

Knowledge Graph Analysis

Exercise Sheet 8

Dr. Asja Fischer, Prof. Jens Lehmann

January 30, 2017

1 IN CLASS

1. Latent Distant Models

- What kind of distance measures (i.e. metrics) do you know? List at least two.
- Recall the Structured Embedding model

$$f_{ijk}^{\text{SE}} = -\|\mathbf{W}_k^s \mathbf{a}_i - \mathbf{W}_k^o \mathbf{a}_j\|_1 .$$

Explain the role of the parameters in the score-function. Do the same for the TransE model

$$f_{ijk}^{\text{TransE}} = -d(\mathbf{a}_i + \mathbf{r}_k, \mathbf{a}_j) .$$

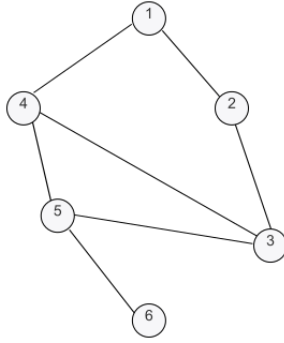
- Given $d(\cdot) = \|\cdot\|_2$ and $\|\mathbf{a}_i\| = \|\mathbf{a}_j\| = 1$, show that the score function of the TransE model can be written as:

$$f_{ijk}^{\text{TransE}} = -(2\mathbf{r}_k^T(\mathbf{a}_i - \mathbf{a}_j) - 2\mathbf{a}_i^T \mathbf{a}_j + \|\mathbf{r}_k\|_2^2) .$$

2. Graph theory

Look at the undirected graph given in Figure ??.

- What is the neighborhood of node 5?
- Give an example for a path between node 2 and node 6.



- c) Consider a random walk starting in node 1 and moving in each step to a neighbor of the current node chosen uniformly at random. What is the probability of the path $\{(1, 4), (4, 5), (5, 6)\}$?

3. Graph Feature Models

- a) Explain the idea behind the **Common neighbors** index

$$f_{ij} = |N(e_i) \cap N(e_j)| ,$$

the **Adamic/Adar** index

$$f_{ij} = \sum_{e_h \in N(e_i) \cap N(e_j)} \frac{1}{\log(|N(e_h)|)} ,$$

and the **Katz** index

$$f_{ij} = \sum_{l=1}^{\infty} \beta^l \cdot |\Pi_{e_i, e_j}^l| ,$$

where $\Pi_{e_i, e_j}^l = \{\text{paths of length } l \text{ from } e_i \text{ to } e_j\}$.

- b) Are these local or global similarity indices for uni-relational data?
 c) Recall the score function of the **path ranking algorithm (PRA)**

$$f_{ijk}^{\text{PRA}} = \mathbf{r}_k^T \mathbf{x}^{\text{PRA}} ,$$

with

$$\mathbf{x}_{ijk}^{\text{PRA}} = [P(\Pi_{e_i, e_j}(t)) : \Pi_{e_i, e_j}(t) \subset \Pi_{ijk}] .$$

Consider the following relation sequences/paths and the weights learned by the PRA for the prediction of the college a person attended (i.e. the existence of triples of the form (p, college, c):

- $t_1 = (\text{parents}, \text{education}, \text{institution}), r_{k1} = 1.37$
- $t_2 = (\text{placeOfBirth}, \text{peopleBornHere}, \text{education}), r_{k2} = 6.4$
- $t_3 = (\text{profession}, \text{peopleWithProfession}, \text{education}, \text{institution}), r_{k3} = 2.19$

To which logic rules (i.e. Horn clauses) do the paths correspond and which is the most predictive for the existence of the relation?