Robustness of RDF2Vec

Mike Laszkiewicz
Faculty of Mathematics, Ruhr-University Bochum

Motivation

- RDF2Vec aims to produce useful **word embeddings** for entities in a RDF graph.
- This technique builds on **random walks**.
- Are the embeddings **robust** on the choice of walks?

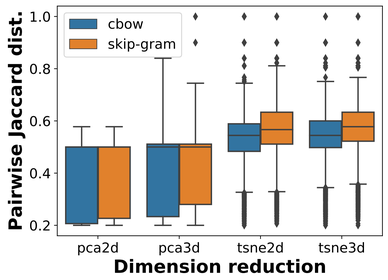
Experimental Setup

How to investigate the robustness?

1. **Jaccard distance** of the **neighborhoods**:
$$J(N_1, N_2) := \frac{N_1 \cap N_2}{N_1 \cup N_2} \in [0, 1].$$
2. **Predictive power** of the embeddings: 20-Nearest-Neighborhood **classification** of the entities type.

Results

Pairwise similarities: aifb

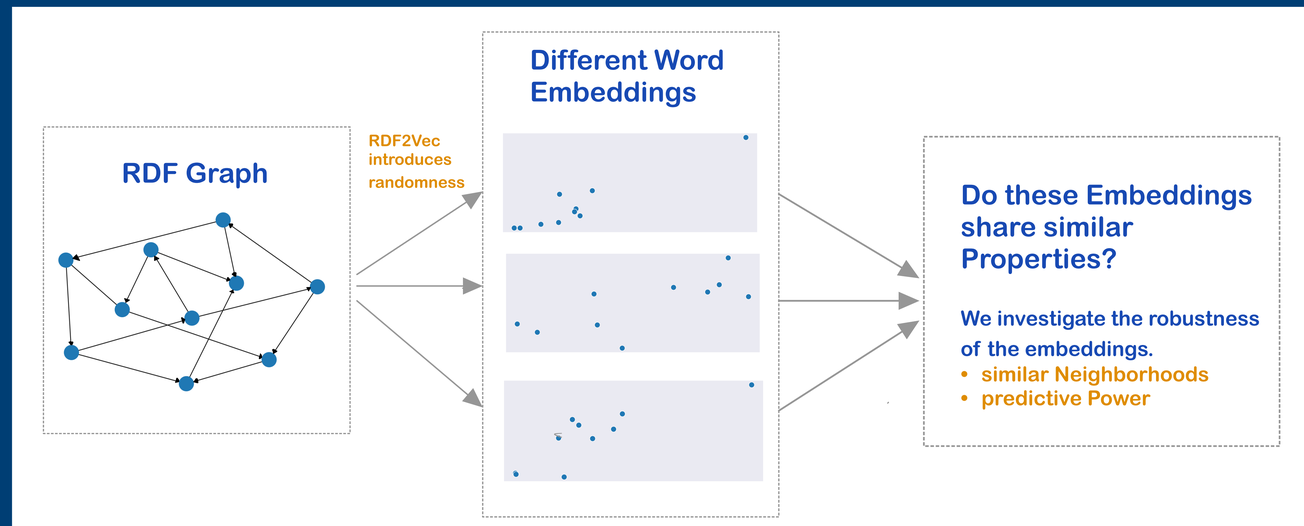


Future Extensions:

- Increase the **number of random walks** per entity;
- **weighted random walks**;
- Other **measures of similarity**.

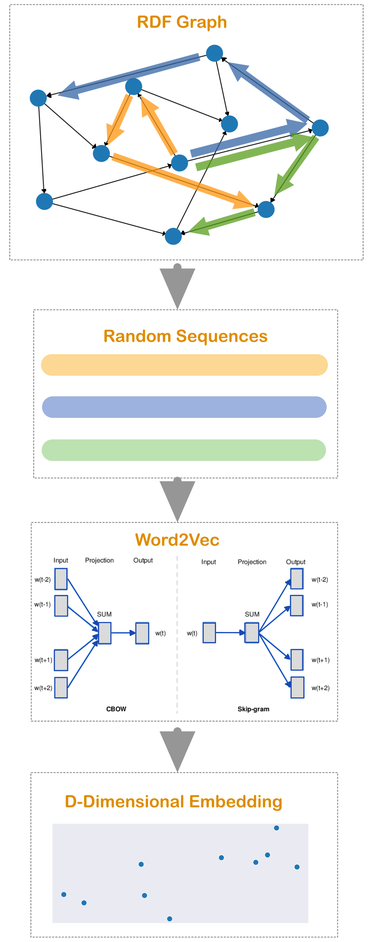
RUHR UNIVERSITÄT BOCHUM RUB

Randomness in **RDF2Vec** leads to **ambiguous** word embeddings. The method **fails** to produce **robust embeddings.**



Take a picture to get to the **github** repository, which includes **details**, **further results**, and the **code**.

RDF2Vec in a nut shell:



Second experiment:

Do all embeddings promote the same **prediction** results?

Similar predictions: aifb