Superhighway: Bypass Data Sparsity in Cross-Domain Collaborative Filtering

Kwei-Herng Lai*, Ting-Hsiang Wang*, Heng-Yu Chi†, Yian Chen†, Ming-Feng Tsai❦, Chuan-Ju Wang*
*Academia Sinica, †KKBOX Inc., ❦National Chengchi University

Motivation

In modern e-commerce, **cross-domain CF** is proposed to alleviate **data sparsity** in the target domain by leveraging information from related source domains. Existing methods often pose **density prerequisite** or require **domain-specific knowledge**. In addition, most methods report **unilateral improvement** on the target domain only.

Problem:

- ✘ Density prerequisite for the source domain limits the range of addressable problems.
- ✘ Pre-compiled knowledge may not exist for a given recommendation scenario.
- ❓ What if nuanced differences in each domains can be used to promote mutual improvement in cross-domain CF?

Application:

- Cross-platform recommender localization
- Cross-region recommender localization
- And their converse

Conclusion

Remarks:

- Superhighway …
  - ✓ poses no density prerequisite
  - ✓ leverages in-system information only
  - ✓ shows improvement in the source domain
- Superhighway achieves improvement not through intra-domain interactions, thus **bypass** data sparsity.
- Experiments verify the effectiveness of superhighway in cross-platform and cross-region recommendations.

Takeaway:

- Cross-domain connectivity enhancement may be a worthy approach to address cross-domain CF due to its **undirected** and **self-contained** property.

Experiments

Dataset:

- Cross-platform movie dataset (Movielens—Netflix)
- Cross-region music dataset (KKBOX_R1—KKBOX_R2)

Baseline:

- **Single:** CF on target domain only
- **Highway:** source and target joined
- **Superhighway:** highway with upgrade
- **Pre-train and fine-tune:** source then target

Item-based recommendation:

	MAP@10	MF	DeepWalk	HPE
Music	Single (Pretrained)	30.4 (30.3)	19.6 (22.2)	14.2 (27.8)
	Highway	30.5	0.193	0.2
	Superhighway	32.4	22.6	31.1
Movie	Single (Pretrained)	5.5 (5.3)	2.8 (1.7)	4.2 (6.3)
	Highway	1.4	2.0	0.014
	Superhighway	6.8	4.0	7.4

- Observation 1: Superhighway improves performance in target domain across setups.
- Observation 2: Results are unstable for all other compared baselines.
- Observation 3: Similar results are observed in source domain.

Codes: Network Embedding

Superhighway Construction

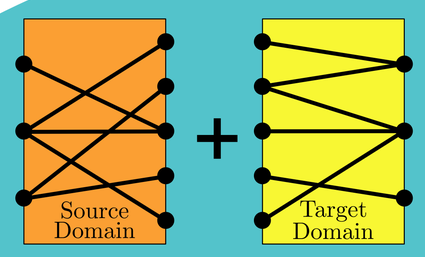
Formulation:

- WOLOG, given a source and a target domain with non-disjoint items and disjoint users, **superhighway construction** seeks to enhance **cross-domain connectivity** by inferring cross-domain user-user relations based on shared items.

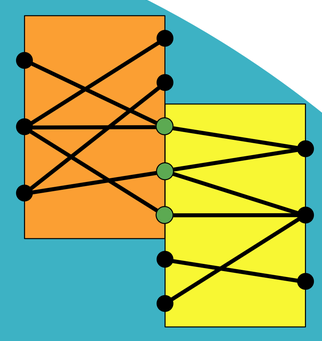
Structure:

- **Highway:** existing user-item-user cross-domain metapath
- **Superhighway:** user-user cross-domain relation inferred from highways

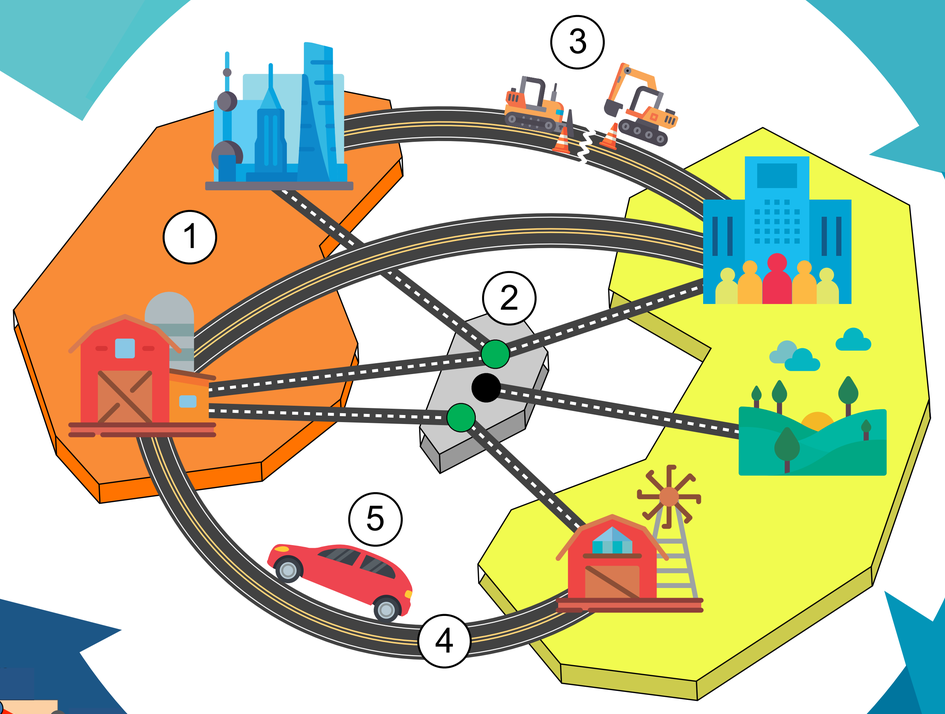
Paper: arXiv



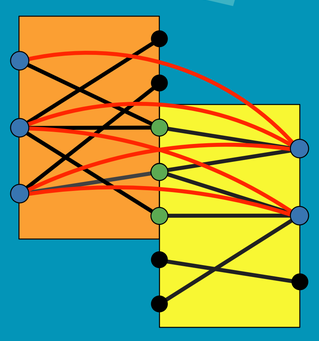
1 Cross-domain CF utilizes information in the source domain to assist CF in the target domain.



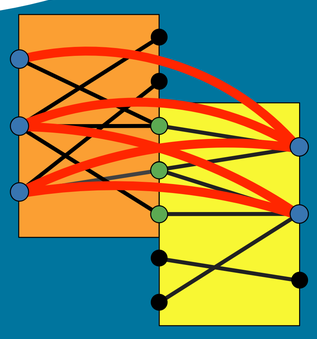
Identify shared nodes serving as midpoints for the **highway** structure.



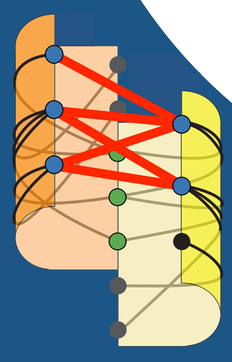
2 Infer the ***superhighway*** structure based on shared midpoints abiding to a smoothness threshold.



3 Scale ***superhighway*** weight to control cross-domain sampling likelihood.



4 ***Superhighway*** provides a new dimension for aligning domains.



5

Control factor:

- Ensure data **smoothness** during highway upgrade via:

③ $$\hat{U}_d = \left\{ u \,\middle|\, u \in U_d, \frac{|\mathcal{N}(u) \cap \tilde{I}|}{|\mathcal{N}(u)|} \geq \alpha \right\}$$

- Enhance "domain stitching" effect of superhighway via:

④ $$w = \beta \times |\mathcal{N}(u_i) \cap \mathcal{N}(u_j)|$$

Codes: Superhighway Construction

Contact: thwang@citi.sinica.edu.tw