inputs then
7      $B_M \leftarrow B_M \cup \{x\}$ ;
8  $R \leftarrow 0$ ;
9 foreach  $x \in B_M$  do
10  compute the detached scale  $\tau_x$  using Equation (11);
11  compute  $z_x(e)$  using Equation (12) for all
       $e \in N_x \cup \bigcup_{j=1}^J O_x^{(j)}$ ;
12   $\bar{U}_O(x) \leftarrow \frac{1}{J} \sum_{j=1}^J U_\beta(x, O_x^{(j)})$ ;
13   $R \leftarrow R + [U_\beta(x, N_x) - \bar{U}_O(x)]_+^2$ ;
14  $\mathcal{L}_{\text{MEOR}} \leftarrow \begin{cases} 0, & B_M = \emptyset, \\ R/|B_M|, & \text{otherwise,} \end{cases}$ ;
15  $\mathcal{L}(\theta_{m,k,u}; \mathcal{B}_u) \leftarrow \mathcal{L}_{\text{host}}(\theta_{m,k,u}; \mathcal{B}_u) + \lambda \mathcal{L}_{\text{MEOR}}$ ;

```

This comparison isolates the contribution of structural matching. Replay uses the same host refinement with zero regularization and therefore isolates the contribution of the intervention.

A persistent admission-cohort calibrator provides a score-only control. Let $c(e)$ denote the update at which entity e first entered the graph. For training realization k , $\alpha_{k,c,r,d} > 0$ is a multiplicative coefficient and $g_{k,c,r,d} \in \mathbb{R}$ is a dimensionless offset coefficient. Both are fitted from replay validation scores for admission cohort c , relation r , and prediction mode d when the cohort is admitted, then retained at later updates. The detached statistic $\bar{\tau}_{k,u,r,d}^{\text{val}}$ is the median finite validation scale for the same relation and prediction mode at update u , with prediction-mode and update-level fallbacks. When this scale is available, the calibrator multiplies the raw replay score $s_{\text{Replay},k,u}(q, e)$ by $\alpha_{k,c(e),r(q),d}$ and adds the scale-adjusted offset $g_{k,c(e),r(q),d} \bar{\tau}_{k,u,r,d}^{\text{val}}$. If no finite scale summary is available, the raw replay score is retained. The initial cohort uses the identity transform, $\alpha = 1$ and $g = 0$. This control adds no training trajectory and separates an intervention applied during training from a post hoc adjustment of cohort score scales.

All controls share the host model, update data, replay information, optimization budget, and trainable parameter scope. The matched maximum control and MEOR also share their matched reference identifiers. Their concrete training and calibration settings appear in Section 5.

4.5 Interpretation and Computational Cost

Equation (15) has zero gradient whenever newcomer competition is no greater than the mean matched reference. When active, the smooth aggregate distributes pressure across several above-answer candidates rather than only the largest one. The loss acts on scores, so the evaluation measures historical current-universe ranking and target-newcomer acquisition directly. These outcomes capture whether score changes cross the answer and whether legitimate newcomer learning is preserved.

If one candidate score requires $O(p)$ operations, candidate scoring for one refinement batch without score reuse is

$$O\left(p \sum_{x \in B_M} ((1+J)|N_x| + K_{\text{scale}})\right). \quad (20)$$

The complete newcomer cohort and every reference cohort are scored, so computational cost increases with the size of the admission batch. Approximate cohort scoring lies outside the scope of this study.

5 Experiments and Results

The evaluation centers on whether MEOR improves historical ranking by reducing candidate-set interference. ENTITY-ComplEx provides the primary comparison with replay, persistent calibration, MMR, and UOR. Direct ablations identify the contribution of each component, while experiments on FBInc-S and FBInc-L examine transfer across different rates of entity admission, embedding models, and continual KGE hosts. All comparative claims are drawn from paired runs within the same experimental setting.

5.1 Experimental Setting

Our experiments use four five-snapshot entity-growth streams. ENTITY is the primary FB15K-237 [24] stream introduced with LKGE [3]. FBInc-S and FBInc-L are the small-growth and large-growth streams distributed with the informed-initialization benchmark [16]. WN-CKGE is the WordNet stream used by FastKGE [14]. It represents the applicability boundary discussed in Section 6 and is not treated as a positive efficacy setting. Snapshot 0 initializes the model, and Snapshots 1 through 4 form the continual sequence. Table 1 reports the cumulative graph size and the facts introduced at each snapshot.

We evaluate filtered head and tail prediction after every update. The filter removes other positives visible in the current graph prefix while retaining the recorded answer, and exact score ties are resolved by increasing canonical entity identifier. Each endpoint assigns equal weight to the update and prediction-direction combinations defined in Section 3. Historical current-universe MRR H_{cur} measures the ranking of established answers against every entity available at the current update. Historical old-universe MRR H_{old} restricts the same evaluation to previously known candidates. Their difference, $D_{\text{MCI}} = H_{\text{old}} - H_{\text{cur}}$, measures the rank loss caused by candidate admission. Target-newcomer MRR A_{TN} evaluates queries whose correct answer is newly admitted.

Although MEOR is independent of the embedding architecture, the primary study uses ComplEx with 200 complex coordinates and float64 arithmetic. Every method optimizes the same soft-plus link-prediction objective with Adam, a learning rate of 10^{-4} ,

Table 1: Dataset statistics. Entity and relation counts are cumulative; Facts is the number of triples introduced at each snapshot.

Dataset	Snapshot 0			Snapshot 1			Snapshot 2			Snapshot 3			Snapshot 4		
	$ \mathcal{E} $	$ \mathcal{R} $	Facts	$ \mathcal{E} $	$ \mathcal{R} $	Facts	$ \mathcal{E} $	$ \mathcal{R} $	Facts	$ \mathcal{E} $	$ \mathcal{R} $	Facts	$ \mathcal{E} $	$ \mathcal{R} $	Facts
ENTITY	2,909	233	46,388	5,817	236	72,111	8,725	236	73,785	11,633	237	70,506	14,541	237	47,326
FBInc-S	2,909	233	46,388	2,919	233	235	2,930	233	152	2,940	233	180	2,950	233	239
FBInc-L	2,909	233	46,388	3,010	233	3,110	3,110	234	2,736	3,211	234	2,908	3,312	234	3,105
WN-CKGE	24,567	11	55,801	28,660	11	9,300	32,754	11	9,300	36,848	11	9,300	40,943	11	9,302

$(\beta_1, \beta_2) = (0.9, 0.999)$, $\epsilon = 10^{-8}$, and no weight decay. Training uses batches of 2,048 positive facts with ten negative samples per fact. Base training runs for at most 200 epochs with an early-stopping patience of three, and replay retains at most 2,048 historical facts. Continual refinement makes one ordered pass over the facts introduced at the current snapshot. Previously learned entity and relation embeddings remain fixed during refinement, so the intervention updates only newly admitted parameters.

We instantiate MEOR with $\beta = 5$, four matched reference cohorts, and at most 256 old entities in each scale sample. Reference construction matches prediction role, degree bin, and incident relation-role signature before falling back to role and degree and then to role alone. The regularization coefficient for each ENTITY method is fixed before evaluation by the score-blind gradient-balance rule. Replay uses zero regularization. Persistent calibration applies the affine transformation $\alpha s + g\bar{r}^{\text{val}}$ to newcomer scores over the fixed grid

$$\alpha \in \{0.5, 0.75, 1, 1.25, 1.5\}, \quad g \in \{-1, -0.5, 0, 0.5, 1\},$$

where $\bar{r}^{\text{val}}$ is the corresponding replay-validation score scale. Calibration retains transformations whose A_{TN} is no more than 0.005 below replay and selects the retained transformation with the highest H_{cur} .

The primary, ablation, and transfer studies each use eight paired runs. Within a pair, the compared methods share initialization, graph stream, fact order, replay sample, negative samples, stopping decisions, and evaluation queries. Comparisons involving matched references also share the relevant reference identities. The direct ablations use the same coefficient-selection rule and change only the component named in Table 4. The transfer study adds MEOR to replay with ComplEx [25], DistMult [27], and TransE [1] and to the LKGE [3] and IncDE [13] update procedures with their TransE backbones. Comparisons are made only within the same dataset, host, and backbone.

For the primary and ablation studies, LB denotes the one-sided 95% lower confidence bound computed from the eight unrounded paired effects. A positive lower bound supports an improvement in H_{cur} or a reduction in D_{MCI} . Preservation requires the lower bound for ΔH_{old} or ΔA_{TN} to remain above -0.005 . W/T/L counts positive, zero, and negative full-precision H_{cur} effects. The transfer study reports two-sided paired 95% confidence intervals for H_{cur} and the same one-sided lower bound for ΔA_{TN} . These intervals are unadjusted summaries of prespecified, cell-specific comparisons rather than simultaneous family-wise intervals.

Table 2: Absolute endpoint performance of MEOR and mechanism-matched controls in the primary ENTITY-ComplEx setting.

Method	$H_{\text{cur}} \uparrow$	$H_{\text{old}} \uparrow$	$D_{\text{MCI}} \downarrow$	$A_{\text{TN}} \uparrow$
Replay	0.0830	0.1199	0.0369	0.1700
Persistent calibration	0.0816	–	–	0.1753
MMR	0.0836	0.1200	0.0364	0.1695
UOR	0.0842	0.1200	0.0358	0.1698
MEOR	0.0887	0.1201	0.0314	0.1715

5.2 Primary Results on ENTITY

Table 2 reports the endpoint means for ENTITY with ComplEx. At 0.0887, MEOR attains the highest H_{cur} . It also records the highest H_{old} and the lowest D_{MCI} , while increasing A_{TN} above replay, MMR, and UOR. Persistent calibration attains the highest A_{TN} but a lower H_{cur} . The calibration analysis reports only H_{cur} and A_{TN} ; H_{old} and D_{MCI} were not recorded for this control. All endpoints are equal-weight means over update and prediction-direction combinations.

The absolute endpoints show the aggregate ordering, while the paired analysis determines whether that ordering is consistent across runs. Table 3 reports the paired effects of MEOR relative to each mechanism-matched control.

MEOR improves H_{cur} against every control. The mean gain is 0.0057 over replay, 0.0071 over persistent calibration, 0.0051 over MMR, and 0.0045 over UOR. The corresponding lower bounds are 0.0052, 0.0065, 0.0047, and 0.0030, respectively. The effect is positive in all eight paired runs against replay, persistent calibration, and MMR and in seven of eight runs against UOR.

The decomposition against replay locates the gain in candidate admission rather than parameter retention. The 0.0057 increase in H_{cur} combines a much smaller 0.0002 change in H_{old} with a 0.0055 reduction in D_{MCI} . The same pattern appears against MMR, where a 0.0051 gain combines a 0.0001 old-universe change with a 0.0050 interference reduction, and against UOR, where the corresponding effects are 0.0045, 0.0001, and 0.0044. Every interference-reduction lower bound is positive. These paired decompositions indicate that the gains arise primarily from reduced competition by newly admitted entities, while the mean changes in old-universe ranking are small.

Our results meet the prespecified criterion for retained and newcomer performance. The H_{old} lower bounds against replay, MMR, and UOR are -0.0001 , -0.0003 , and -0.0011 , respectively, all above

Table 3: Paired effects of MEOR relative to mechanism-matched controls on ENTITY-CompLex across eight paired training realizations.

Comparator	$\Delta H_{\text{cur}} \uparrow$		$\Delta H_{\text{old}} \uparrow$		$D_{\text{MCI}} \text{ reduction} \uparrow$		$\Delta A_{\text{TN}} \uparrow$		W/T/L
	Mean	LB	Mean	LB	Mean	LB	Mean	LB	
Replay	0.0057	0.0052	0.0002	-0.0001	0.0055	0.0051	0.0015	0.0008	8/0/0
Persistent calibration	0.0071	0.0065	-	-	-	-	-0.0038	-0.0043	8/0/0
MMR	0.0051	0.0047	0.0001	-0.0003	0.0050	0.0049	0.0020	0.0015	8/0/0
UOR	0.0045	0.0030	0.0001	-0.0011	0.0044	0.0041	0.0017	0.0012	7/0/1

the -0.005 margin. MEOR improves A_{TN} against replay, MMR, and UOR by 0.0015, 0.0020, and 0.0017, with positive lower bounds of 0.0008, 0.0015, and 0.0012. Persistent calibration retains a mean A_{TN} advantage of 0.0038. The one-sided lower bound for the corresponding MEOR-minus-calibration effect is -0.0043 , which also remains above the preservation margin.

5.3 Mechanism Analysis

The primary controls distinguish the two central components of MEOR. MMR preserves query-specific matched references and old-reference centering but replaces smooth tail aggregation with a maximum. UOR preserves smooth aggregation and query-specific drawing but uses role-only old references. The shuffled-reference comparison breaks query-specific assignment while retaining the matched pools, and the uncentered comparison removes subtraction of the old reference. All four ablations use the same score-blind coefficient rule and paired inputs. Table 4 reports the resulting effects.

Each component makes a distinct contribution. Smooth tail aggregation produces the largest gain over its matched alternative, with a mean H_{cur} effect of 0.0051 and a lower bound of 0.0047. Structural matching follows closely with a mean effect of 0.0045 and a lower bound of 0.0030. Old-reference centering contributes 0.0011, and query-specific assignment contributes 0.0008, with lower bounds of 0.0006 and 0.0005, respectively. Every ablation also yields a positive lower bound for reducing D_{MCI} . The A_{TN} effects are positive on average in all four contrasts, and their lower bounds remain above the preservation margin. The complete method therefore benefits from each element of its reference construction and aggregation.

5.4 Transfer Across Streams, Backbones, and Hosts

The transfer study treats MEOR as an intervention within a named host rather than as a replacement continual KGE architecture. Replay is evaluated with CompLex, DistMult, TransE, LKGE-TransE, and IncDE-TransE. Each row of Table 5 is a separate paired comparison within the specified dataset, host, and backbone. The H_{cur} column reports the mean effect and its two-sided 95% interval; the final column reports the one-sided lower bound for ΔA_{TN} .

Adding MEOR improves H_{cur} in all ten transfer settings, and every 95% interval excludes zero. The replay gains range from 0.0020 to 0.0026 across the two streams and three backbones. The gains within LKGE are 0.0023 on FBInc-S and 0.0026 on FBInc-L, while

those within IncDE are 0.0029 and 0.0032. Four settings improve in all eight paired runs, and each remaining setting improves in at least seven.

The A_{TN} effects are slightly negative, ranging from -0.0001 to -0.0015 , but every lower bound remains above the -0.005 preservation margin. The transfer study therefore shows that MEOR improves historical current-universe ranking across the reported growth regimes, embedding models, and continual hosts while preserving newcomer acquisition under the stated criterion.

Together, the primary results, direct ablations, and transfer study support a coherent account of MEOR. Its historical-ranking gains arise chiefly from reduced candidate-set interference, each component contributes to that reduction, and the intervention remains effective across the named host configurations without materially impairing newly admitted entities.

6 Limitations

MEOR is designed for entity-growth streams in which newly admitted entities become overly competitive answers to historical queries. WN-CKGE defines the complementary boundary. In the registered WN-CKGE diagnostic, MEOR and MMR each had zero active records among 64 calibration records, indicating that aggregate newcomer pressure did not exceed the matched old-candidate reference under the fixed construction. Their regularization terms therefore contributed neither loss nor gradient in the evaluated records. The suppression mechanism had no direct optimization effect because there was no excess candidate pressure to remove. This selective activation is central to the method. MEOR intervenes when entity admission disturbs historical ranking and otherwise leaves refinement governed by the host objective. WN-CKGE thus documents an inactive applicability regime rather than a positive efficacy comparison.

The method does not address every source of degradation in a continual knowledge graph. Relation-only or fact-only updates can still alter established representations, but they do not create the entity-candidate expansion captured by D_{MCI} . MEOR is consequently a complement to replay, distillation, and parameter-stability methods rather than a replacement for them. Its role is specific and measurable. It controls admission-induced competition while the host learner remains responsible for retaining prior knowledge and acquiring new facts.

The quality of the reference depends on the structural information available for matching. Prediction role, degree, and incident relation-role signatures provide a score-blind basis for comparison,

Table 4: Direct ablations of MEOR’s reference construction and aggregation on ENTITY–Complex.

Isolated property	Comparator	$\Delta H_{\text{cur}} \uparrow$		$D_{\text{MCI}} \text{ reduction} \uparrow$		$\Delta A_{\text{TN}} \uparrow$	
		Mean	LB	Mean	LB	Mean	LB
Smooth tail aggregation	MMR (maximum)	0.0051	0.0047	0.0050	0.0049	0.0020	0.0015
Structural matching	UOR (role-only references)	0.0045	0.0030	0.0044	0.0041	0.0017	0.0012
Query-specific assignment	Shuffled matched references	0.0008	0.0005	0.0007	0.0002	0.0002	−0.0002
Old-reference centering	Uncentered smooth penalty	0.0011	0.0006	0.0009	0.0003	0.0003	−0.0002

Table 5: Paired effects of adding MEOR to replay hosts and source-bound continual-learning adapters on FBInc-S and FBInc-L.

Dataset	Host family	Host–backbone	$\Delta H_{\text{cur}} \uparrow$	Paired 95% CI	W/T/L	$\Delta A_{\text{TN}} \uparrow$	$\Delta A_{\text{TN}} \text{ LB}$
FBInc-S	Replay	Complex	0.0026	[0.0017, 0.0043]	8/0/0	−0.0004	−0.0011
FBInc-S	Replay	DistMult	0.0022	[0.0010, 0.0034]	7/1/0	−0.0002	−0.0008
FBInc-S	Replay	TransE	0.0020	[0.0009, 0.0031]	7/1/0	−0.0001	−0.0006
FBInc-S	Source-Bound	LKGE–TransE	0.0023	[0.0011, 0.0035]	7/0/1	−0.0012	−0.0024
FBInc-S	Source-Bound	IncDE–TransE	0.0029	[0.0016, 0.0042]	8/0/0	−0.0010	−0.0021
FBInc-L	Replay	Complex	0.0026	[0.0019, 0.0045]	8/0/0	−0.0008	−0.0016
FBInc-L	Replay	DistMult	0.0023	[0.0011, 0.0035]	7/1/0	−0.0003	−0.0009
FBInc-L	Replay	TransE	0.0021	[0.0010, 0.0032]	7/1/0	−0.0002	−0.0007
FBInc-L	Source-Bound	LKGE–TransE	0.0026	[0.0012, 0.0040]	7/0/1	−0.0015	−0.0029
FBInc-L	Source-Bound	IncDE–TransE	0.0032	[0.0018, 0.0046]	8/0/0	−0.0014	−0.0027

but they do not form a causal counterfactual. Sparse structural cells also require coarser fallbacks, which reduce the specificity of the reference cohort. The precision of the reference is therefore bounded by the structural resolution available in each update. Computational cost presents a separate boundary because exact cohort scoring grows with the number of admitted entities and matched draws. Very large admission batches therefore require efficient batching or approximate cohort evaluation. The reported experiments establish effectiveness on ENTITY, FBInc-S, and FBInc-L, while WN-CKGE identifies the inactive boundary of the intervention. The empirical claim is limited to the reported streams, backbones, hosts, and fixed aggregation and matching settings. Other prediction tasks require candidate and reference definitions appropriate to their ranking spaces.

7 Conclusion

Continual knowledge graph embedding must account for more than changes to previously learned representations. When the entity vocabulary grows, every admitted entity becomes a new candidate for historical queries and can lower the rank of an established answer even when the ordering among old entities is preserved. We formalized this effect as candidate-set interference and separated it from parameter forgetting through a same-checkpoint rank decomposition. We then introduced MEOR, a host-level regularizer that compares answer-relative newcomer competition with score-blind, structurally matched old references. Its one-sided objective suppresses only the excess competition attributable to admission, while the host loss continues to learn legitimate new knowledge.

The experiments show that this intervention improves the ranking quantity it is designed to control. On ENTITY–Complex, MEOR increases historical current-universe MRR by 0.0057 over replay and reduces D_{MCI} by 0.0055, while satisfying the preservation criteria for old-universe ranking and newcomer acquisition. Comparisons with persistent calibration, MMR, and UOR show that the gain persists relative to score rescaling, maximum-only control, and unmatched references. Direct ablations further support the contributions of smooth tail aggregation, structural matching, query-specific assignment, and old-reference centering. Across FBInc-S and FBInc-L, adding MEOR improves historical current-universe MRR in all ten reported combinations of stream, host, and backbone, with every paired confidence interval excluding zero and every newcomer-acquisition bound remaining within the preservation margin.

These findings establish candidate-set interference as a distinct and controllable source of historical rank loss in continual KGE. They also show that it can be addressed without redesigning the underlying embedding architecture or continual learner. By making admission-induced candidate pressure part of refinement, MEOR extends continual learning from preserving representations to managing the ranking consequences of graph growth.

Ethical Considerations

This work studies continual knowledge graph embedding using established research benchmarks and does not involve human participants or the collection of new personal data. Its direct ethical risk is therefore limited. Nevertheless, knowledge graphs may contain inaccurate, outdated, sensitive, or socially biased information,

and link prediction can reproduce these properties or infer unrecorded associations. Because MEOR regulates the competitive pressure introduced by newly admitted entities, an improperly configured deployment could also favor established entities and delay the recognition of legitimate new facts or underrepresented entities. The one-sided intervention and the explicit preservation of newcomer acquisition reduce this risk, but they do not replace careful dataset and deployment auditing. Applications involving people or consequential decisions should examine subgroup ranking behavior, verify data provenance, restrict access to sensitive predictions, provide correction and removal procedures, and retain appropriate human oversight. The method is intended for research on knowledge graph ranking and should not be used as the sole basis for high-stakes decisions.

References

- [1] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In *Advances in Neural Information Processing Systems*, C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger (Eds.), Vol. 26. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2013/file/1cecc7a77928ca8133fa24680a88d2f9-Paper.pdf
- [2] Jiahang Cao, Jinyuan Fang, Zaiqiao Meng, and Shangsong Liang. 2024. Knowledge Graph Embedding: A Survey from the Perspective of Representation Spaces. *ACM Comput. Surv.* 56, 6, Article 159 (March 2024), 42 pages. doi:10.1145/3643806
- [3] Yuanning Cui, Yuxin Wang, Zequn Sun, Wenqiang Liu, Yiqiao Jiang, Kexin Han, and Wei Hu. 2023. Lifelong embedding learning and transfer for growing knowledge graphs. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 37. AAAI Press, 4217–4224. doi:10.1609/aaai.v37i4.25539
- [4] Angel Daruna, Mehul Gupta, Swayambhu Nath Ray, and Sonia Chernova. 2021. Continual Learning of Knowledge Graph Embeddings. *IEEE Robotics and Automation Letters* 6, 2 (2021), 1128–1135. doi:10.1109/LRA.2021.3056071
- [5] Shib Sankar Dasgupta, Swayambhu Nath Ray, and Partha Talukdar. 2018. HyTE: Hyperplane-based Temporally aware Knowledge Graph Embedding. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii (Eds.). Association for Computational Linguistics, Brussels, Belgium, 2001–2011. doi:10.18653/v1/D18-1225
- [6] Frederik Michel Dekking, Cornelis Kraaikamp, Hendrik Paul Lopuhaä, and Ludolf Erwin Meester. 2005. *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Springer London, London, United Kingdom, 234–236. doi:10.1007/1-84628-168-7
- [7] Rishab Goel, Seyed Mehran Kazemi, Marcus Brubaker, and Pascal Poupart. 2020. Diachronic Embedding for Temporal Knowledge Graph Completion. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. AAAI Press, 3988–3995. doi:10.1609/aaai.v34i04.5815
- [8] Matt Jones and Bradley C. Love. 2007. Beyond common features: The role of roles in determining similarity. *Cognitive Psychology* 55, 3 (2007), 196–231. doi:10.1016/j.cogpsych.2006.09.004
- [9] Xiaoyu Kou, Yankai Lin, Shaobo Liu, Peng Li, Jie Zhou, and Yan Zhang. 2020. Disentangle-based Continual Graph Representation Learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 2961–2972. doi:10.18653/v1/2020.emnlp-main.237
- [10] Kyung-Hwan Lee and Dong-Wan Choi. 2026. STARK: Structure-Aware and Adaptive Representation Learning for Continual Knowledge Graph Embedding. In *Proceedings of the ACM Web Conference 2026* (United Arab Emirates) (WWW '26). Association for Computing Machinery, New York, NY, USA, 3941–3951. doi:10.1145/3774904.3792391
- [11] Linyu Li, Zhi Jin, Yuanpeng He, Dongming Jin, Yichi Zhang, Haoran Duan, Xuan Zhang, Zhengwei Tao, and Tashi Nyima. 2026. Learning to Evolve: Bayesian-Guided Continual Knowledge Graph Embedding. In *Proceedings of the ACM Web Conference 2026* (United Arab Emirates) (WWW '26). Association for Computing Machinery, New York, NY, USA, 4372–4383. doi:10.1145/3774904.3792692
- [12] Yifei Li, Lingling Zhang, Hang Yan, Tianzhe Zhao, Zihan Ma, Muye Huang, and Jun Liu. 2025. SAGE: Scale-Aware Gradual Evolution for Continual Knowledge Graph Embedding. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2* (Toronto ON, Canada) (KDD '25). Association for Computing Machinery, New York, NY, USA, 1600–1611. doi:10.1145/3711896.3737115
- [13] Jiajun Liu, Wenjun Ke, Peng Wang, Ziyu Shang, Jinhua Gao, Guozheng Li, Ke Ji, and Yanhe Liu. 2024. Towards Continual Knowledge Graph Embedding via Incremental Distillation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. AAAI Press, 8759–8768. doi:10.1609/aaai.v38i8.28722
- [14] Jiajun Liu, Wenjun Ke, Peng Wang, Jiahao Wang, Jinhua Gao, Ziyu Shang, Guozheng Li, Zijie Xu, Ke Ji, and Yining Li. 2024. Fast and Continual Knowledge Graph Embedding via Incremental LoRA. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, Kate Larson (Ed.). International Joint Conferences on Artificial Intelligence Organization, 2198–2206. doi:10.24963/ijcai.2024/243 Main Track.
- [15] Mehrnoosh Mirtaheeri, Mohammad Rostami, and Aram Galstyan. 2023. History repeats: Overcoming catastrophic forgetting for event-centric temporal knowledge graph completion. In *Findings of the Association for Computational Linguistics: ACL 2023*, Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 7740–7755. doi:10.18653/v1/2023.findings-acl.490
- [16] Gerard Pons, Besim Bilalli, and Anna Queralt. 2026. Improving continual knowledge graph embedding via informed initialization. *Neurocomputing* 695 (2026), 134045. doi:10.1016/j.neucom.2026.134045
- [17] Gerard Pons, Carlos Escolano, Besim Bilalli, and Anna Queralt. 2026. Revisiting Catastrophic Forgetting in Continual Knowledge Graph Embedding. arXiv:2604.19401 [cs.LG] doi:10.48550/arXiv.2604.19401
- [18] Jing Qi, Yuxiang Wang, Zhiyuan Yu, Xiaoliang Xu, Yuanshi Zheng, and Tianxing Wu. 2026. Multi-Faceted Continual Knowledge Graph Embedding for Semantic-Aware Link Prediction. In *Proceedings of the 49th International ACM SIGIR Conference on Research and Development in Information Retrieval (Australia) (SIGIR '26)*. Association for Computing Machinery, New York, NY, USA, 1462–1472. doi:10.1145/3805712.3809683
- [19] David Rolnick, Arun Ahuja, Jonathan Schwarz, Timothy Lillicrap, and Gregory Wayne. 2019. Experience Replay for Continual Learning. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.), Vol. 32. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2019/file/fa7cdfad1a5aaf8370ebda47a1ffc3-Paper.pdf
- [20] Andrea Rossi, Denilson Barbosa, Donatella Firmani, Antonio Matinata, and Paolo Merialdo. 2021. Knowledge Graph Embedding for Link Prediction: A Comparative Analysis. *ACM Trans. Knowl. Discov. Data* 15, 2, Article 14 (Jan. 2021), 49 pages. doi:10.1145/3424672
- [21] Andrea Rossi, Donatella Firmani, Paolo Merialdo, and Tommaso Teofili. 2022. Explaining Link Prediction Systems based on Knowledge Graph Embeddings. In *Proceedings of the 2022 International Conference on Management of Data (Philadelphia, PA, USA) (SIGMOD '22)*. Association for Computing Machinery, New York, NY, USA, 2062–2075. doi:10.1145/3514221.3517887
- [22] Xiangjun Shi, Chong Mu, Lizong Zhang, Jinchuan Zhang, and Qianghua Yuan. 2026. DTD: Dynamic temperature distillation for continual knowledge graph embedding. *Neurocomputing* 701 (2026), 134563. doi:10.1016/j.neucom.2026.134563
- [23] Ailin Song, Yuhong Chen, Yusong Wang, Shuai Zhong, and Mingkun Xu. 2025. Orchestrating Plasticity and Stability: A Continual Knowledge Graph Embedding Framework with Bio-Inspired Dual-Mask Mechanism. In *Proceedings of the 16th Asian Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 260)*, Vu Nguyen and Hsuan-Tien Lin (Eds.). PMLR, 1208–1223. <https://proceedings.mlr.press/v260/song25a.html>
- [24] Kristina Toutanova and Danqi Chen. 2015. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, Alexandre Allauzen, Edward Grefenstette, Karl Moritz Hermann, Hugo Larochelle, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Beijing, China, 57–66. doi:10.18653/v1/W15-4007
- [25] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Eric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *Proceedings of The 33rd International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 48)*, Maria Florina Balcan and Kilian Q. Weinberger (Eds.). PMLR, New York, New York, USA, 2071–2080. <https://proceedings.mlr.press/v48/trouillon16.html>
- [26] Changyi Xiao, Xiangnan He, and Yixin Cao. 2023. Knowledge Graph Embedding by Normalizing Flows. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. AAAI Press, 4756–4764. doi:10.1609/aaai.v37i4.25600
- [27] Bishan Yang, Scott Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. In *Proceedings of the International Conference on Learning Representations (ICLR) 2015*. <https://www.microsoft.com/en-us/research/publication/embedding-entities-and-relations-for-learning-and-inference-in-knowledge-bases/>
- [28] Yuchao Zhang, Xiangjie Kong, Zhehui Shen, Jianxin Li, Qiuhua Yi, Guojiang Shen, and Bo Dong. 2024. A survey on temporal knowledge graph embedding: Models and applications. *Knowledge-Based Systems* 304 (2024), 112454. doi:10.1016/j.knsys.2024.112454
- [29] Lijing Zhu, Dong Hyun Jeon, Wenbo Sun, Li Yang, Yixin Xie, and Shuteng Niu. 2024. Flexible Memory Rotation (FMR): Rotated Representation with Dynamic

- Regularization to Overcome Catastrophic Forgetting in Continual Knowledge Graph Learning. In *2024 IEEE International Conference on Big Data (BigData)*. 6180–6189. doi:10.1109/BigData62323.2024.10825244
- [30] Lijing Zhu, Qizhen Lan, Qing Tian, Wenbo Sun, Li Yang, Lu Xia, Yixin Xie, Xi Xiao, Tiehang Duan, Cui Tao, and Shuteng Niu. 2025. ETT-CKGE: Efficient Task-Driven Tokens for Continual Knowledge Graph Embedding. In *Machine Learning and Knowledge Discovery in Databases. Research Track*, Rita P. Ribeiro, Bernhard Pfahringer, Nathalie Japkowicz, Pedro Larrañaga, Alípio M. Jorge, Carlos Soares, Pedro H. Abreu, and João Gama (Eds.). Springer Nature Switzerland, Cham, 481–496.