# Trans-X and Voting Strategy for Link Prediction

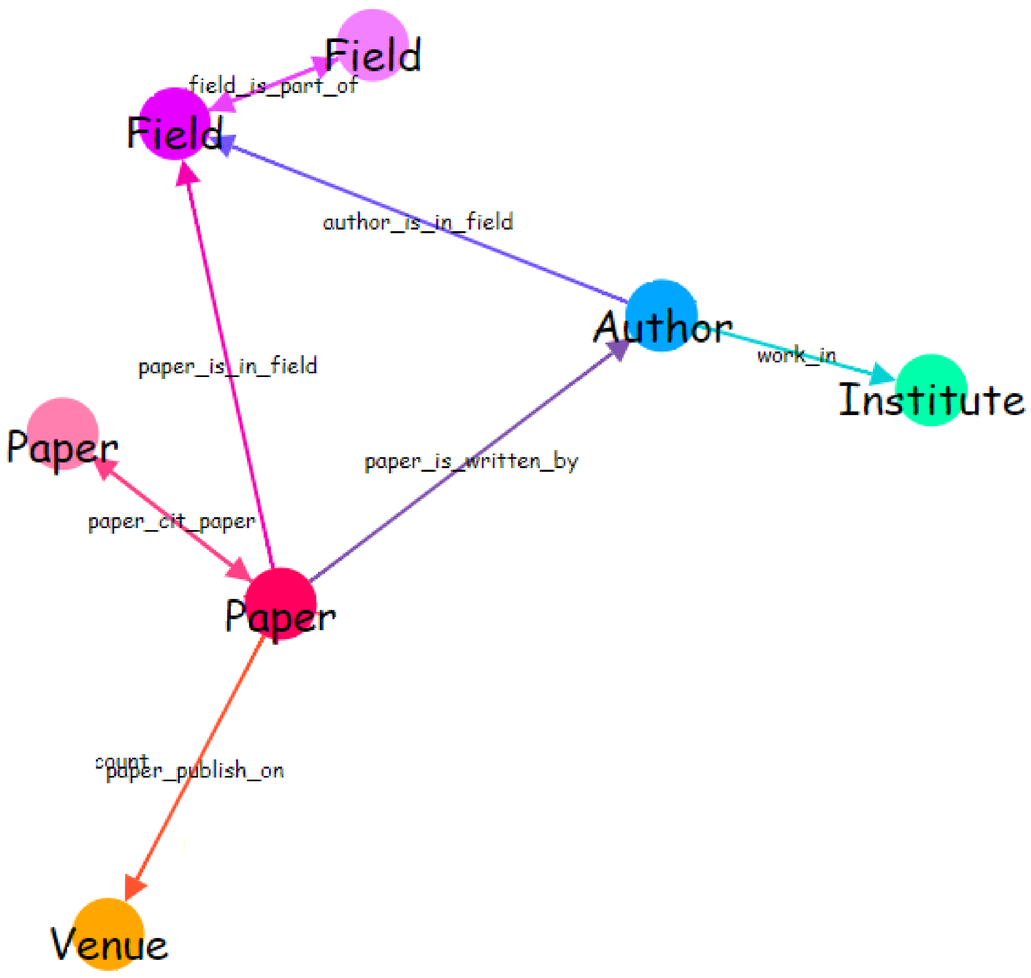
Caiyuan Li (caiyuanlee@foxmail.com)
EE-448, Shanghai Jiao Tong University

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

## Background

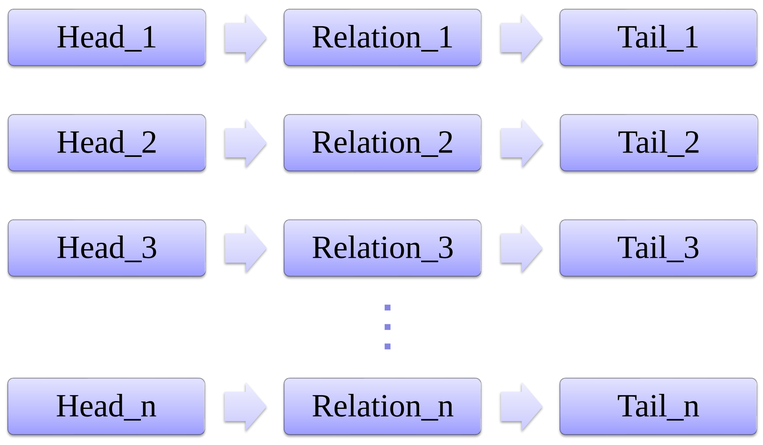
### Knowledge graphs

Knowledge graph is essentially a semantic network. It is a graph-based data structure composed of nodes and edges. In the Knowledge graph, each node represents the "entity" that exists in the real world, and each edge is the "relationship" between the entity and the entity. Knowledge graphs are the most effective representation of relationships. In layman's terms, Knowledge graph is a network of relations that connect all different kinds of information together.

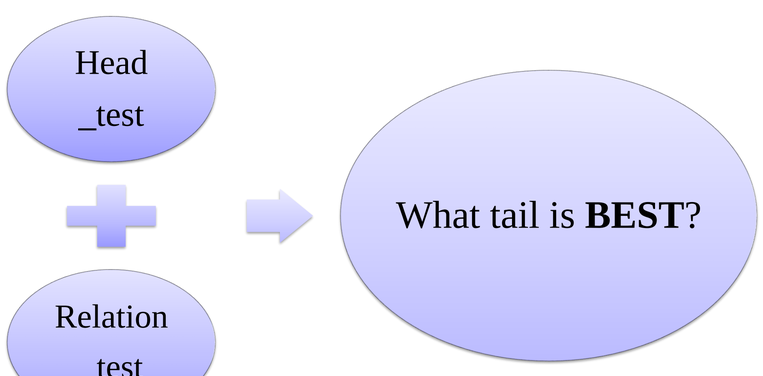



## Task

### train




### test

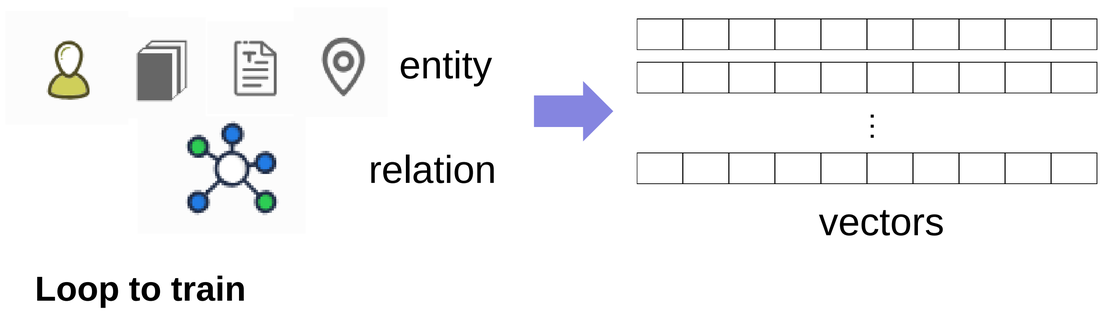



An instance of edge is a triplet of fact (head entity, relation, tail entity) (denoted as (h, r, t)). The training set provided in the course includes 146,917 triples. We need use it to infer missing links in an observed academic knowledge graph.

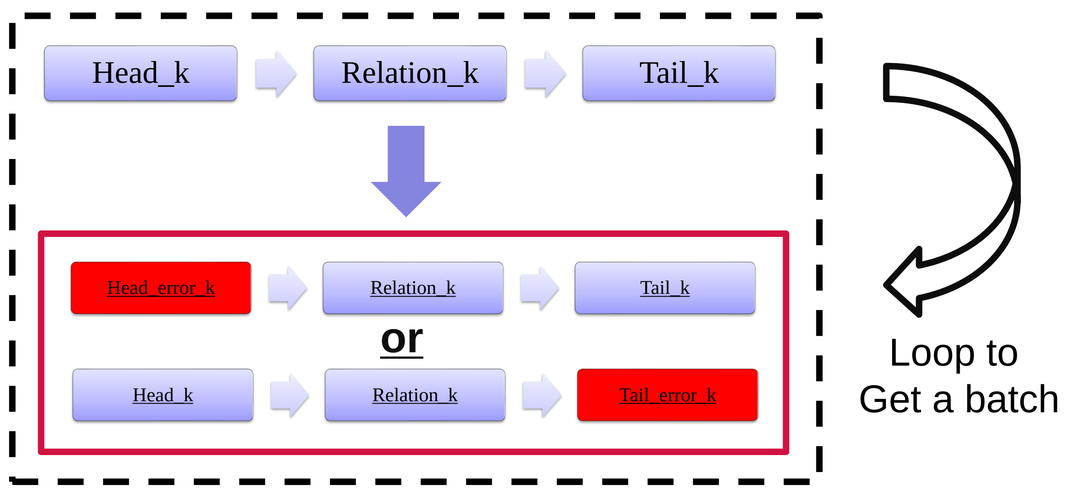
## My Approach

### Trans-E

#### Initialization




#### Loop to train




We define loss function as:

$$\mathcal{L} = \sum_{(h,\ell,t)\in S} \sum_{(h',\ell,t')\in S'_{(h,\ell,t)}} \left[\gamma + d(\boldsymbol{h}+\boldsymbol{\ell},\boldsymbol{t}) - d(\boldsymbol{h'}+\boldsymbol{\ell},\boldsymbol{t'})\right]_+$$

Where:

$$S'_{(h,\ell,t)} = \{(h',\ell,t)|h' \in E\} \cup \{(h,\ell,t')|t' \in E\}$$

Loop to update **loss function**

#### Voting Strategy

Strategy 1: Use n different results and get 3*n predicted tails for each testing head and relation. Find the 3 tails that appear most often and regard them as a new result.
Strategy 2: Add weights based on Strategy 1. The weight for each result is its own score. For example, if AAAAAAAA appears two times in it, its value is 2 using Strategy 1. However,if the score of these two results is 0.28 and 0.32, AAAAAAAA's value is 0.60. Experiments show that Strategy 2 is significantly better than Strategy 1 when the number of results used for voting is large.

## Experiments

| Result | score |
|---|---|
| TransE: dim.=100, marg.=1, L1-norm | 0.24 |
| TransH: dim.=100, marg.=1, L1-norm | 0.23 |
| TransR: dim.=100, marg.=1, L1-norm | 0.25 |
| TransE: dim.= 50 , marg.=1, L1-norm | 0.21 |
| TransE: dim.=100, marg.=2, L1-norm | 0.26 |
| TransE: dim.=200, marg.=2, L1-norm | 0.28 |
| Deep walk: dim.=128, number-walks=80 | 0.28 |
| Deep walk: dim.=200, number-walks=40 | 0.34 |
| Voting strategy 1 | 0.35 |
| Voting strategy 2 | 0.38 |