# GENE: Global Event Network Embedding

**Qi Zeng[1], Manling Li[1], Tuan Lai[1], Heng Ji[1], Mohit Bansal[2], Hanghang Tong[1]**
[1]University of Illinois Urbana-Champaign [2]UNC Chapel Hill
{qizeng2, manling2, tuanml2, hengji, htong}@illinois.edu
mbansal@cs.unc.edu

## Abstract

Current methods for event representation ignore related events in a corpus-level global context. For a deep and comprehensive understanding of complex events, we introduce a new task, Event Network Embedding, which aims to represent events by capturing the connections among events. We propose a novel framework, Global Event Network Embedding (GENE), that encodes the event network with a multi-view graph encoder while preserving the graph topology and node semantics. The graph encoder is trained by minimizing both structural and semantic losses. We develop a new series of structured probing tasks, and show that our approach effectively outperforms baseline models on node typing, argument role classification, and event coreference resolution. [1]

## 1 Introduction

Understanding events is a fundamental human activity. Our minds represent events at various granularity and abstraction levels, which allows us to quickly access and reason about related scenarios. A typical event mention includes an *event trigger* (the word or phrase that most clearly expresses an event occurrence) and its *arguments* (i.e., participants in events). The lexical embedding of a trigger is usually not sufficient, because the type of an event often depends on its arguments (Ritter and Rosen, 2000; Xu and Huang, 2013; Weber et al., 2018). For example, the support verb "get" may indicate a Transfer.Ownership event ("Ellison to spend $10.3 billion to **get** his company.") or a Movement.Transport event ("Airlines are **getting** flyers to destinations on time more often."). In Figure 1, the event type triggered by "execution" is Life.Die instead of project implementation. However, such kind of atomic event representation is

still overly simplistic since it only captures local information and ignores related events in the global context. Real-world events are inter-connected, as illustrated in the example in Figure 1. To have a comprehensive representation of the set fire event on an embassy, we need to incorporate its causes (e.g., the preceding execution event) and recent relevant events (e.g., the protests that happened before and after it). To capture these inter-event relations in a global context, we propose the following two assumptions.

**Assumption 1. Two events can be connected through the entities involved.** On schema or type level, two event types can be connected through multiple paths and form a coherent story (Li et al., 2020). This observation is also valid on instance level. For the example in Figure 1, one of the relations between the Set Fire event and the Execution event is the blue path ⟨*Set Fire, target, Saudi Embassy, affiliation, Saudi Arabia, agent, Execution*⟩, which partially supports the fact that angry protesters revenge the death of Nimr al-Nimr against Saudi Arabia by attacking its embassy. This approximation for event-event relations lessens the problems of coarse classification granularity and low inter-annotation agreement (which may be as low as 20% as reported in (Hong et al., 2016)). Hence, we propose to construct an **Event Network**, where each event node represents a unique instance labeled with its type, arguments, and attributes. These nodes are connected through multiple instantiated meta-paths (Sun et al., 2011) consisting of their entity arguments and the entity-entity relations. These entities can be co-referential (e.g., two protests on different dates that both occur in Tehran, Iran) or involved in the same semantic relations (both protests targeted the Saudi embassy, which is affiliated with the location entity "Saudi Arabia").

**Assumption 2. The representation of one event depends on its neighboring events in the**

---

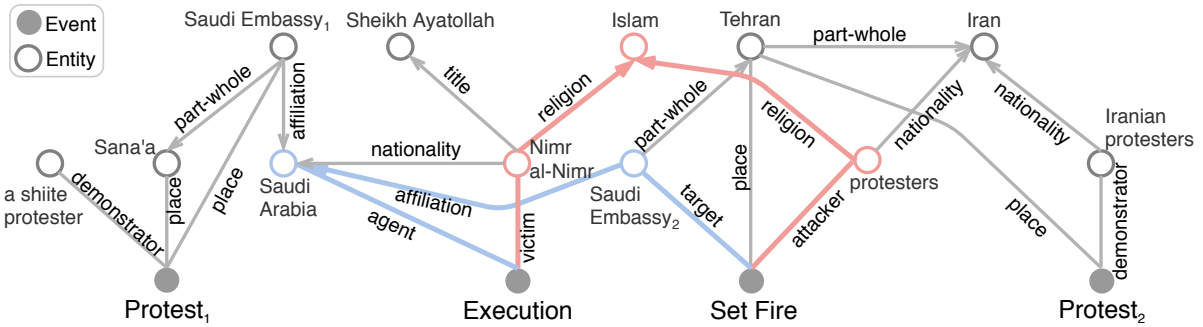[1]Our code is released at https://github.com/pkuzengqi/GENE

Figure 1: An example of Event Network constructed from one VOA news article, where events are connected through entities involved. Each node is an event or entity and each edge represents an argument role or entity-entity relation. In this example, *Execution* event and *Set Fire* event are connected through two paths, which tell the story of angry protesters revenge the death of Nimral-Nimr against Saudi Arabia by attacking its embassy.

**event network.** In Figure 1, a good representation of the Set Fire event should involve the Execution event because the latter clarifies the grievance motivating the former. We further enrich event representations by introducing more context from the entire event network. Compared with other methods to connect events (e.g., with event-event relations (Pustejovsky et al., 2003; Cassidy et al., 2014; Hong et al., 2016; Ikuta et al., 2014; O'Gorman et al., 2016)), our representation of each event grounded in an event network is semantically richer.

Based on these two hypotheses, we introduce a new task of **Event Network Embedding**, aiming at representing events with low-dimensional and informative embeddings by incorporating neighboring events. We also propose a novel **Global Event Network Embedding Learning (GENE)** framework for this task. To capture network topology and preserve node attributes in the event representations, GENE trains a graph encoder by minimizing both structural and semantic losses. To promote relational message passing with focus on different parts of the graph, we propose an innovative multi-view graph encoding method.

We design **Event Network Structural Probes**, an evaluation framework including a series of structural probing tasks, to check the model's capability to implicitly incorporate event network structures. In this work, the learned node embeddings are intrinsically evaluated with node typing and event argument role classification tasks, and applied to the downstream task of event coreference resolution. Experimental results on the augmented Automatic Content Extraction (ACE) dataset show that leveraging global context can significantly enrich

the event representations. GENE and its variants significantly outperform the baseline methods on various tasks.

In summary, our contributions are:

- We formalize the task of event network embedding and accordingly propose a novel unsupervised learning framework, which trains the multi-view graph encoder with topology and semantics learning losses.
- We design a series of incrementally structural probing tasks, including node typing, argument role classification, and event coreference resolution, to comprehensively evaluate the event network embedding models.
- We demonstrate that our event network embedding is effective and general enough to enhance downstream applications.

## 2   Problem Formulation

**Event Network.** An event network with $n$ nodes is a heterogeneous attributed network denoted as $G = \{V, E\}$, where $V$ and $E$ are node and edge sets, respectively. Each node $v_i = \langle a_i, b_i, s_i, l_i \rangle \in V$ represents an event or entity mention, where $a_i$ and $b_i$ are the start and end word indices in sentence $s_i$, and $l_i$ is the node type label. Each edge $e_{ij} = \langle i, j, l_{ij} \rangle \in E$ represents an event-entity or entity-entity relation, where $i$ and $j$ are indices of the involved nodes and $l_{ij}$ is the edge type label.

In this work, we initialize the semantic representation of each node $v_i$ with an $m$-dimensional attribute vector $x_i$ derived from sentence context using a pretrained BERT model (Devlin et al., 2019).

**Semantic Proximity (Gao and Huang, 2018).** Given an event network $G = \{V, E\}$, the semantic proximity of node $v_i$ and node $v_j$ is determined by

43

the similarity of node attribute vectors $x_i$ and $x_j$. If two nodes are semantically similar in the original space, they should stay similar in the new space.

**Local Neighborhood.** Given $G = \{V, E\}$, the local (one-hop) neighborhood $\mathcal{N}_i$ of node $v_i$ is defined as $\mathcal{N}_i = \{v_j \in V \mid e_{ij} \in E\}$. For example, the local neighborhood of one event is composed of its argument entities. Given event-entity node pairs, the task of argument role classification is to label the local neighborhood of events.

**Global Neighborhood.** Given $G = \{V, E\}$, node $v_j$ belongs to the global ($k$-hop with $k \geq 2$) neighborhood of node $v_i$, if node $v_i$ can walk to node $v_j$ in $k$ hops. For example, two events are 3-hop neighbors when there is a path from one event to the other through two entity nodes.

**Event Network Embedding.** Given an event network $G = \{V, E\}$ with $n$ nodes, the task of event network embedding aims to learn a mapping function $f : \{V, E\} \to Y$ or $f : \mathbb{R}^{n \times m} \times \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times d}$, where $Y = [y_i] \in \mathbb{R}^{n \times d}$ is the node representation, $d$ is the embedding dimension, and $Y$ should preserve the Semantic Proximity, Local Neighborhood and Global Neighborhood.

## 3 Model

### 3.1 Approach Overview

Compared to other network embedding tasks, there are three challenges in event network embedding:

- Data Sparsity: We rely on supervised Information Extraction (IE) techniques to construct the event network, because they provide high-quality knowledge elements. However, due to the limited number of types in pre-defined ontologies, the constructed event network tends to be sparse.
- Relational Structure: The event network is heterogeneous with edges representing relations of different types. Relation types differ in semantics and will influence message passing.
- Long-Distance Dependency: Global neighborhood preservation requires node embedding to capture the distant relations between two nodes.

We first initiate the event network by event and entity extraction, event argument role labeling and entity-entity relation extraction. The nodes in the event network are events and entities. If entity coreference resolution results are available, we merge coreferential entity mentions and label the

mention text with the first occurring mention. For each node $v_i$, we derive its $m$-dimensional attribute vector $x_i$ with its mention text by averaging the corresponding contextual token embeddings from a pretrained bert-base model. The edges in the event network come from the event argument roles connecting event mentions and entities, and the entity-entity relations. In addition, to alleviate the data sparsity problem we enrich the event network with external Wikipedia entity-entity relations and event narrative orders as a data preprocessing step detailed in Section 5.1.

We propose an unsupervised Global Event Network Embedding (GENE) learning framework for this task (Figure 2). We first encode the graph with a Relational Graph Convolutional Network (RGCN) (Schlichtkrull et al., 2018) based multi-view graph encoder, in which the multi-view component puts focus on various perspectives of the graph. To capture both the semantic and topological contexts, i.e. the node attributes and graph structure, in event node representation, GENE trains the graph encoder by minimizing semantic reconstruction loss and relation discrimination loss.

### 3.2 Multi-View Graph Encoder

Given an event network $G = \{V, E\}$, the graph encoder projects the nodes into a set of embeddings $Y$ while preserving the graph structure and node attributes. As shown in Figure 2, we first feed different views of $G$ to the graph encoder, then integrate encoded node embeddings into $Y$.

**RGCN.** Because of the relational structure of event network, we apply RGCN (Schlichtkrull et al., 2018), a relational variant of GCN (Kipf and Welling, 2017), as the graph encoder. RGCN induces the node embeddings based on the local neighborhood with operations on a heterogeneous graph. It differs from GCN in the type-specific weights in message propagation.

We stack two RGCN layers in the encoder. The hidden state of node $v_i$ in the first layer is initiated with node attribute $x_i$. The output of the former layer serves as the input of the next layer. Formally, in each RGCN layer the hidden state $h$ of node $v_i$ is updated through message propagation with the hidden states of neighbors (and itself) from the last layer and message aggregation with an addition
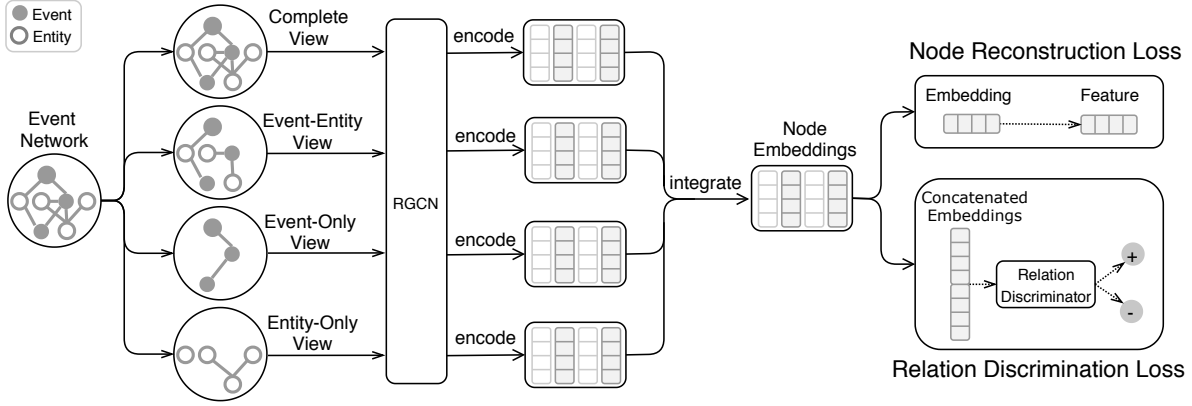
Figure 2: An overview of the proposed GENE framework. The event network is encoded by a relational graph convolutional network, which is trained with node reconstruction loss and relation discrimination loss.

operation and an element-wise activation function.

$$h_i^{(l+1)} = \sigma(W_0^{(l)} h_i^{(l)} + \sum_{r \in R} \sum_{j \in \mathcal{N}_i^r} \frac{1}{c_{i,r}} W_r^{(l)} h_j^{(l)}),$$

where $h_i^{(l)} \in \mathbb{R}^{d^{(l)}}$ is the hidden state of node $v_i$ at the $l$-th layer of RGCN, $d^{(l)}$ is the dimension of the hidden state at the $l$-th layer, $R$ is the edge relation set, $\mathcal{N}_i^r$ is the neighborhood of node $v_i$ under relation type $r \in R$, $W_r^{(l)}$ is the trainable weight matrix of relation type $r$ at the $l$-th layer, $c_{i,r} = |\mathcal{N}_i^r|$ is a normalization constant, and $\sigma$ is Leaky ReLU.

**Weight Decomposition.** In order to reduce the growing model parameter size and prevent the accompanying over-fitting problem, we follow (Schlichtkrull et al., 2018) and perform basis decomposition on relation weight matrix:

$$W_r^{(l)} = \sum_{b=1}^{B} a_{rb}^{(l)} V_b^{(l)},$$

where the edge weight $W_r^{(l)}$ is a linear combination of basis transformations $V_b^{(l)} \in \mathbb{R}^{d^{(l+1)} \times d^{(l)}}$ with coefficients $a_{rb}$. This basis decomposition method reduces model parameters by using a much smaller base set $B$ to compose relation set $R$ and can be seen as a way of weight sharing between different relation types.

**Multiple Views.** The structure of event networks can be viewed in multiple different perspectives. For example, when entity-entity relations are masked out, an event network degenerates to pieces of isolated events and only local neighborhood will be observed. The advantage of separate modeling

is that it enables the graph encoder to focus on different perspectives of the graph and lessens the over-smoothing problem (the tendency of indistinguishable encoded node embeddings). Therefore, we propose to encode the network $G = \{V, E\}$ from the following views:

(1) Complete View: We keep all nodes and all edges in this view.

(2) Event-Entity View: We keep all nodes and only event-entity relations in this view. Events are isolated as single subgraphs, each of which only includes the corresponding event and its argument entities.

(3) Entity-Only View: We only keep entity nodes and entity-entity relations in this view. Information is flowed only among entity nodes and will not be influenced by events.

(4) Event-Only View: We only keep event nodes and event-event relations in this view. Similarly, events are isolated from entities.

We feed the event network in different views as separate inputs to the graph encoder, and integrate the encoded results in three ways:

**Concatenation.** Node embeddings of $\frac{d}{v}$ dimensions from $v$ views are directly concatenated with $y_{cat} = [y^0 \cdot y^1 \cdots y^{v-1}]$.

**Averaging.** Node embeddings of $d$ dimensions from $v$ views are averaged with $y_{avg} = \frac{1}{v} \sum_{j=1}^{v} y^j$.

**Weighted Averaging.** Node embeddings of $d$ dimensions from $v$ views are averaged with $y_{wavg} = \frac{1}{v} \sum_{j=1}^{v} W_v^j y^j$, where $W_v^j$ is a trainable matrix.

### 3.3 Topology Learning

To capture neighborhood information, we train the graph encoder with relation discrimination loss to

45

learn the graph topology.

$$\mathcal{L}_T = \sum_i (\sum_{r \in R} \sum_{j \in \mathcal{N}_i^r} \mathbb{E}[\log \mathcal{D}_r(y_i, y_j)]$$
$$+ \sum_{r \in R} \sum_{j' \notin \mathcal{N}_i^r} \mathbb{E}[\log(1 - \mathcal{D}_r(y_i, y_{j'}))])$$

The relation-specific discriminator $\mathcal{D}_r$ determines the probability score for one node's being connected with another node in relation $r$:

$$\mathcal{D}_r(y_i, y_j) = \sigma(y_i^T W_{\mathcal{D}}^r y_j)$$

where $W_{\mathcal{D}}^r$ is a trainable bi-linear scoring matrix and $\sigma$ is Sigmoid funtion. We choose binary discriminator over multi-class classifier to capture features required for independent classification decisions.

### 3.4 Semantics Learning

To preserve the node semantics, we perform node attribute reconstruction with a two-layer feed-forward neural network:

$$\mathcal{L}_S = \sum_i \|x_i - \phi(y_i)\|_2$$

where $x_i$ represents the attributes of node $v_i$, $y_i$ represents the encoded embedding of node $v_i$, and $\phi : \mathbb{R}^{n \times d} \to \mathbb{R}^{n \times m}$ denotes the non-linear transformation function. $\mathcal{L}_S$ loss evaluates how much information required to reconstruct node attributes is preserved in the encoded node embeddings.

### 3.5 Training

To encourage the graph encoder to learn both the graph topology and node semantics, we combine the structural loss and semantics loss as the final objective function:

$$\mathcal{L} = \mathcal{L}_T + \lambda \mathcal{L}_S$$

where $\lambda$ is a weight normalization hyper-parameter.

## 4 Structural Probes for Event Network

As there is no existing work on comprehensive event representation evaluation, in this work we design an evaluation framework with a series of probing tasks to comprehensively evaluate the model's capability to capture network structures and preserve node attributes. Structural Probes are models trained to predict certain properties from inferred representations, and have been used to understand

linguistic properties (Hewitt and Manning, 2019; Conneau et al., 2018).

The task of event network embedding requires the embedded distributional node representations to preserve semantic proximity, local neighborhood and global neighborhood. Accordingly, we intrinsically evaluate the semantics preservation with node typing and assess the local neighborhood preservation with event argument role classification. We also apply the node embeddings to a downstream task, event coreference resolution, to extrinsically evaluate the global neighborhood preservation.

Node Typing and Event Argument Role Classification are conducted under the same evaluation setting: given the learned node embeddings, predict the labels with a multi-layer perceptron (MLP) based classifier. If the input of the classifier is of different dimension to the event network embeddings, it will be first projected into the same dimension. The classifier is a two-layer feed-forward neural network with a linear transformation layer, a non-linear activation operation, a layer normalization, a dropout operation, and another linear transformation layer. The classifier is designed to be simple on purpose so that it will be limited in reasoning ability and thus the evidence for classification will be mainly derived from the node embeddings.

### 4.1 Node Typing

The event or entity type of each node can be inferred from the sentence context of its mentions. As the node attribute vector $x_i$ for node $v_i$ comes from the contextual word embeddings, $x_i$ naturally implies its node type. This characteristic is supposed to be preserved after the node has been further embedded and the embedding dimension has been reduced.

We evaluate the node semantics preservation by checking whether the node types can be recovered from the node embeddings. Given one event or entity node, our evaluation model predicts its type out of 45 labels, which includes 7 coarse-grained entity types, 5 value types, and 33 event types as defined in the NIST Automatic Content Extraction (ACE) task. The performance on this task is compared in terms of multi-label classification Micro F1 score.

### 4.2 Event Argument Role Classification

We detect local neighborhood preservation by evaluating whether the event-entity relation (event argument role) can be recovered from the node embeddings. Given one event node and one entity

node, we predict the relation type between each pair of nodes out of 238 labels. Each label consists of an event type and an argument role type as defined in ACE. For example, the argument role label "Justice:Arrest-Jail:Agent" can only be correctly selected when the event node implies the type "Justice:Arrest-Jail" and the entity node implies its role being the "Agent". Compared to the traditional argument role labeling procedure, this setting skips the step of mention identification, which has been done in network construction process. The performance is reported with multi-label classification Micro F1 score.

### 4.3 Event Coreference Resolution

The goal of event coreference resolution is to determine which event mentions refer to the same real-world event. The features for similarity computation used in previous work are typically limited to event triggers, arguments and sentence-level contexts (Chen et al., 2009; Chen and Ji, 2009; Sammons et al., 2015; Lu and Ng, 2016; Chen and Ng, 2016; Duncan et al., 2017; Lai et al., 2021). However, event arguments are often distributed across the content of an article. Therefore a global event network can ground event mentions into a wider context with related events and help cluster coreferential mentions more accurately.

In this task we evaluate the impact of applying event network embedding as additional features on enhancing event coreference resolution. We concatenate the event embeddings learned by the event network and by a fine-tuned SpanBERT model (Joshi et al., 2020) as the input for the scoring function. The training procedure is the same as that in (Joshi et al., 2019).

We report F1 scores in terms of B$^3$ (Bagga and Baldwin, 1998), MUC (Vilain et al., 1995), CEAF$_e$ (Luo, 2005), BLANC (Recasens and Hovy, 2011) metrics, and also their averaged results (AVG).

## 5 Results and Analysis

### 5.1 Dataset

We construct corpus-level graphs for training, development, and test sets from the English subset of Automatic Content Extraction (ACE) 2005 dataset[2]. We follow the pre-processing steps in (Lin et al., 2020) and show the dataset statistics in Table 1.

We perform automatic entity linking (Pan et al., 2017) to link entities to Wikipedia. Entity nodes linked to the same Wikipedia entity are merged into one node. We further retrieve entity-entity relations from Wikidata and enrich the event network with these connections, such as the part-whole relation between Tehran and Iran in Figure 1. We also add narrative event-event relations by connecting every pair of events within one document as edges in the graph.

### 5.2 Baseline

**Non-Graph Event Representation Methods.** Mention-based method represents events with contextual representations inferred by BERT (Devlin et al., 2019). Tuple-based method uses the averaged contextual representations of event mentions and its arguments.

**Graph Representation Methods.** Skipgram (Mikolov et al., 2013) learns graph topology by increasing the predicted similarity of adjacent node embeddings and decreasing the similarity of irrelevant node embeddings with random negative sampling:

$$\mathcal{L}_G = \sum_i (\sum_{j \in \mathcal{N}_i} \log \sigma(y_j^T y_i) + \sum_{j' \notin \mathcal{N}_i} \log \sigma(-y_{j'}^T y_i))$$

Deep Graph Infomax (Velickovic et al., 2019) captures graph topology by maximizing the mutual information between patch representations and higher-level subgraph summary:

$$\mathcal{L}_D = \sum_i (\sum_{j \in \mathcal{N}_i} \mathbb{E}[\log \mathcal{D}(y_i, s)] + \sum_{j' \notin \mathcal{N}_i} \mathbb{E}[\log(1 - \mathcal{D}(y_{j'}, s))])$$

where the subgraph summary $s$ is read out as the average of node embeddings and $\mathcal{D}$ is the discriminator deciding the probability score for node's being contained in the summary.

For fair comparison, we train the same framework with the following graph representation learning methods.

**Event Coreference Resolution.** Besides existing methods (Bejan and Harabagiu, 2010b; Liu et al., 2014) we implement the model architecture (Lee et al., 2017) that has achieved the current state-of-the-art results in entity coreference resolution (Joshi et al., 2019) and cross-document event

| | | Article | Node | | Edge | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Event | Entity | Event-Entity | Entity-Entity | | Event-Event | |
| | | | | | | Original | Wiki* | Narrative* | Coref |
| ACE | train | 521 | 4,353 | 3,688 | 7,888 | 6,856 | 7,040 | 70,992 | 912 |
| | dev | 30 | 494 | 667 | 938 | 723 | 853 | 12,572 | 144 |
| | test | 40 | 424 | 750 | 897 | 796 | 1,543 | 6,154 | 121 |

Table 1: Statistics for the enhanced ACE 2005 dataset. *Wiki* and *Narrative* are enriched event-event relations.

coreference resolution (Cattan et al., 2020). We use SpanBERT (Joshi et al., 2020) for contextual embeddings. The detailed methods about the baseline event corefernece resolution framework are described in (Lai et al., 2021). In this experiment, we compare the performance with and without our event network embeddings as additional features.

### 5.3 Training Details

All models are implemented with Deep Graph Library and Pytorch framework. We train each models for 10 epochs and apply an early stopping strategy with a patience of 3 epochs (if the model does not outperform its best checkpoint for 3 epochs on validation set we will stop the training process). The batch size is 64.

The hyper-parameters are selected based on model performance on development set. The model is optimized with the Adam optimizer with a learning rate of $1e-5$ and a dropout rate of $0.1$. The embedding dimension is $256$ and the hidden dimension is $512$. The lambda in loss function is $1.0$. On average it takes approximately four hours to train a model until converge with one Tesla V100 GPU with 16GB DRAM. To improve training efficiency, neighbor pre-sampling is performed for all topology learning losses.

### 5.4 Results and Analysis

We conclude the results shown in Table 2 with the following observations:

**GENE preserves node semantics well with low-dimensional and informative embeddings.** Though with only one third of embedding dimension (typically $256$, comparing to $768$ in other event representation baselines), our models have higher performance on Node Typing, which shows the node semantics has been well preserved.

**Topology learning loss is crucial to event neighborhood proximity preservation.** We propose to use relation discrimination loss to learn the graph structure and exam it with argument role classification task. Methods without topology learning objectives (Event as Mention, Event as Tuple, and

GENE w/ $\mathcal{L}_T$) have a significant drop of performance on this task, while our proposed model has the best performance because of the similarity and transferability between argument role classification and argument role discrimination in $\mathcal{L}_T$. Another reason is that only $\mathcal{L}_T$ is designed for heterogeneous graphs while SKG and DGI do not consider relation types.

**In general Multi-view encoder is beneficial.** Compared to the single-view variants, our multi-view encoder has overall better performance. Keeping complete view has the most closed performance, while discarding event-entity relations yields significant drop on argument role classification.

**Averaging multi-view embeddings is better than Weighted Averaging.** Intuitively weighted averaging captures the correlations among different embedding dimensions, promotes salient dimensions and/or teases out unimportant ones within the same view by performing a linear transformation within each view before averaging over views. However, results show that it is not comparable with averaging and concatenation multi-view encoders. One possible reason is that the distribution of embedding within each view is greatly restricted by the input embedding distribution.

**GENE improves the performance on event coreference resolution by connecting events through related entities.** SpanBERT model is a strong baseline with better performance compared with the former methods. We show that using our embeddings as additional features, SpanBERT can further improve all event coreference resolution scores. In the following example, SpanBERT model fails to detect the coreference link between event *sell* and event *buy* while GENE succeeds by discovering the relation between the entity arguments.

> ... The Times said Vivendi Universal was negotiating to **sell** its flagship theme parks to New York investment firm Blackstone Group as a first step toward dismantlingits entertainment empire . Vivendi Universal officials in the United States were not immediately available for comment on Friday . Under the reported plans , Blackstone Group would **buy** Vivendi ' s theme park division , including ...

| Model | Node Typing | Argument Classification | Event Coreference | | | | |
|---|---|---|---|---|---|---|---|
| | | | MUC | $B^3$ | $CEAF_e$ | BLANC | AVG |
| Event Mention | 80.58 | 71.57 | 61.81 | 87.79 | 84.24 | 74.97 | 77.20 |
| Event Tuple | 68.40 | 72.13 | 63.10 | 89.06 | 85.23 | 77.6 | 78.75 |
| Skip-gram(Mikolov et al., 2013) | 75.55 | 93.42 | 59.81 | 88.09 | 83.40 | 77.30 | 77.15 |
| Deep Graph Infomax(Velickovic et al., 2019) | 74.96 | 95.32 | 59.36 | 87.05 | 82.19 | 73.41 | 75.50 |
| HDP(Bejan and Harabagiu, 2010b) | - | - | - | 83.8 | 76.7 | - | - |
| (Liu et al., 2014) | - | - | 50.98 | 89.38 | **86.47** | 70.43 | 74.32 |
| SpanBERT(Joshi et al., 2020) | - | - | 65.72 | 89.48 | 85.35 | 79.82 | 80.09 |
| **GENE** | **81.26** | **95.76** | 68.99 | 89.53 | 85.86 | 80.38 | 81.19 |
| · w/o $\mathcal{L}_T$ | 78.78 | 79.04 | **70.63** | 89.03 | 84.88 | **81.13** | 81.42 |
| · w/o $\mathcal{L}_S$ | 78.02 | 95.32 | 68.14 | 89.53 | 86.17 | 79.90 | 80.94 |
| · w/ Event-Entity view | 80.82 | 92.64 | 60.09 | 87.97 | 84.75 | 70.67 | 75.87 |
| · w/ Event-only & Entity-only views | 74.79 | 72.02 | 60.86 | 88.46 | 85.15 | 75.64 | 77.53 |
| · w/ Complete view | 79.42 | 90.52 | 63.01 | 88.11 | 84.94 | 75.50 | 77.89 |
| · w/ Concatenated integration | 78.45 | 93.87 | 70.08 | **89.81** | 85.85 | 81.08 | **81.71** |
| · w/ Weighted integration | 74.53 | 94.31 | 66.36 | 88.99 | 85.81 | 76.97 | 79.53 |

Table 2: Results on test set of ACE dataset. Node typing and argument role classification results are reported in micro F1 scores(%). Event Coreference are performed with our embeddings as additional features.

**Remaining Challenges.** One of the unsolved challenges is to capture the long distance relation in the encoder in addition to the two encoder layers. Another challenge is the limited ability in entity coreference resolution. In some failing cases, GENE model does not link two events because some of their connecting arguments are expressed as pronouns. This limitation is inherited from the upstream event extraction.

# 6 Related Work

**Event Representation.** Some previous efforts enrich event representations by introducing arguments (Levin, 1993; Goldberg, 1995; Ritter and Rosen, 2000; Huang and Ahrens, 2000; Iwata, 2005; Goldberg, 2006; Xu and Huang, 2013; Bies et al., 2016; Do et al., 2017; Kalm et al., 2019), intent and sentiment (Ding et al., 2019), and temporal information (Tong et al., 2008). (Weber et al., 2018) proposes a tensor-based event composition approach to combine a trigger and arguments to represent each event. We extend the definition of scenario to multiple inter-connected events. (Modi, 2016) captures statistical dependencies between events but limits to script data sets where the events are naturally organized in sequential temporal order. Our approach captures a rich variety of explicit semantic connections among complex events. (Hong et al., 2018) learns distributed event representations using supervised multi-task learning, while our framework is based on unsupervised learning.

**Network Embedding.** Our work falls into the scope of unsupervised learning for heterogeneous attributed network embeddings. Heterogeneous network embedding methods (Chang et al., 2015; Dong et al., 2017; Wang et al., 2019) jointly model nodes and edges. Attributed network embedding approaches (Gao and Huang, 2018; Yang et al., 2015) on the other hand put focus on preserving node attributes when encoding the networks.

**Event Coreference Resolution.** Most existing methods (Chen et al., 2009; Chen and Ji, 2009; Bejan and Harabagiu, 2010a; Zhang et al., 2015; Peng et al., 2016; Lai et al., 2021) only exploit local features including trigger, argument and sentence context matching. To prevent error propagation, some models perform joint inference between event extraction and event coreference resolution (Lee et al., 2012; Araki and Mitamura, 2015; Lu and Ng, 2017) or incorporate document topic structures (Choubey and Huang, 2018). To the best of our knowledge our method is the first to leverage the entire event networks to compute similarity features.

# 7 Conclusions and Future Work

We propose a novel continuous event representation called Event Network Embedding to capture the connections among events in a global context. This new representation provides a powerful framework for downstream applications such as event coreference resolution and event ordering.

In the future we aim to improve the ability to capture the long-distance relations in the graph encode by introducing event-event relation in the form of multiple meta-paths. The relations, or the event evolution patterns, extracted from large-scale corpora can guide event-related reasoning and act as shortcut linking event nodes. Another direction is to explore a unified automatic evaluation benchmark for event representation.

## Acknowledgement

## References

Jun Araki and Teruko Mitamura. 2015. Joint event trigger identification and event coreference resolution with structured perceptron. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 2074–2080, Lisbon, Portugal. Association for Computational Linguistics.

Amit Bagga and Breck Baldwin. 1998. Algorithms for scoring coreference chains. In The first international conference on language resources and evaluation workshop on linguistics coreference, volume 1, pages 563–566. Citeseer.

Cosmin Bejan and Sanda Harabagiu. 2010a. Unsupervised event coreference resolution with rich linguistic features. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 1412–1422, Uppsala, Sweden. Association for Computational Linguistics.

Cosmin Adrian Bejan and Sanda M. Harabagiu. 2010b. Unsupervised event coreference resolution with rich linguistic features. In ACL 2010, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, July 11-16, 2010, Uppsala, Sweden, pages 1412–1422. The Association for Computer Linguistics.

Ann Bies, Zhiyi Song, Jeremy Getman, Joe Ellis, Justin Mott, Stephanie Strassel, Martha Palmer, Teruko Mitamura, Marjorie Freedman, Heng Ji, and Tim O'Gorman. 2016. A comparison of event representations in DEFT. In Proceedings of the Fourth Workshop on Events, pages 27–36, San Diego, California. Association for Computational Linguistics.

Taylor Cassidy, Bill McDowell, Nathanael Chambers, and Steven Bethard. 2014. An annotation framework for dense event ordering. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 501–506, Baltimore, Maryland. Association for Computational Linguistics.

Arie Cattan, Alon Eirew, Gabriel Stanovsky, Mandar Joshi, and Ido Dagan. 2020. Streamlining cross-document coreference resolution: Evaluation and modeling. CoRR, abs/2009.11032.

Shiyu Chang, Wei Han, Jiliang Tang, Guo-Jun Qi, Charu C. Aggarwal, and Thomas S. Huang. 2015. Heterogeneous network embedding via deep architectures. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, August 10-13, 2015, pages 119–128.

Chen Chen and Vincent Ng. 2016. Joint inference over a lightly supervised information extraction pipeline: Towards event coreference resolution for resource-scarce languages. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI'16, page 2913–2920. AAAI Press.

Zheng Chen and Heng Ji. 2009. Graph-based event coreference resolution. In Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing (TextGraphs-4), pages 54–57, Suntec, Singapore. Association for Computational Linguistics.

Zheng Chen, Heng Ji, and Robert Haralick. 2009. A pairwise event coreference model, feature impact and evaluation for event coreference resolution. In Proceedings of the Workshop on Events in Emerging Text Types, pages 17–22, Borovets, Bulgaria. Association for Computational Linguistics.

Prafulla Kumar Choubey and Ruihong Huang. 2018. Improving event coreference resolution by modeling correlations between event coreference chains and document topic structures. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 485–495, Melbourne, Australia. Association for Computational Linguistics.

Alexis Conneau, German Kruszewski, Guillaume Lample, Loïc Barrault, and Marco Baroni. 2018. What you can cram into a single $&!#* vector: Probing sentence embeddings for linguistic properties. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 2126–2136, Melbourne, Australia. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Xiao Ding, Kuo Liao, Ting Liu, Zhongyang Li, and Junwen Duan. 2019. Event representation learning enhanced with external commonsense knowledge. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 4894–4903, Hong Kong, China. Association for Computational Linguistics.

Quynh Ngoc Thi Do, Steven Bethard, and Marie-Francine Moens. 2017. Improving implicit semantic role labeling by predicting semantic frame arguments. In Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 90–99, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Yuxiao Dong, Nitesh V. Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, August 13 - 17, 2017, pages 135–144.

Chase Duncan, Liang-Wei Chan, Haoruo Peng, Hao Wu, Shyam Upadhyay, Nitish Gupta, Chen-Tse Tsai, Mark Sammons, and Dan Roth. 2017. UI CCG TAC-KBP2017 submissions: Entity discovery and linking, and event nugget detection and co-reference. In Proceedings of the 2017 Text Analysis Conference, TAC 2017, Gaithersburg, Maryland, USA, November 13-14, 2017. NIST.

Hongchang Gao and Heng Huang. 2018. Deep attributed network embedding. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden, pages 3364–3370.

Adele E. Goldberg. 1995. Constructions: A Construction Grammar Approach to Argument Structure. Chicago: University of Chicago Press.

Adele E. Goldberg. 2006. Constructions at Work: the Nature of Generalization in Language. Oxford: Oxford University Press.

John Hewitt and Christopher D. Manning. 2019. A structural probe for finding syntax in word representations. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4129–4138, Minneapolis, Minnesota. Association for Computational Linguistics.

Xudong Hong, Asad Sayeed, and Vera Demberg. 2018. Learning distributed event representations with a multi-task approach. In Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, pages 11–21, New Orleans, Louisiana. Association for Computational Linguistics.

Yu Hong, Tongtao Zhang, Tim O'Gorman, Sharone Horowit-Hendler, Heng Ji, and Martha Palmer. 2016. Building a cross-document event-event relation corpus. In Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016), pages 1–6, Berlin, Germany. Association for Computational Linguistics.

Chu-Ren Huang and Kathleen Ahrens. 2000. The module-attribute representation of verbal semantics. In Proceedings of the 14th Pacific Asia Conference on Language, Information and Computation, pages 109–120, Waseda University International Conference Center, Tokyo, Japan. PACLIC 14 Organizing Committee.

Rei Ikuta, Will Styler, Mariah Hamang, Tim O'Gorman, and Martha Palmer. 2014. Challenges of adding causation to richer event descriptions. In Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation, pages 12–20, Baltimore, Maryland, USA. Association for Computational Linguistics.

Seizi Iwata. 2005. Locative alternation and two levels of verb meaning. Cognitive Linguistics, 16(2):355–407.

Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S. Weld, Luke Zettlemoyer, and Omer Levy. 2020. SpanBERT: Improving pre-training by representing and predicting spans. Transactions of the Association for Computational Linguistics, 8:64–77.

Mandar Joshi, Omer Levy, Luke Zettlemoyer, and Daniel Weld. 2019. BERT for coreference resolution: Baselines and analysis. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 5803–5808, Hong Kong, China. Association for Computational Linguistics.

Pavlina Kalm, Michael Regan, and William Croft. 2019. Event structure representation: Between verbs and argument structure constructions. In Proceedings of the First International Workshop on Designing Meaning Representations, pages 100–109, Florence, Italy. Association for Computational Linguistics.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings.

Tuan Lai, Heng Ji, Trung Bui, Quan Hung Tran, Franck Dernoncourt, and Walter Chang. 2021. A context-dependent gated module for incorporating symbolic semantics into event coreference resolution. In Proc. The 2021 Conference of the North American Chapter of the Association

for Computational Linguistics - Human Language Technologies (NAACL-HLT2021).

Heeyoung Lee, Marta Recasens, Angel Chang, Mihai Surdeanu, and Dan Jurafsky. 2012. Joint entity and event coreference resolution across documents. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pages 489–500, Jeju Island, Korea. Association for Computational Linguistics.

Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer. 2017. End-to-end neural coreference resolution. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 188–197, Copenhagen, Denmark. Association for Computational Linguistics.

Beth Levin. 1993. English Verb Classes and Alternations: a Preliminary Investigation. Chicago: University of Chicago Press.

Manling Li, Qi Zeng, Ying Lin, Kyunghyun Cho, Heng Ji, Jonathan May, Nathanael Chambers, and Clare Voss. 2020. Connecting the dots: Event graph schema induction with path language modeling. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics.

Ying Lin, Heng Ji, Fei Huang, and Lingfei Wu. 2020. A joint neural model for information extraction with global features. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 7999–8009, Online. Association for Computational Linguistics.

Zhengzhong Liu, Jun Araki, Eduard H. Hovy, and Teruko Mitamura. 2014. Supervised within-document event coreference using information propagation. In Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014, pages 4539–4544. European Language Resources Association (ELRA).

Jing Lu and Vincent Ng. 2016. Event coreference resolution with multi-pass sieves. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), pages 3996–4003, Portorož, Slovenia. European Language Resources Association (ELRA).

Jing Lu and Vincent Ng. 2017. Joint learning for event coreference resolution. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 90–101, Vancouver, Canada. Association for Computational Linguistics.

Xiaoqiang Luo. 2005. On coreference resolution performance metrics. In Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, pages 25–32.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings.

Ashutosh Modi. 2016. Event embeddings for semantic script modeling. In Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning, pages 75–83, Berlin, Germany. Association for Computational Linguistics.

Tim O'Gorman, Kristin Wright-Bettner, and Martha Palmer. 2016. Richer event description: Integrating event coreference with temporal, causal and bridging annotation. In Proceedings of the 2nd Workshop on Computing News Storylines (CNS 2016), pages 47–56, Austin, Texas. Association for Computational Linguistics.

Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji. 2017. Cross-lingual name tagging and linking for 282 languages. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1946–1958, Vancouver, Canada. Association for Computational Linguistics.

Haoruo Peng, Yangqiu Song, and Dan Roth. 2016. Event detection and co-reference with minimal supervision. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 392–402, Austin, Texas. Association for Computational Linguistics.

James Pustejovsky, Patrick Hanks, Roser Saurí, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, and Marcia Lazo. 2003. The TIMEBANK corpus. Corpus Linguistics, 2003:40.

Marta Recasens and Eduard Hovy. 2011. Blanc: Implementing the rand index for coreference evaluation. Natural Language Engineering, 17(4):485–510.

Elizabeth Ritter and Sara Thomas Rosen. 2000. Event structure and ergativity. In Events as Grammatical Objects, pages 187–238.

Mark Sammons, Haoruo Peng, Yangqiu Song, Shyam Upadhyay, Chen-Tse Tsai, Pavankumar Reddy, Subhro Roy, and Dan Roth. 2015. Illinois CCG TAC 2015 event nugget, entity discovery and linking, and slot filler validation systems. In Proceedings of the 2015 Text Analysis Conference, TAC 2015, Gaithersburg, Maryland, USA, November 16-17, 2015, 2015. NIST.

Michael Sejr Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings, pages 593–607.

Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S. Yu, and Tianyi Wu. 2011. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. Proc. VLDB Endow., 4(11):992–1003.

Hanghang Tong, Yasushi Sakurai, Tina Eliassi-Rad, and Christos Faloutsos. 2008. Fast mining of complex time-stamped events. In Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008, Napa Valley, California, USA, October 26-30, 2008, pages 759–768.

Petar Velickovic, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R. Devon Hjelm. 2019. Deep graph infomax. In 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019. OpenReview.net.

Marc B. Vilain, J. Burger, J. Aberdeen, D. Connolly, and L. Hirschman. 1995. A model-theoretic coreference scoring scheme. In MUC.

Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S. Yu. 2019. Heterogeneous graph attention network. In The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019, pages 2022–2032.

Noah Weber, Niranjan Balasubramanian, and Nathanael Chambers. 2018. Event representations with tensor-based compositions. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018, pages 4946–4953.

Hongzhi Xu and Chu-Ren Huang. 2013. Primitives of events and the semantic representation. In Proceedings of the 6th International Conference on Generative Approaches to the Lexicon (GL2013), pages 54–61, Pisa, Italy. Association for Computational Linguistics.

Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Y. Chang. 2015. Network representation learning with rich text information. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015, pages 2111–2117.

Tongtao Zhang, Hongzhi Li, Heng Ji, and Shih-Fu Chang. 2015. Cross-document event coreference resolution based on cross-media features. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 201–206, Lisbon, Portugal. Association for Computational Linguistics.