CareWeavers BiKE 2026: BERT-Initialised Graph Embeddings for Natural Product Knowledge Extraction

Khyati Patni[†1], Daksh Sammi[†1], Rahul Kharkwal[1] and Diksha Kashyap[1]

[1]*Microcrispr Pvt. Ltd., New Delhi, 110001, India*

Abstract

We present CareWeavers BiKE 2026, a BERT-initialised knowledge graph embedding pipeline for biochemical link prediction on the NatUKE natural product extraction benchmark. The input to our system is a set of biochemical paper DOIs. For each DOI, we retrieve title and abstract metadata from CrossRef and encode paper nodes and property-value label nodes using SciBERT and PubMedBERT into 768-dimensional vectors. We construct a five-relation heterogeneous knowledge graph over 142 DOI nodes and five NuBBE property types: compound name, bioactivity, collection species, collection site, and isolation type.

We evaluate 11 configurations: two BERT cosine baselines, two topology-only walk baselines, one spectral heat-kernel baseline, five SciBERT-initialised graph embedding models, and one normalised ensemble. Seven of the eleven configurations use BERT-derived representations either directly as cosine baselines or as graph-model initialisation. Under the official 3rd-stage 60%/40% NatUKE split, the best CareWeavers configuration per task exceeds the published NatUKE reference baseline on 4 of 5 property types. RotatE achieves Hits@1 = 0.931 on isolation type (EPHEN reference: 0.754), Hits@5 = 0.523 on bioactivity (EPHEN reference: 0.656). GraphSAGE achieves Hits@20 = 0.688 on collection site (EPHEN reference: 0.550). PubMedBERT cosine similarity achieves Hits@50 = 0.778 on species (Metapath2Vec reference: 0.429). Metapath2Vec achieves Hits@50 = 0.406 on compound name (Metapath2Vec reference: 0.093).

Keywords

knowledge graph embedding, natural product extraction, NatUKE, SciBERT, PubMedBERT, RotatE, GraphSAGE, BERT initialisation

Code and reproducibility. The code, scripts, and instructions required to reproduce the submitted results are available at: https://github.com/khyati-patni/Bike_Challenge_2026

1. Introduction

Natural products remain a major source of pharmacologically relevant compounds, with approximately 67% of approved drugs derived directly or indirectly from natural sources. However, decades of biodiversity and natural product chemistry research remain locked in unstructured scientific articles, while many species and their undocumented chemistry face extinction. Manual curation of compound names, bioactivities, collection species, collection sites, and isolation types cannot scale with the publication rate. The BiKE 2026 / NatUKE task therefore provides a timely benchmark for evaluating automated knowledge extraction and link prediction methods over natural product literature.

The NatUKE benchmark [1] defines a link-prediction task over a heterogeneous knowledge graph derived from biochemical papers in the NuBBE database, encompassing over 2,500 natural product records. Given a paper DOI as a head entity, a model must rank candidate property values (tail entities) for five relation types. The challenge follows an official evaluation protocol at the 3rd stage with a 60%/40% training/test split, measuring how well methods generalise when labelled data is moderately abundant.

BiKE'26: Third International Biochemical Knowledge Extraction Challenge, May 10–May 14, 2026, co-located with the 23rd European Semantic Web Conference (ESWC), Dubrovnik, Croatia

[†]These authors contributed equally to this work.

khyati.patni@microcrispr.com (K. Patni[†]); daksh.sammi@microcrispr.com (D. Sammi[†]); rahul.kharkwal@microcrispr.com (R. Kharkwal); diksha.kashyap@microcrispr.com (D. Kashyap)

0009-0000-9708-2505 (K. Patni[†]); 0009-0007-7780-5674 (D. Sammi[†]); 0009-0002-3649-2548 (R. Kharkwal); 0009-0009-4211-9636 (D. Kashyap)

The original NatUKE benchmark compares DeepWalk, Node2Vec, Metapath2Vec, and EPHEN. While the first three are topology-driven graph embedding baselines, EPHEN is a propagation-based baseline that combines initial BERT embeddings with graph structure through iterative regularisation. BiKE v2 asks a natural follow-up question: how much do domain-specific BERT representations of paper text improve over structure-only baselines when used to initialise graph embeddings?

Our contributions are:

1. A fully reproducible pipeline that retrieves paper title and abstract metadata via CrossRef API, encodes them with SciBERT [2] and PubMedBERT, and constructs a five-relation heterogeneous KG over 791 nodes for the evaluated 3rd-stage split, derived from the broader NatUKE/BiKE dataset of 2,521 natural product records.
2. Comprehensive evaluation of 11 configurations—two BERT cosine baselines, two topology-only walk baselines, one spectral heat-kernel baseline, five SciBERT-initialised graph models, and one ensemble—under the official 3rd-stage NatUKE protocol. Seven configurations employ BERT-derived representations.
3. Systematic comparison of structure-only baselines versus BERT-derived models, showing that the best CareWeavers configuration exceeds the published NatUKE reference on 4 of 5 official 3rd-stage property metrics. We further analyse performance as a function of property label-space structure, showing that BERT-initialised graph models are most effective for small and structured label spaces such as isolation type, while sparse and highly variable labels such as compound names remain better served by typed walk or cosine-based methods.
4. Analysis of RotatE's relational rotation modelling on structured biochemical property relations, with strong results on isolation type (H@1 = 0.931).

2. Data and Knowledge Graph

2.1. Source Data

The broader NatUKE/BiKE challenge data contains 2,521 natural product records. For the officially evaluated 3rd stage split, the challenge provides train/test CSV splits for five NuBBE property types at the 60%/40% ratio. Each row is a triple (doi, property_value, relation_type), linking the 142 DOI nodes in our constructed graph to five NuBBE property types: compound name, bioactivity, collection species, collection site, and isolation type.

For each DOI, we query the CrossRef API and construct a paper text field from the available title and abstract. When the abstract is unavailable, the title is retained if present; if neither title nor abstract is available, the DOI string itself is used as fallback text. No chemical structures, molecular descriptors, fingerprints, or topic models are used at any stage of the pipeline. Unlike topic-model-based reconnection strategies such as BERTopic-style clustering, our method directly connects papers and property labels through shared SciBERT/PubMedBERT embedding spaces and graph-based link prediction.

2.2. Knowledge Graph Construction

We build a single heterogeneous KG by pooling all five tasks. Table 1 summarises the graph.

Node types are `doi` (papers) and `label` (property values). Each edge connects a paper to a property value via one of the five relation types. The graph has a two-hop diameter for any pair of papers sharing at least one property value, enabling neighbourhood-aggregation models to propagate semantic information across topically related papers.

The five property types differ substantially in label-space size, which strongly affects ranking difficulty. Isolation type has only five distinct labels, making it closer to a constrained classification problem, whereas compound name spans 445 distinct labels and is therefore much sparser. This asymmetry helps explain the divergent behaviour of RotatE, GraphSAGE, cosine baselines, and Metapath2Vec across the five tasks.

Statistic	Value
Total nodes	791
DOI nodes (papers)	142
Property-value nodes (labels)	649
Relation types	5
Total triples	1,104
bioActivity triples (train/test)	137 / 64
collectionSite triples (train/test)	96 / 58
collectionSpecie triples (train/test)	100 / 56
collectionType triples (train/test)	91 / 57
name triples (train/test)	338 / 107

Table 1
Knowledge graph statistics (3rd stage: 60%/40% split).

3. Approach

3.1. Text Encoding

Both encoders use mean-pooling of the final hidden layer (max length 512 tokens, batch size 16):

- **SciBERT** [2] (`allenai/scibert_scivocab_uncased`): trained on 1.14M papers from Semantic Scholar across broad scientific domains; 768-d output.
- **PubMedBERT** [11] (`microsoft/BiomedNLP-BiomedBERT-base-uncased-abstract-fulltext`): pre-trained from scratch on PubMed abstracts and full text; 768-d output, closely domain-matched to the biochemistry papers in NatUKE.

Both DOI nodes (paper embeddings) and label nodes (property-value embeddings) are encoded in the same 768-d space. Cosine similarity in this shared space is used directly as the two cosine-only baselines.

PubMedBERT is particularly useful for property types whose values appear verbatim or near-verbatim in biomedical abstracts, such as collection species. This motivates evaluating both direct cosine retrieval and graph-based refinement over the same text-derived embedding space.

3.2. Graph Models

All graph models are trained from scratch using the 60% training edges at the 3rd stage. A frequency-prior calibration penalty $\lambda \log(\text{freq} + \varepsilon)$ ($\lambda = 0.3$) is applied at ranking time to de-bias over-represented property values.

3.2.1. Topology-only baselines

These match the NatUKE baseline setup for topology-driven graph embedding: embeddings are learned purely from graph connectivity, without any BERT initialisation.

- **Node2Vec** [3]: biased random walks ($p = 1.0, q = 0.5$, length 80, 10 walks/node) trained with SkipGram (5 epochs, window 10, dim 128). Scored by cosine similarity between walk embeddings.
- **Metapath2Vec** [4]: typed walks on the heterogeneous graph along DOI→Label→DOI and Label→DOI→Label metapaths (length 100, 5 walks/node, neg. sampling, dim 128). Scored by dot product.

3.2.2. Spectral Heat-Kernel baseline

As an additional CareWeavers non-BERT structural baseline, we compute a normalised graph Laplacian $L = I - D^{-1/2}AD^{-1/2}$, derive a heat-kernel spectral representation $K = e^{-tL}$ with $t = 5$, use the 128 smallest eigenvectors, L_2-normalise the resulting 256-dimensional embeddings, and score DOI–label pairs by cosine similarity.

3.2.3. NatUKE propagation baseline

The NatUKE benchmark includes a text-seeded embedding propagation method that combines textual embeddings with graph structure, distinct from purely topology-driven methods.

- **EPHEN** [12]: Following the NatUKE benchmark, EPHEN is an embedding-propagation baseline on a heterogeneous knowledge graph. NatUKE describes EPHEN as a regularisation-based method that distributes initial BERT embeddings over the graph, combining textual node representations with graph structure in an unsupervised setting. Node states are initialised from available seed embeddings, updated from degree-normalised neighbour information, and re-anchored to the original embedding at each iteration through a mixing parameter. We treat EPHEN as a text-seeded propagation baseline rather than a topology-only embedding method.

3.2.4. BERT-initialised models (SciBERT node features)

SciBERT 768-d embeddings are used to initialise node representations; the exact projection mechanism differs per model as detailed below. Relation embeddings are randomly initialised and trained.

- **RotatE** [7]: each relation is a rotation in complex space, $\text{score}(h, r, t) = -\|h \circ r - t\|$. Node features are projected to phase space via a fixed random projection (768→256 d). Self-adversarial negative sampling; 400 epochs, dim 256, margin 6.0, batch 256.
- **R-GCN** [6]: two-layer relational graph convolution with per-relation weight matrices, bilinear scoring $h^\top W_r t$. Node features are projected from 768 d to 256 d via a learnable linear layer trained jointly with the graph model. 300 epochs, dropout 0.2, 5 negatives/positive.
- **TransE** [10]: translational model $h + r \approx t$, L2 margin-ranking loss. Node features are projected to 256 d via a fixed random projection. 400 epochs, dim 256, margin 2.0, batch 256.
- **ComplEx** [8]: complex bilinear model, $\text{score}(h, r, t) = \text{Re}(\langle e_h, w_r, \overline{e}_t \rangle)$. Node features are projected via a fixed random projection to 512 d (two 256-d halves representing the real and imaginary parts). Softplus BCE + Adagrad (L2 reg. 10^{-3}); 400 epochs, complex dim 256, batch 256.
- **GraphSAGE** [5]: inductive two-layer mean aggregator, L2-normalised layer outputs, bilinear per-relation scoring $h^\top W_r t$. Raw 768-d SciBERT embeddings are fed directly as node features; the first aggregation layer concatenates each node's self-embedding with its sampled neighbourhood mean (input size $2 \times 768 = 1536$) and projects to 256 d via a learnable linear layer. Max-margin loss; 300 epochs, 15 sampled neighbours, margin 2.

3.2.5. Ensemble

Scores from RotatE, R-GCN, TransE, ComplEx, and GraphSAGE are min-max normalised then averaged. The best available cosine score (PubMedBERT preferred) is added to the sum before normalisation.

4. Evaluation

Following NatUKE [1] exactly for the 3rd stage:

- **Split:** 60%/40% training/test ratio.

- **DOI-level splitting:** a DOI is either fully in train or fully in test across all five tasks simultaneously. This DOI-level split prevents information leakage because all labels associated with a paper are assigned consistently to either training or testing across the five property types.
- **Metrics:** Hits@k for $k \in \{1, 5, 20, 50\}$ and MRR.
- **Primary k:** Hits@50 for C and S; Hits@5 for B; Hits@20 for L; Hits@1 for T.
- **Model training:** All trainable graph models are trained from scratch using only the 3rd-stage training edges; cosine baselines need no retraining.

Table 2 reports the official 3rd-stage Hits@primary-k results for all NatUKE reference baselines and all CareWeavers configurations.

Method	C (H@50)	B (H@5)	L (H@20)	S (H@50)	T (H@1)
NatUKE Reference Baselines					
DeepWalk	0.000	0.109	0.431	0.143	0.175
Node2Vec	0.000	0.078	0.328	0.161	0.053
Metapath2Vec	0.093	0.109	0.448	0.429	0.263
EPHEN	0.037	0.656	0.550	0.179	0.754
CareWeavers Cosine Baselines					
Cosine SciBERT	0.400	0.090	0.310	0.680	0.000
Cosine PubMedBERT	0.390	0.080	0.300	**0.778**	0.000
CareWeavers Topology-Only Models					
Node2Vec	0.320	0.060	0.250	0.650	0.000
Metapath2Vec	**0.406**	0.090	0.300	0.670	0.000
CareWeavers Spectral Propagation Baseline					
Spectral Heat-Kernel	0.280	0.080	0.270	0.710	0.000
CareWeavers BERT-Initialised KG Embedding Models					
R-GCN	0.090	0.190	0.440	0.620	0.000
RotatE	0.170	**0.523**	0.641	0.620	**0.931**
TransE	0.350	0.080	0.300	0.680	0.000
ComplEx	0.260	0.080	0.300	0.730	0.000
GraphSAGE	0.310	0.470	**0.688**	0.600	0.070
CareWeavers Ensemble					
Normalised Ensemble	0.350	0.060	0.280	0.730	0.020

Table 2
Official 3rd-stage NatUKE results under the 60%/40% split. C = compound name, B = bioactivity, L = collection site, S = species, and T = isolation type. Primary metrics follow the NatUKE protocol: H@50 for C and S, H@5 for B, H@20 for L, and H@1 for T. Bold indicates the best score per property among CareWeavers configurations.

Code and reproducibility. The code, scripts, and instructions required to reproduce the submitted results are available at: https://github.com/khyati-patni/Bike_Challenge_2026

5. Related Work

The present work is related to three lines of research: natural product knowledge extraction benchmarks, graph representation learning for link prediction, and domain-specific language models for scientific and biomedical text.

NatUKE provides the benchmark setting for this study by formulating natural product knowledge extraction from academic literature as a link-prediction task over a heterogeneous knowledge graph [1]. In this setup, paper DOI nodes are linked to property-value nodes through relation types corresponding to NuBBE properties such as compound name, bioactivity, collection species, collection site, and isolation type. The NatUKE reference baselines include topology-driven graph embedding methods

such as DeepWalk, Node2Vec, and Metapath2Vec, as well as EPHEN, a propagation-based baseline that combines initial BERT embeddings with graph structure through iterative regularisation. Our work follows the same benchmark framing and official 3rd-stage evaluation protocol, but specifically studies whether scientific and biomedical text encoders can improve or complement structure-only graph representations.

Random-walk-based graph embedding methods form an important baseline family for this task. DeepWalk introduced the idea of learning node representations by treating truncated random walks as sentence-like sequences [9]. Node2Vec extends this family through biased random walks that control local versus more exploratory neighbourhood traversal [3]. Metapath2Vec adapts random-walk representation learning to heterogeneous networks by using metapaths over typed nodes and edges [4]. These methods are relevant to NatUKE because the graph is heterogeneous and bipartite, with DOI nodes connected to property-value label nodes. In our evaluation, Node2Vec and Metapath2Vec serve as structure-only baselines that learn from graph topology without access to paper text.

Knowledge graph embedding models provide a second family of related approaches. TransE represents relations as translations between head and tail entity embeddings [10]. ComplEx uses complex-valued embeddings for link prediction and can model both symmetric and antisymmetric relations [8]. RotatE represents each relation as a rotation in complex space and was designed for link prediction in knowledge graphs [7]. R-GCN extends graph convolution to multi-relational data by using relation-specific transformations, making it suitable for graphs with multiple edge types [6]. GraphSAGE learns inductive node representations by sampling and aggregating local neighbourhood features [5]. These methods motivate our BERT-initialised graph models, where SciBERT embeddings provide initial node features and graph training refines them for relation-specific prediction.

Scientific and biomedical language models are also central to this work. SciBERT is a BERT-based model pretrained on a large corpus of scientific publications and is designed for scientific NLP tasks [2]. PubMedBERT is pretrained from scratch on biomedical text, including PubMed abstracts and PubMed Central full-text articles, and was introduced to study domain-specific language model pretraining for biomedical NLP [11]. In our pipeline, these models encode both paper metadata and property-value labels into a shared semantic space. This enables direct cosine retrieval as well as BERT-initialised graph embedding, allowing us to compare text-only, structure-only, and text-initialised graph-based approaches under the same NatUKE evaluation protocol.

Overall, prior work provides either topology-based graph embeddings, multi-relational knowledge graph models, or domain-specific text encoders. The contribution of this paper is to combine these strands in a controlled NatUKE evaluation: paper and label nodes are initialised with scientific or biomedical BERT representations, then evaluated against structure-only graph baselines and NatUKE reference results across five natural product property types.

6. Discussion

6.1. Isolation Type (T)

RotatE achieves Hits@1 = 0.931, compared to the NatUKE EPHEN reference of 0.754. The strong performance may be attributed to RotatE's relational rotation scoring in complex space, which is well-suited to the small and highly structured label space of isolation types (few distinct values) compared to spectral or walk-based approaches. With only five distinct isolation-type labels, the ranking task is highly constrained. RotatE is well suited to this setting because its relation-specific rotations can map paper embeddings toward a small set of categorical label embeddings, effectively separating the possible isolation types in complex embedding space.

6.2. Species (S)

PubMedBERT cosine similarity achieves Hits@50 = 0.778, compared to the NatUKE Metapath2Vec baseline at 0.429 on the 3rd stage. The result suggests that species names are distinctive enough in

abstract text that semantic similarity alone, without any graph training, retrieves the correct answer within the top 50 candidates most of the time. The strong PubMedBERT cosine result suggests that collection species are often recoverable directly from text. Species names are usually distinctive Latinised binomials that occur in titles or abstracts, so graph training adds less value than semantic retrieval in this case.

6.3. Collection Site (L)

GraphSAGE leads at the 3rd stage with Hits@20 = 0.688, exceeding the NatUKE EPHEN reference of 0.550. RotatE achieves Hits@20 = 0.641 on this property. The neighbourhood aggregation in GraphSAGE benefits from the moderate connectivity in the 60% training graph, allowing it to effectively leverage relational structure for geographic disambiguation. Collection site benefits from neighbourhood aggregation because geographic labels often co-occur with related species and bioactivity labels in the same papers. GraphSAGE can exploit this multi-hop context by propagating information across papers that share property-value neighbours.

6.4. Bioactivity (B)

RotatE achieves Hits@5 = 0.523 on bioactivity, compared to the NatUKE EPHEN reference of 0.656. The EPHEN baseline's advantage suggests that regularisation-based embedding propagation can better transfer BERT-seeded information across the heterogeneous graph for bioactivity than pure relational rotation alone. Bioactivity labels are highly diverse and multi-valued, making this the most challenging property in the benchmark. Bioactivity remains difficult because labels are multi-valued and semantically overlapping. A single paper may report multiple biological activities, and these labels are not always lexically distinct in the abstract text. This explains why EPHEN's text-seeded graph regularisation remains competitive and why future work should consider multi-label ranking formulations.

6.5. Compound Name (C)

This is the hardest property due to the large and sparse label space. Metapath2Vec achieves Hits@50 = 0.406, compared to the NatUKE reference of 0.093. The improvement suggests that typed random walks over the pooled heterogeneous graph capture useful paper–property co-occurrence patterns, which may be especially helpful when compound-name labels are sparse and highly variable. Compound name is sparse because many labels are rare and highly specific. Typed metapath walks help by capturing paper-property co-occurrence patterns, whereas BERT-initialised graph models may struggle when compound names are too specialised or absent from the available title/abstract text.

6.6. Ensemble Underperformance

The ensemble underperforms the individual best models on bioactivity and collection site. This is because the frequency-calibration parameter $\lambda = 0.3$ and the ensemble weights are fixed globally, not tuned per property. A model-selection strategy or per-property λ tuning would likely improve ensemble performance.

7. Conclusion

This work evaluates BERT-initialised knowledge graph embeddings on the NatUKE natural product extraction benchmark. We present a comprehensive evaluation of 11 configurations combining cosine baselines, structure-only graph methods, and BERT-initialised graph models. RotatE achieves Hits@1 = 0.931 on isolation type (EPHEN reference: 0.754) and Hits@5 = 0.523 on bioactivity (EPHEN reference: 0.656). PubMedBERT cosine similarity achieves Hits@50 = 0.778 on species (Metapath2Vec reference: 0.429), and GraphSAGE achieves Hits@20 = 0.688 on collection site (EPHEN reference: 0.550). Metapath2Vec achieves Hits@50 = 0.406 on compound name (reference: 0.093).

Under the official 3rd-stage 60%/40% setting, the best configuration per property type exceeds the published NatUKE reference on 4 of 5 property types, while bioactivity remains the main unresolved challenge. Per-property model selection appears more effective than ensemble approaches for this benchmark. Future work includes section-aware abstract encoding, multi-hop neighbourhood reasoning, domain-specific bioactivity models, and per-property hyperparameter optimisation.

Overall, the results indicate that no single model dominates all property types. Per-property model selection is more appropriate than a fixed ensemble: RotatE is strongest for structured categorical labels, PubMedBERT cosine retrieval is strongest for textually distinctive species labels, GraphSAGE is strongest for collection-site neighbourhood signals, and Metapath2Vec remains useful for sparse compound-name prediction.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: Grammar and spelling check. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] Paulo Viviurka do Carmo, Edgard Marx, Ricardo Marcacini, Marilia Valli, João Victor Silva e Silva, and Alan Pilon. NatUKE: A benchmark for natural product knowledge extraction from academic literature. In *17th IEEE International Conference on Semantic Computing (ICSC)*. IEEE, 2023.

[2] Iz Beltagy, Kyle Lo, and Arman Cohan. SciBERT: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676*, 2019.

[3] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2016.

[4] Yuxiao Dong, Nitesh V. Chawla, and Ananthram Swami. metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2017.

[5] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[6] Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *Proceedings of the European Semantic Web Conference (ESWC)*, 2018.

[7] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. RotatE: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197*, 2019.

[8] Theo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, 2016.

[9] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. DeepWalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2014.

[10] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems 26 (NeurIPS)*, 2013.

[11] Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare*, 2021.

[12] Paulo do Carmo and Ricardo Marcacini. Embedding propagation over heterogeneous event net-

works for link prediction. In *2021 IEEE International Conference on Big Data (Big Data)*, pp. 4812–4821, 2021. DOI: 10.1109/BigData52589.2021.9671645.