On the Scalability of Graph Kernels Applied to Collaborative Recommenders

Jérôme Kunegis, Andreas Lommatzsch,

Christian Bauckhage, Şahin Albayrak

DAI-Labor, Technische Universität Berlin

18th European Conference on Artificial Intelligence, Patras

Workshop on Recommender Systems

CC IRML
Information Retrieval
& Machine Learning

Collaborative Filtering – Definition

Users rate items

- Task: Recommend items to users
- Task: Find similar users
- Task: Find similar items
- Task: Predict missing ratings

…using only ratings, not the content

Examples: Amazon, Netflix, …

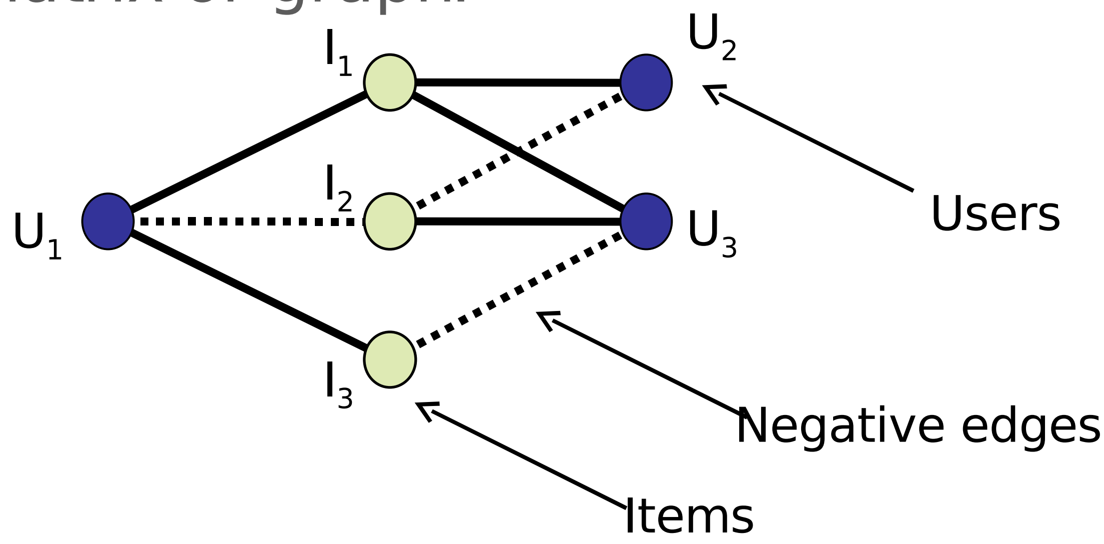
Dipl.-Inform. Jérôme Kunegis I kunegis@dai-labor.de

Rating Models

Ratings modeled as matrix or graph:

R=

I\U	U_1	U_2	U_3
I_1	+1	+1	+1
I_2	-1	-1	+1
I_3	+1		-1



A = [0 R; R^T 0] is the adjacency matrix of G

- The rating graph is bipartite with weighted edges
- The adjacency matrix is symmetric, sparse

Dipl.-Inform. Jérôme Kunegis I kunegis@dai-labor.de

Dimensionality Reduction

Find a low-rank approximation to A

$$A \approx \tilde{A} = U_k \Lambda U_k^T$$

Entry $\tilde{A}_{ui}$ contains prediction for A_{ui}

The parameter k is the reduced dimension.

Dipl.-Inform. Jérôme Kunegis I kunegis@dai-labor.de

Kernels

A kernel is a similarity function that is positive semi-definite

Represented as a matrix: $K = GG^T$

(when the underlying space is finite)

The matrix $\tilde{A}$ is a kernel when $\Lambda \geq 0$.

Dipl.-Inform. Jérôme Kunegis I kunegis@dai-labor.de

Exponential Diffusion Kernel

The matrix exponential of A

$$\exp(\alpha A) = \Sigma_i (\alpha^i/i!) A^i$$

Is a weighted sum of the powers A^i

- A^i contains the number of paths of length i between all node pairs
- exp(A) is a weighted mean of a path counts, weighted by the inverse factorial of their length, and powers of the parameter α.

Laplacian Kernel

Define the Laplacian matrix L

$$L = D - A \text{ with } D_{ii} = \Sigma_j A_{ij}$$

It's pseudo-inverse $K = L^+$ is a kernel

- Using $K = GG^T$, we can consider the nodes in the space given by G
- Distances in this space correspond to the resistance distance

Dipl.-Inform. Jérôme Kunegis I kunegis@dai-labor.de

Laplacian Kernel – Variants

Other Laplacian kernels:

$(D - A)^{+}$	resistance distance kernel
$(1 - \alpha)(I - \alpha D^{-1} A)$	stochastic diffusion kernel
$\exp(-\alpha L)$	Laplacian exponential diffusion kernel
$(I + \alpha(\gamma D - A))^{-1}$	regularized Laplacian kernel

Dipl.-Inform. Jérôme Kunegis I kunegis@dai-labor.de

Signed Laplacian Kernels

- Laplacian kernels are suited for positive data
- Ratings are signed: use signed Laplacian kernels

Let $L^s = D^s - A$

with $D^s_{ii} = \Sigma_j |A_{ij}|$

(see my presentation about *Modeling Collaborative Similarity with the Signed Resistance Distance Kernel* on Thursday)

Dipl.-Inform. Jérôme Kunegis I kunegis@dai-labor.de

Computation

All kernels are computed using the eigenvalue decomposition of A or L.

Exponentiation and pseudoinversion are simple to apply to the decomposition:

$$\exp(U\Lambda U^T) = U \exp(\Lambda) U^T$$

$$(U\Lambda U^T)^+ = U\Lambda^+ U^T$$

Dipl.-Inform. Jérôme Kunegis I kunegis@dai-labor.de

Prediction and Recommendation

- Normalization: $r_{norm} = (2r - \mu_u - \mu_i) / (\sigma_u + \sigma_i)$

(where mu/i and su/i are the mean and standard deviation of u/is ratings)

Rating prediction using the nearest-neighbor/weighted-mean algorithm:

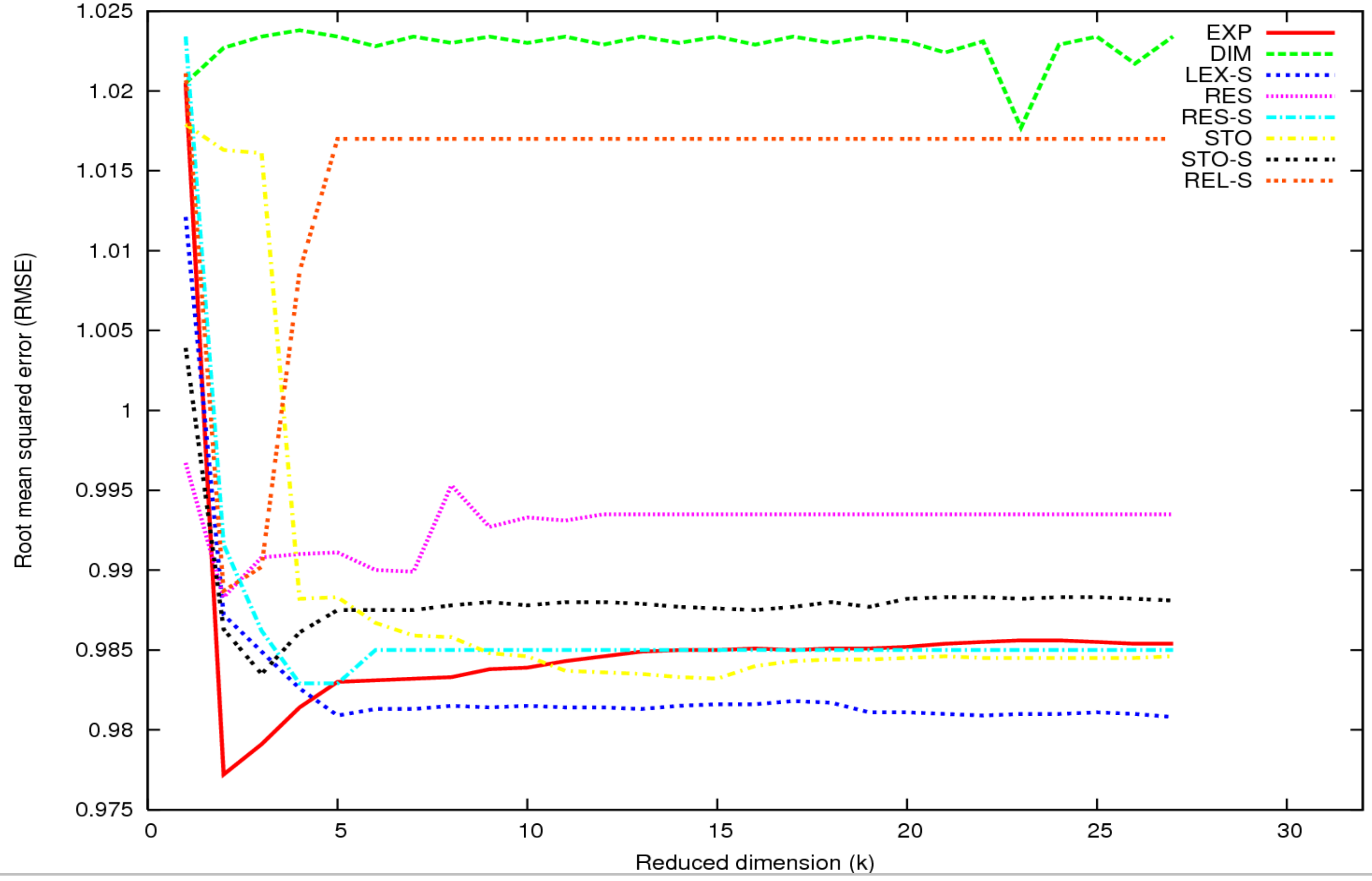
- Find nearest neighbors according to the kernel that have already rated the item

- Compute weighted mean of their ratings, using the kernel as weight

For **recommendation**, rank all items by their prediction

Evaluation – Prediction with Netflix

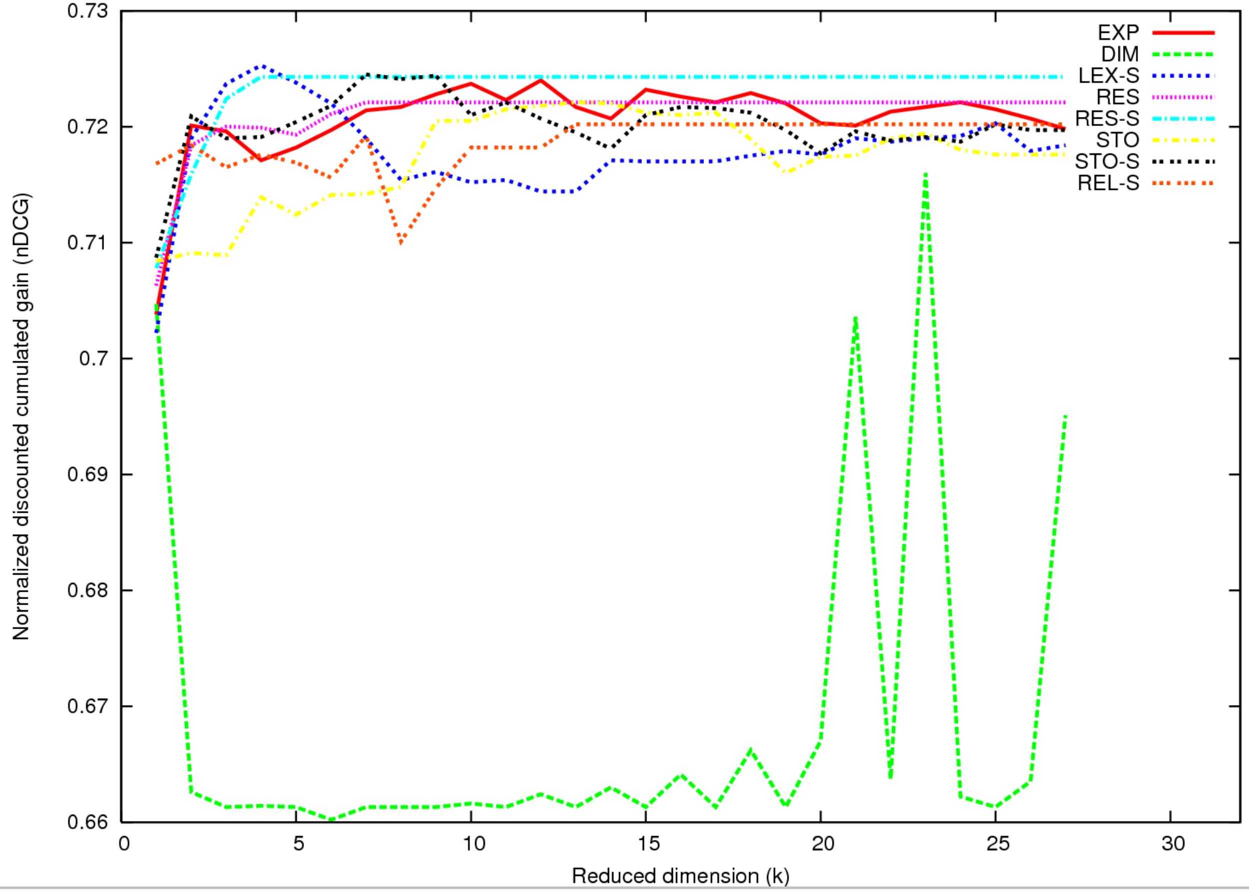
Root mean squared error on a subset of the Netflix Prize rating prediction task



DAI-Labor TU Berlin

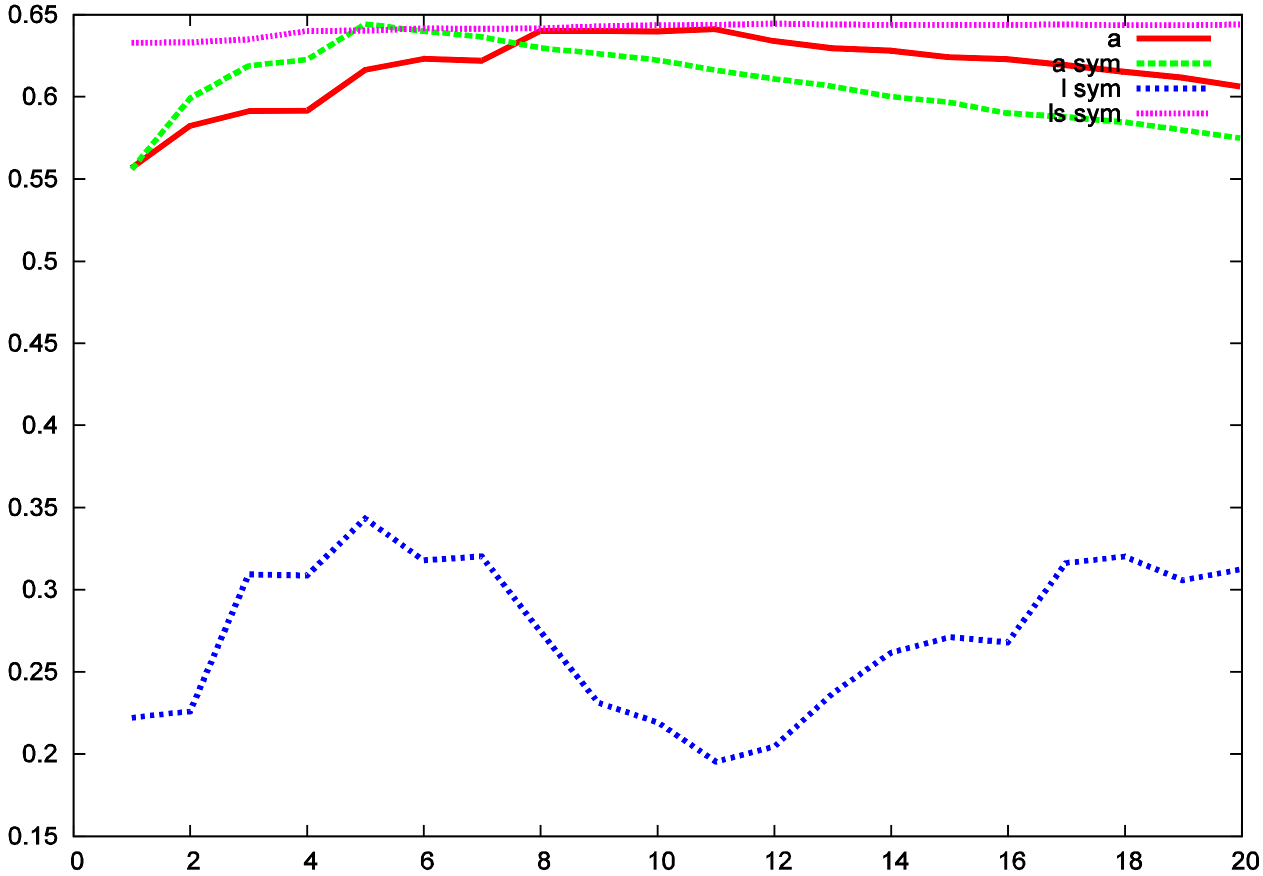
Evaluation – Recommendation with Netflix

Recommendation on the Netflix data, showing the normalized discounted cumulated gain



Evaluation – Prediction with the Slashdot Zoo

Work in progress: predict whether users are friends or foes.
Show mean score (−1 on error, +1 correct prediction)



DAI-Labor
TU Berlin

Observations 1

- Best performance may be achieved with a specific k or is best asymptotically for big k
- General trend: simple dimensionality reduction suffers from overfitting. Laplacian and to a lesser extent exponential kernels are good for large k.
- Best k is attained (if not at infinity), at k = 2, 3 for Netflix and k = up to 12 for Slashdot Zoo
- The Laplacians are smoother, making the choice of k less relevant
- Good recommender/bad predictor: Prediction seems to be harder that recommendation

Dipl.-Inform. Jérôme Kunegis I kunegis@dai-labor.de

Observations 2

- Good recommender/bad predictor: Prediction seems to be harder that recommendation. What is more relevant in practice?
- Signed kernels perform better than equivalent unsigned kernels.

THANK YOU!

(and don't forget my presentation on Thursday about the *Signed Resistance Distance* in CMI4 at 11:30)

Dipl.-Inform. Jérôme Kunegis I kunegis@dai-labor.de