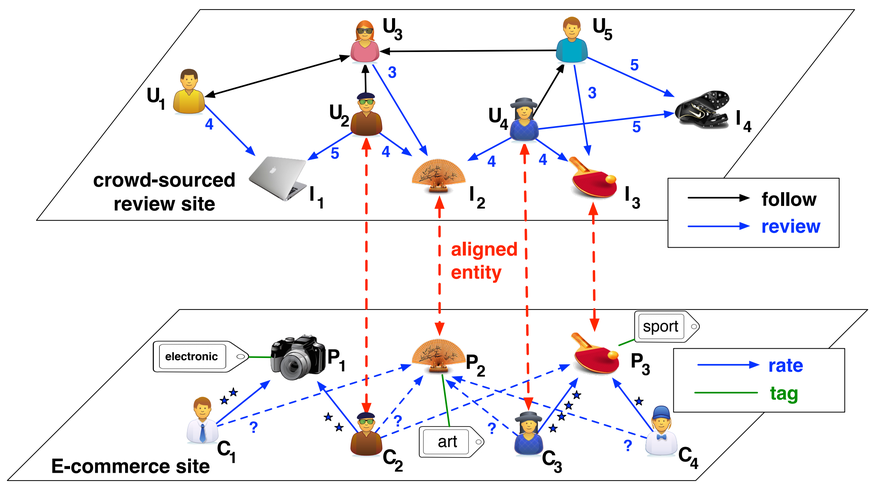
Item Recommendation for Emerging Online Businesses

Chun-Ta Lu (clu29@uic.edu), Sihong Xie, Weixiang Shao, Lifang He and Philip S. Yu

1.Item Recommendation for Emerging Businesses



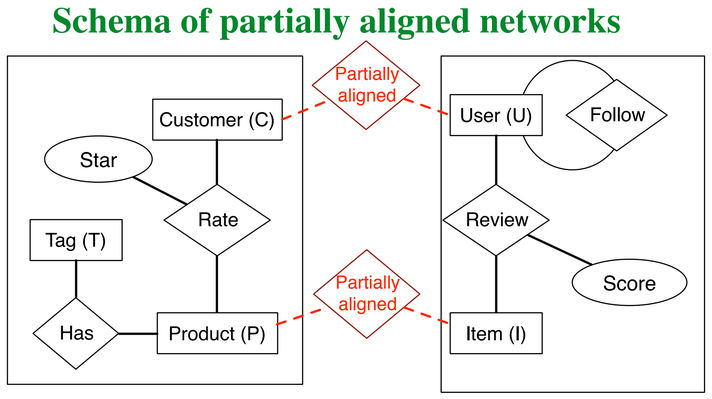
(I) Emerging Business: new network with very sparse information.
(II) Partially Aligned Networks: networks sharing common entities

- How to measure similarity between users/items with sparse information?
- How to utilize information from aligned networks to improve performance?

Hint: Transfer knowledge through aligned entities.

2. Augmented Meta Path-based Similarity (AmpSim)

Schema of partially aligned networks



Examples of Augmented Meta Path (AMP)

Intra-network augmented meta path

$(C \xrightarrow{[star]} P \xleftarrow{[star]} C)$,$(P \xleftarrow{[star]} C \xrightarrow{[star]} P)$, $(P \to T \leftarrow P)$

$(C \xrightarrow{[star]} P \xleftarrow{[star]} C \xrightarrow{[star]} P \xleftarrow{[star]} C)$

$(P \xleftarrow{[star]} C \xrightarrow{[star]} P \xleftarrow{[star]} C \xrightarrow{[star]} P)$

Inter-network augmented meta path

$(P \leftrightarrow I \xleftarrow{[score]} U \xrightarrow{[score]} I \leftrightarrow P)$

$(C \leftrightarrow U \xrightarrow{[score]} I \xleftarrow{[score]} U \leftrightarrow C)$, $(C \leftrightarrow U \leftrightarrow U \leftrightarrow C)$

both linkage structures and augmented link attributes are taken into account

How to measure similarity via AMPs?

Normalize link attributes to deal with different types of link attributes upon an AMP:

$$\mathbf{M}_{ui} = \frac{Y_{ui} - b_i}{\sqrt{\sum_j (Y_{uj} - b_j)^2}}$$

Y: link attributes

Ordinal attributes: $b_i = \overline{Y_i}$

Other attributes: $b_i = 0$

After normalization, $\mathbf{M}\mathbf{M}^{\mathrm{T}}$ is (adjusted) cosine similarity

AmpSim: AMP-based similarity measure between v_a and v_b based on path $\mathcal{P}$

$$s(v_a, v_b|\mathcal{P}) = \frac{[\prod_{i=1}^{l} \mathbf{M}_i]_{ab} + [\prod_{i=l}^{1} \mathbf{M}_i^T]_{ba}}{[\prod_{i=1}^{l} \mathbf{M}_i]_{a*} + [\prod_{i=l}^{1} \mathbf{M}_i^T]_{b*}} \in [0,1]$$

path direction is taken into account

$\prod_{i=1}^{l} \mathbf{M}_i$ is the product of the normalized link attributes upon $\mathcal{P}$

Average over multiple AMPs:

$$S(v_a, v_b) = \frac{tanh(\sum_i w_i S_i(v_a, v_b))}{tanh(1)} \in [0,1] \qquad \sum_i w_i = 1$$

3. Recommendation Model: AmpCMF

$$\min_{\substack{\mathbf{P}^{(\pi)},\mathbf{Q}^{(\pi)} \geq 0, \\ \pi \in \{s,t\}}} \mathcal{J} = \sum_{\pi \in \{s,t\}} ||\mathbf{I}^{(\pi)} \odot (\mathbf{Y}^{(\pi)} - \mathbf{P}^{(\pi)}\mathbf{Q}^{(\pi)^T})||_F^2 + \beta(\mathcal{R}_A(\mathbf{P}^{(t)}, \mathbf{P}^{(s)}) + \mathcal{R}_A(\mathbf{Q}^{(t)}, \mathbf{Q}^{(s)}))$$

Collective Matrix Factorization (CMF) w/ alignment regularization

$$+ \lambda(\mathcal{R}_G(\mathbf{P}^{(t)}) + \mathcal{R}_G(\mathbf{Q}^{(t)}))$$

geometric regularization

$$+ \frac{\alpha}{2} \sum_{\pi \in \{s,t\}} (||\mathbf{P}^{(\pi)}||_F^2 + ||\mathbf{Q}^{(\pi)}||_F^2)$$

prevent overfitting

$\mathbf{Y}$ and $\mathbf{I} \in \mathbb{R}^{n \times m}$ are user-item feedback and indicator matrices

$\mathbf{P} \in \mathbb{R}^{n \times k}$ and $\mathbf{Q} \in \mathbb{R}^{m \times k}$ are low rank representations of users and items

alignment regularization: $\mathcal{R}_A(\mathbf{P}^{(t)}, \mathbf{P}^{(s)}) = \frac{1}{2} \sum_{(i,g_i) \in \mathcal{A}^{(t,s)}} ||\mathbf{p}_i^{(t)} - \mathbf{p}_{g_i}^{(s)}||_2^2 = \frac{1}{2}||\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{P}^{(t)} - \mathbf{A}\mathbf{P}^{(s)}||_F^2$

geometric regularization: $\mathcal{R}_G(\mathbf{P}) = \frac{1}{2} \sum_{u,v} \mathbf{S}_{uv} ||\mathbf{p}_u - \mathbf{p}_v||_2^2 = \frac{1}{2} Tr(\mathbf{P}^T(\mathbf{D} - \mathbf{S})\mathbf{P}) = \frac{1}{2} Tr(\mathbf{P}^T \mathbf{L}_P \mathbf{P})$

✦ **Solve the optimization problem using multiplicative updates**

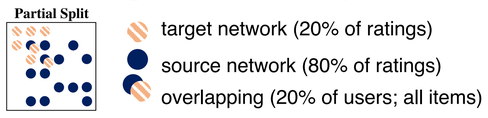
4. Experiment Results

Name	#users	#items	#tags	#ratings	#social links
Yelp	26,618	8,467	770	230,418	183,765
Epinions	21,740	31,678	27	541,108	344,286

Yelp: test a mature business entering a new domain

target network: Nevada w/ 11.5K users; 2.1K items
source network: Arizona w/ 19.5K users; 6.3K items
overlapping: 4.3K users; no items

Epinions: test an emerging business partially aligned with a developed business



WNMF + geometric regularization

- **WNMF:** weighted nonnegative matrix factorization
- **Hete-MF:** WNMF+ PathSim between items only
- **Hete-CF:** WNMF + PathSim between users/items
- **Hete-PRW:** WNMF + Pairwise RW similarity between users/items
- **Amp-MF:** WNMF + AmpSim between users/items

Transfer learning based methods

- **CMF:** collective matrix factorization w/ alignment regularization
- **RMGM:** rating-matrix generative model
- **Amp-CMF:** proposed method

cold-start: test on users who have ≤ 5 ratings in training set
90% (55%) of users are cold-start users in Yelp (Epinions)

Yelp (Overall)

Method	RMSE		MAE	
	$k=10$	$k=20$	$k=10$	$k=20$
WNMF	1.446 ± 0.009	1.429 ± 0.009	1.097 ± 0.006	1.083 ± 0.005
Hete-MF	1.429 ± 0.009	1.351 ± 0.009	1.086 ± 0.005	1.006 ± 0.005
Hete-CF	1.305 ± 0.008	1.199 ± 0.008	0.957 ± 0.005	0.907 ± 0.005
Hete-PRW	1.343 ± 0.008	1.313 ± 0.008	1.018 ± 0.005	0.991 ± 0.005
Amp-MF	$\mathbf{1.191 \pm 0.009}$	$\mathbf{1.187 \pm 0.009}$	$\mathbf{0.899 \pm 0.005}$	$\mathbf{0.897 \pm 0.005}$
CMF	1.294 ± 0.009	1.274 ± 0.010	0.966 ± 0.005	0.949 ± 0.006
RMGM	1.240 ± 0.009	1.238 ± 0.009	0.925 ± 0.005	0.902 ± 0.004
Amp-CMF	$\mathbf{1.134 \pm 0.009}$	$\mathbf{1.127 \pm 0.009}$	$\mathbf{0.854 \pm 0.005}$	$\mathbf{0.847 \pm 0.005}$

Yelp (Cold Start)

Method	RMSE		MAE	
	$k=10$	$k=20$	$k=10$	$k=20$
WNMF	1.535 ± 0.014	1.520 ± 0.013	1.184 ± 0.005	1.170 ± 0.005
Hete-MF	1.518 ± 0.012	1.492 ± 0.011	1.171 ± 0.005	1.148 ± 0.005
Hete-CF	1.378 ± 0.009	1.228 ± 0.009	1.017 ± 0.005	0.935 ± 0.005
Hete-PRW	1.414 ± 0.008	1.382 ± 0.008	1.088 ± 0.005	1.059 ± 0.005
Amp-MF	$\mathbf{1.219 \pm 0.008}$	$\mathbf{1.215 \pm 0.008}$	$\mathbf{0.928 \pm 0.005}$	$\mathbf{0.925 \pm 0.005}$
CMF	1.349 ± 0.012	1.329 ± 0.012	1.015 ± 0.005	0.998 ± 0.006
RMGM	1.316 ± 0.009	1.295 ± 0.009	0.995 ± 0.004	0.974 ± 0.004
Amp-CMF	$\mathbf{1.148 \pm 0.009}$	$\mathbf{1.139 \pm 0.009}$	$\mathbf{0.875 \pm 0.006}$	$\mathbf{0.865 \pm 0.006}$

Epinions (Overall)

Method	RMSE		MAE	
	$k=10$	$k=20$	$k=10$	$k=20$
WNMF	1.551 ± 0.009	1.533 ± 0.009	1.156 ± 0.008	1.153 ± 0.009
Hete-MF	1.402 ± 0.010	1.398 ± 0.009	1.034 ± 0.009	1.030 ± 0.008
Hete-CF	1.148 ± 0.008	1.141 ± 0.008	0.908 ± 0.007	0.901 ± 0.007
Hete-PRW	1.395 ± 0.009	1.392 ± 0.009	1.039 ± 0.005	1.030 ± 0.005
Amp-MF	$\mathbf{1.099 \pm 0.009}$	$\mathbf{1.097 \pm 0.009}$	$\mathbf{0.869 \pm 0.005}$	$\mathbf{0.868 \pm 0.005}$
CMF	1.152 ± 0.007	1.143 ± 0.007	0.870 ± 0.004	0.868 ± 0.005
RMGM	1.246 ± 0.008	1.242 ± 0.010	0.989 ± 0.005	0.983 ± 0.007
Amp-CMF	$\mathbf{1.097 \pm 0.009}$	$\mathbf{1.095 \pm 0.009}$	$\mathbf{0.867 \pm 0.005}$	$\mathbf{0.866 \pm 0.005}$

Epinions (Cold Start)

Method	RMSE		MAE	
	$k=10$	$k=20$	$k=10$	$k=20$
WNMF	1.596 ± 0.016	1.594 ± 0.016	1.219 ± 0.016	1.218 ± 0.013
Hete-MF	1.462 ± 0.017	1.454 ± 0.014	1.106 ± 0.016	1.101 ± 0.013
Hete-CF	1.211 ± 0.011	1.201 ± 0.011	0.961 ± 0.009	0.955 ± 0.009
Hete-PRW	1.434 ± 0.009	1.428 ± 0.009	1.072 ± 0.005	1.066 ± 0.005
Amp-MF	$\mathbf{1.131 \pm 0.009}$	$\mathbf{1.128 \pm 0.009}$	$\mathbf{0.899 \pm 0.005}$	$\mathbf{0.897 \pm 0.005}$
CMF	1.198 ± 0.012	1.185 ± 0.010	0.902 ± 0.009	0.899 ± 0.009
RMGM	1.271 ± 0.005	1.266 ± 0.009	1.013 ± 0.002	1.014 ± 0.006
Amp-CMF	$\mathbf{1.129 \pm 0.009}$	$\mathbf{1.127 \pm 0.009}$	$\mathbf{0.898 \pm 0.005}$	$\mathbf{0.896 \pm 0.005}$

Conclusion:

1. **Methods using more accurate similarity measure achieve better performance**
2. **Combining both geometric and alignment regularization boosts the performance**
3. **Learning from highly overlapped networks can achieve better improvement: improve 21% in Yelp and 29% in Epinions**

5.Acknowledgments

This work is supported in part by NSF through grants III-1526499, and CNS-1115234, NSFC through grant 61503253, and Google Research Award.