

NSCaching: Simple and Efficient Negative Sampling for Knowledge Graph Embedding

Yongqi Zhang*, Quanming Yao^γ, Yingxia Shao[♦], Lei Chen*

*HKUST, ^γ4Paradigm, [♦]BUPT

*{yzhangee, leichen}@cse.ust.hk, ^γyaoquanming@4paradigm.com, [♦]shaoyx@bupt.edu.cn

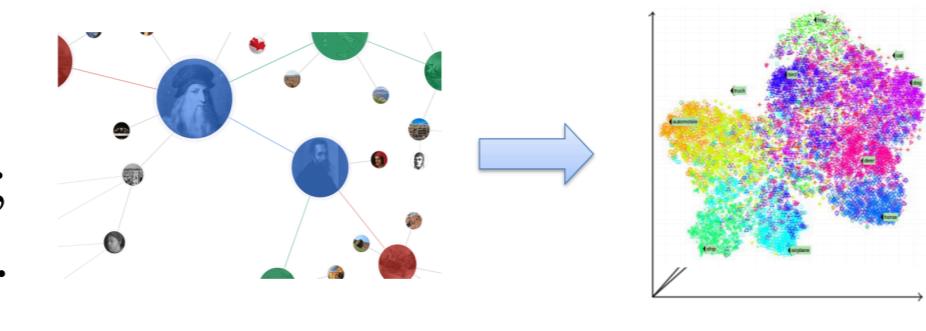
Code: <https://github.com/yzhangee/NSCaching>

KG Embedding

- KG is usually represented as triplets (*head, relation, tail*).
- We aim to learn low-dimensional vectors (embeddings) to represent each **entity** and **relation**.
- Objective:
 - Maximize score on positive triplets;
 - Minimize **score** on negative triplets.

$$\text{TransE: } f(h, r, t) = -\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_1$$

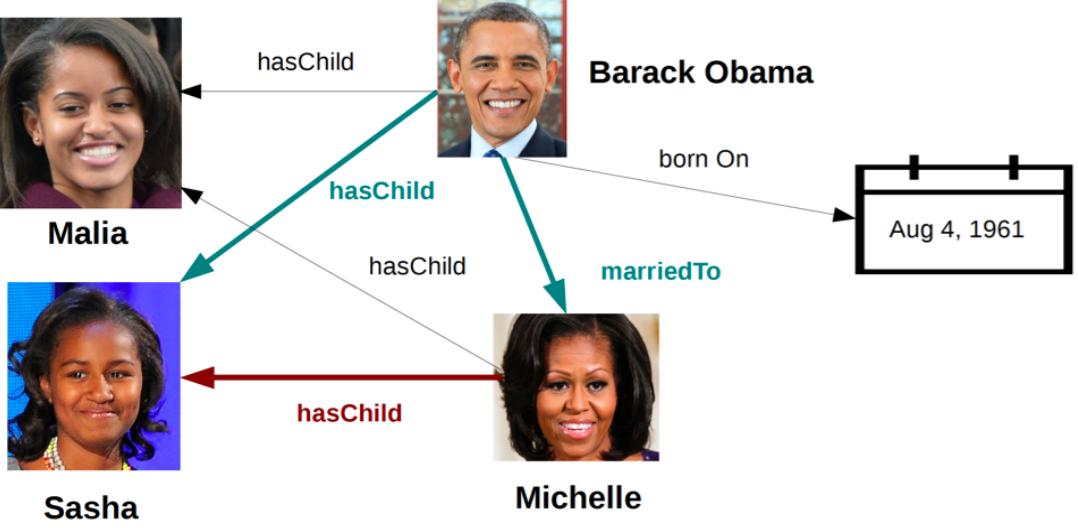
$$\text{DistMult: } f(h, r, t) = \langle \mathbf{h}, \mathbf{r}, \mathbf{t} \rangle$$



Algorithm 1 General framework of KG embedding.
Input: training set $\mathcal{S} = \{(h, r, t)\}$, embedding dimension d and scoring function f ;
 1: initialize the embeddings for each $e \in \mathcal{E}$ and $r \in \mathcal{R}$.
 2: **for** $i = 1, \dots, T$ **do**
 3: sample a mini-batch $\mathcal{S}_{\text{batch}} \subseteq \mathcal{S}$ of size m ;
 4: **for** each $(h, r, t) \in \mathcal{S}_{\text{batch}}$ **do**
 5: sample a negative triplet $(\bar{h}, r, \bar{t}) \in \bar{\mathcal{S}}_{(h, r, t)}$;
 // negative sampling
 6: update parameters of embeddings w.r.t. the gradients using (i). translational distance models:

$$\nabla [\gamma - f(h, r, t) + f(\bar{h}, r, \bar{t})]_+, \quad (3)$$
 or (ii). semantic matching models:

$$\nabla \ell (+1, f(h, r, t)) + \nabla \ell (-1, f(\bar{h}, r, \bar{t})) ; \quad (4)$$
 7: **end for**
 8: **end for**



Negative Sampling

- Only contains **observed** facts (**positive** triplets).
- Non-observed ones are assumed to be **negative**.
- Given a positive triplet (h, r, t) , negative triplet is sampled from
 $\bar{\mathcal{S}}_{(h, r, t)} = \{(\bar{h}, r, t) \notin \mathcal{S} | \bar{h} \in \mathcal{E}\} \cup \{(h, r, \bar{t}) \notin \mathcal{S} | \bar{t} \in \mathcal{E}\}$

Positive	Negative
(Obama, <i>marriedTo</i> , Michelle)	(Obama, <i>marriedTo</i> , Sasha), (Sasha, <i>marriedTo</i> , Michelle), (Obama, <i>bornOn</i> , Michelle)
(Michelle, <i>hasChild</i> , Malia)	(Michelle, <i>hasChild</i> , Obama), (Sasha, <i>hasChild</i> , Malia), (Michelle, <i>bornOn</i> , Malia)

- The **quality** of negative triplets matters!

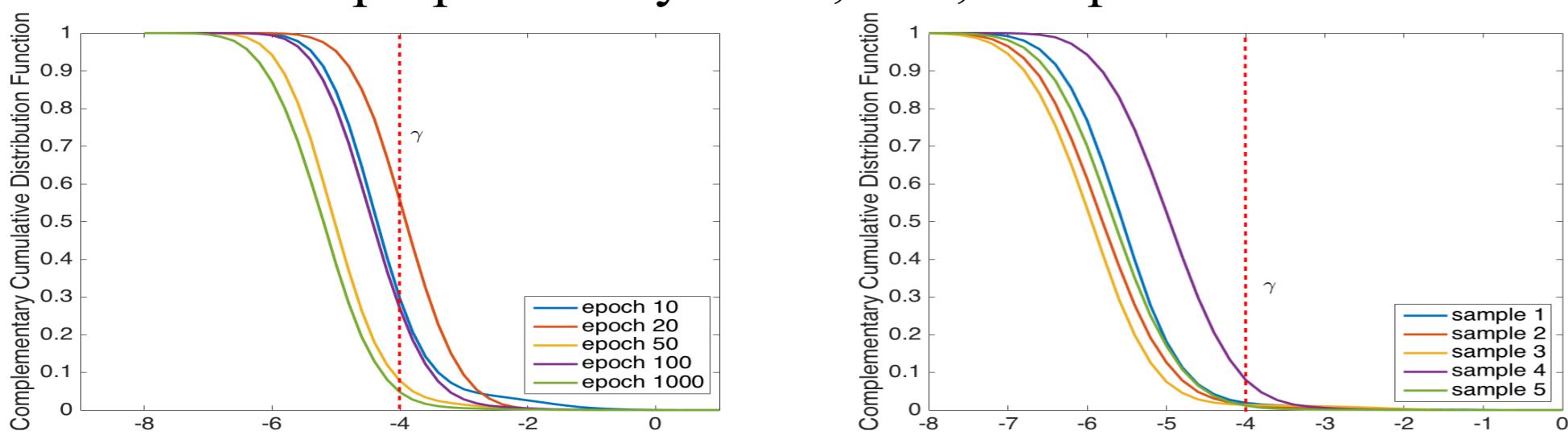
Example:

- Positive: (*Steve Jobs*, *FounderOf*, *Apple Inc.*)
- Low-quality: (*Baseball*, *FounderOf*, *Apple Inc.*)
- High-quality: (*Bill Gates*, *FounderOf*, *Apple Inc.*)

- Quality can be measured by scores

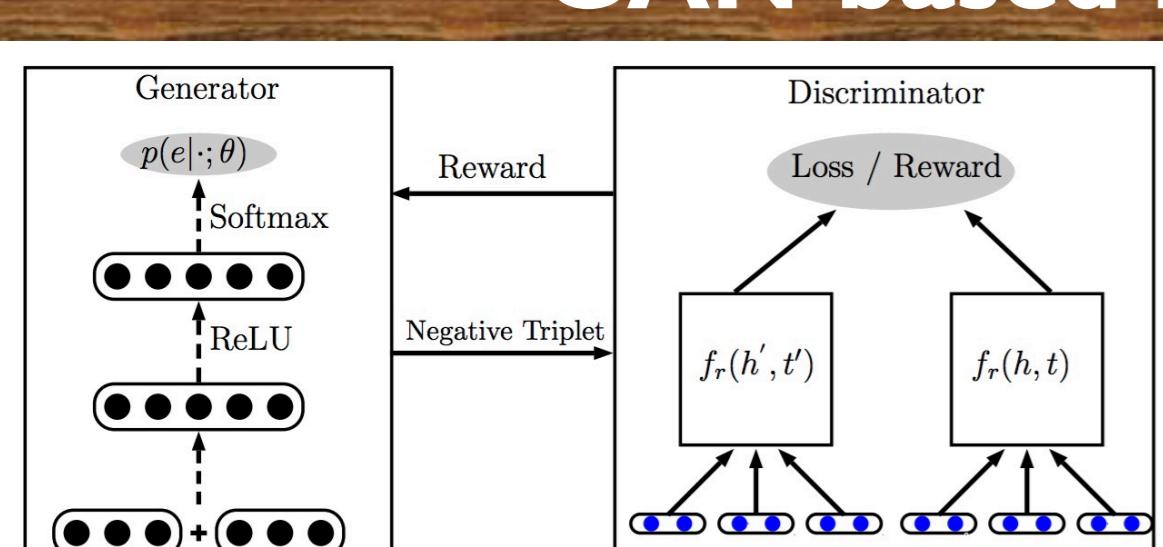
Observations and Challenges

- The score distribution of negative triplets is **highly skewed**.
- Several hard properties: dynamic, rare, complex.



- How to model the **dynamic** distribution of negative triplets.
- How to sample high-quality negative triplets in an efficient way.

GAN-based baselines



IGAN [Wang et.al. AAAI 2018]
 KBGAN [Cai et.al. NAACL 2018]
 Self-space NE [Gao et al. KDD 2018]

Drawbacks:

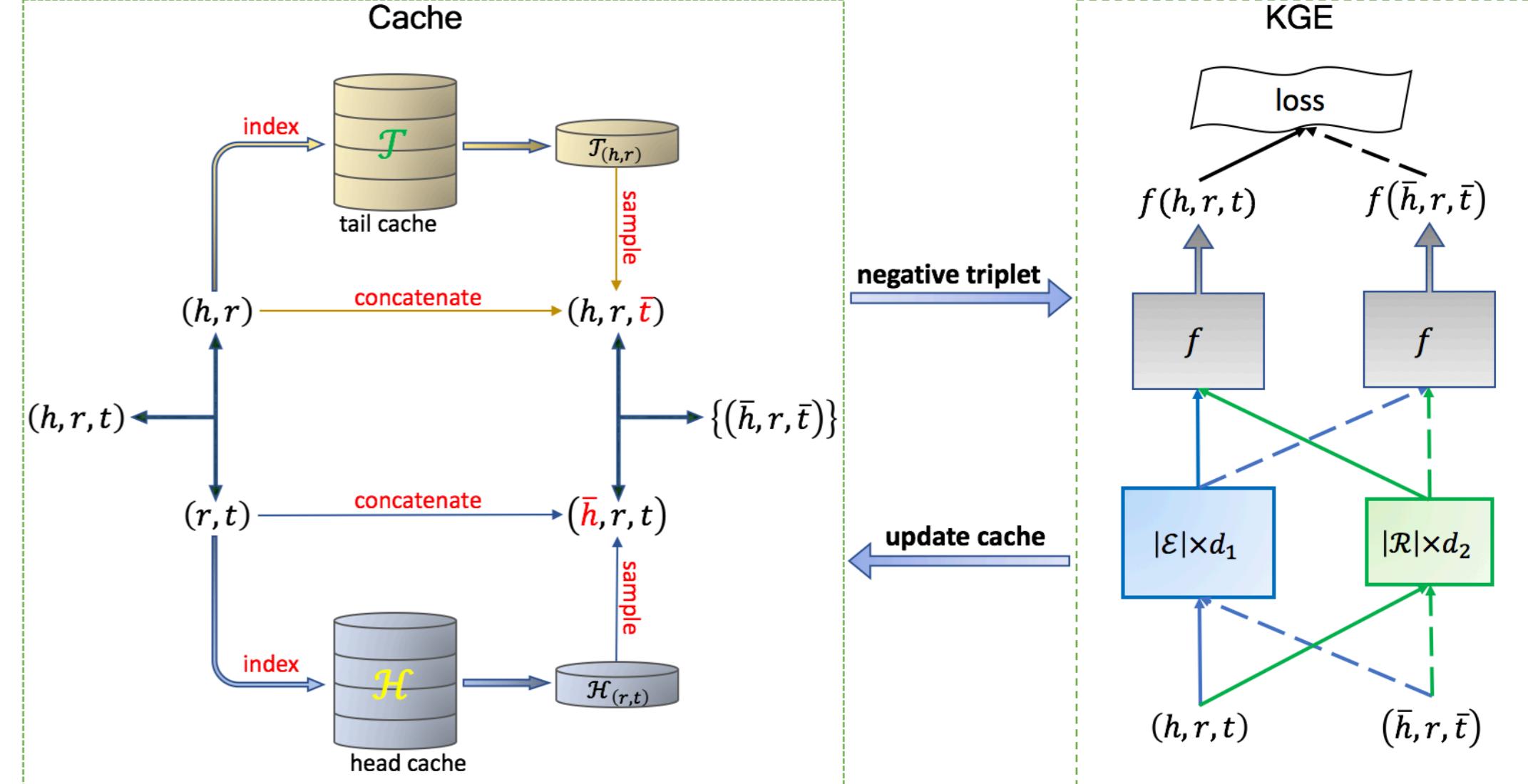
- Needs to train an **extra** model;
- Sampling is **not efficient**;
- Training suffers from **instability**.

NSCaching

Motivation:

- Use a small amount of extra memory, which caches negative triplets with large scores during training.
- Sample the negative triplet directly from the cache.

Framework:



Design schemes

Sampling from the cache

- Uniform sampling
- Importance sampling
- Selecting the top

Updating the cache

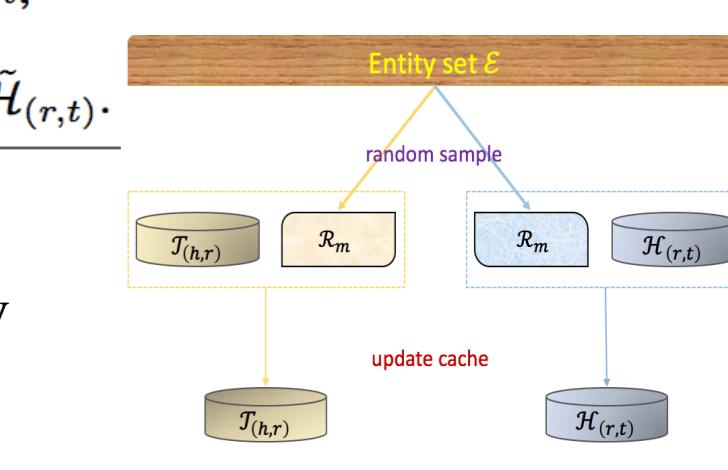
- Selecting top-k
- Importance sampling

Consideration

- Balance Exploration and Exploitation.
- Learn from easy samples first and gradually change to hard ones.

Algorithm 3 Updating head-cache (step 8).
Input: head cache $\mathcal{H}_{(r,t)}$ of size N_1 , triplet $(h, r, t) \in \mathcal{S}$.

- initialize $\tilde{\mathcal{H}}_{(r,t)} \leftarrow \emptyset$;
- uniformly sample a subset $\mathcal{R}_m \subset \mathcal{E}$ with N_2 entities;
- $\tilde{\mathcal{H}}_{(r,t)} \leftarrow \mathcal{H}_{(r,t)} \cup \mathcal{R}_m$;
- compute the score $f(\bar{h}, r, t)$ for all $\bar{h} \in \tilde{\mathcal{H}}_{(r,t)}$;
- for** $i = 1, \dots, N_1$ **do**
- sample $\bar{h} \in \tilde{\mathcal{H}}_{(r,t)}$ with probability in Equation (6);
- remove \bar{h} from $\tilde{\mathcal{H}}_{(r,t)}$;
- $\tilde{\mathcal{H}}_{(r,t)} \leftarrow \tilde{\mathcal{H}}_{(r,t)} \cup \bar{h}$;
- end for**
- update by $\mathcal{H}_{(r,t)} \leftarrow \tilde{\mathcal{H}}_{(r,t)}$.

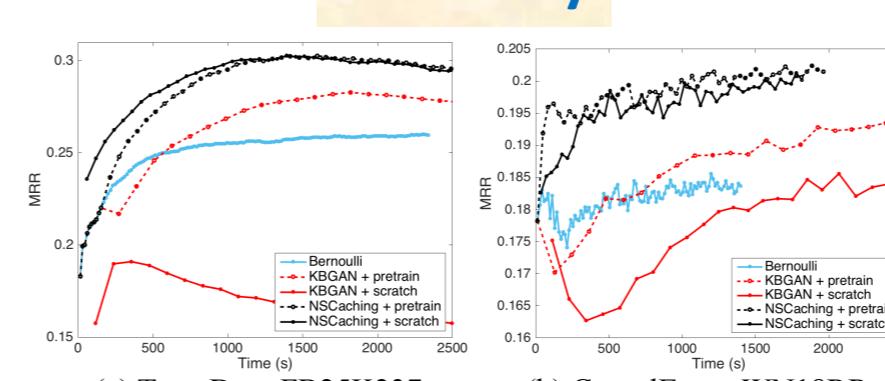


Experiments

Measurements:

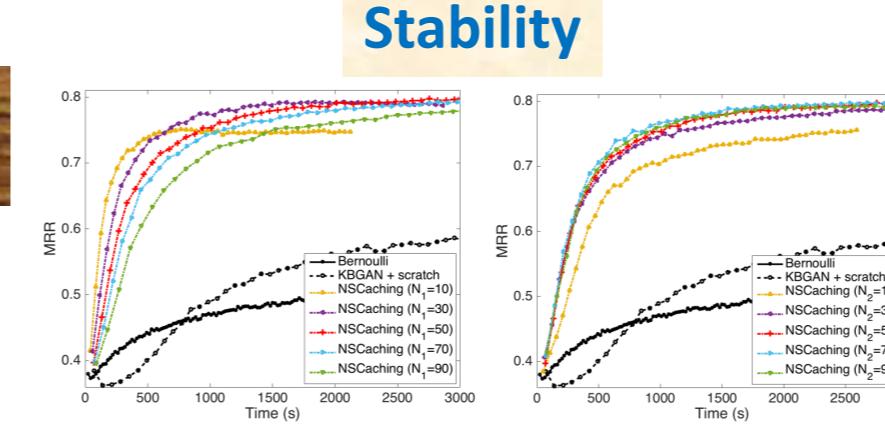
- Given a triplet (h, r, t) ,
- compute $(h', r, t), \forall h' \in \mathcal{E}$,
- rank the score of h among all h' ;
- same for t .

Efficiency



(a) TransD on FB2K237. (b) ComplEx on WN18RR.

Stability



(a) Diff. N1. (b) Diff. N2.

Visualization

Tail-cache of triplet (*manorama*, *profession*, *actor*)

epoch	entities in cache
0	<i>allen_clark</i> , <i>jose_gola</i> , <i>ostrava</i> , <i>ben_llilly</i> , <i>hans_zinsser</i>
20	<i>accountant</i> , <i>frank_pais</i> , <i>laura_marx</i> , <i>como</i> , <i>domitila_lepida</i>
100	<i>artist</i> , <i>aviator</i> , <i>hans_zinsser</i> , <i>john_h_cough</i>
200	<i>physician</i> , <i>artist</i> , <i>raich_carter</i> , <i>coach</i> , <i>mark_shivas</i>
500	<i>artist</i> , <i>physician</i> , <i>cavan</i> , <i>sex_worker</i> , <i>attorney_at_law</i>

epoch	entities in cache
0	<i>allen_clark</i> , <i>jose_gola</i> , <i>ostrava</i> , <i>ben_llilly</i> , <i>hans_zinsser</i>
20	<i>accountant</i> , <i>frank_pais</i> , <i>laura_marx</i> , <i>como</i> , <i>domitila_lepida</i>
100	<i>artist</i> , <i>aviator</i> , <i>hans_zinsser</i> , <i>john_h_cough</i>
200	<i>physician</i> , <i>artist</i> , <i>raich_carter</i> , <i>coach</i> , <i>mark_shivas</i>
500	<i>artist</i> , <i>physician</i> , <i>cavan</i> , <i>sex_worker</i> , <i>attorney_at_law</i>

epoch	entities in cache
0	<i>allen_clark</i> , <i>jose_gola</i> , <i>ostrava</i> , <i>ben_llilly</i> , <i>hans_zinsser</i>
20	<i>accountant</i> , <i>frank_pais</i> , <i>laura_marx</i> , <i>como</i> , <i>domitila_lepida</i>
100	<i>artist</i> , <i>aviator</i> , <i>hans_zinsser</i> , <i>john_h_cough</i>
200	<i>physician</i> , <i>artist</i> , <i>raich_carter</i> , <i>coach</i> , <i>mark_shivas</i>
500	<i>artist</i> , <i>physician</i> , <i>cavan</i> , <i>sex_worker</i> , <i>attorney_at_law</i>

epoch	entities in cache
0	<i>allen_clark</i> , <i>jose_gola</i> , <i>ostrava</i> , <i>ben_llilly</i> , <i>hans_zinsser</i>
20	<i>accountant</i> , <i>frank_pais</i> , <i>laura_marx</i> , <i>como</i> , <i>domitila_lepida</i>
100	<i>artist</i> , <i>aviator</i> , <i>hans_zinsser</i> , <i>john_h_cough</i>
200	<i>physician</i> , <i>artist</i> , <i>raich_carter</i> , <i>coach</i> , <i>mark_shivas</i>
500	<i>artist</i> , <i>physician</i> , <i>cavan</i> , <i>sex_worker</i> , <i>attorney_at_law</i>