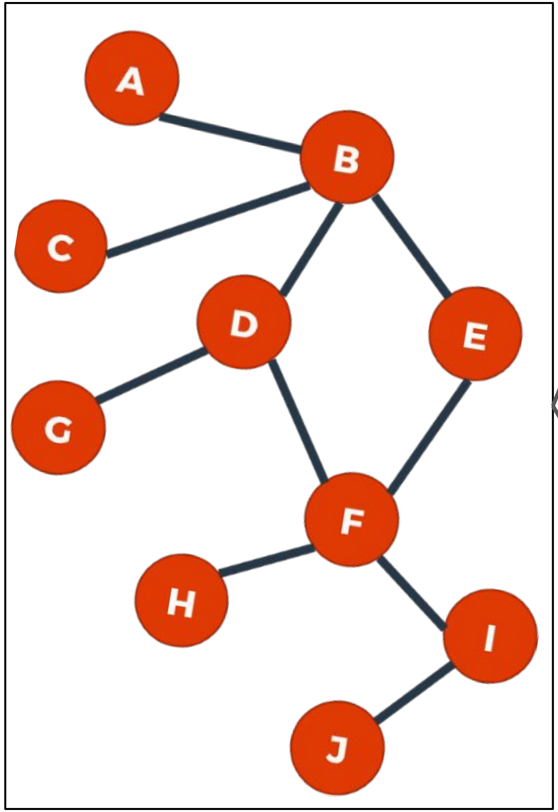
Nonnegative matrix factorization for link prediction in directed complex networks using PageRank and asymmetric link clustering information

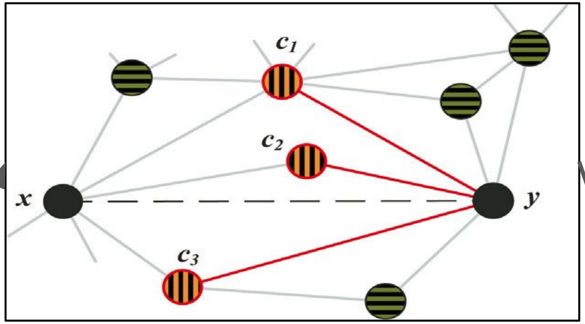
(By Guangfu Chen, Chen Xub, Jingyi Wang, Jianwen Feng, Jiqiang Feng)

Valerii Baianov, Dmitrii Leshchev, Gleb Mezentsev

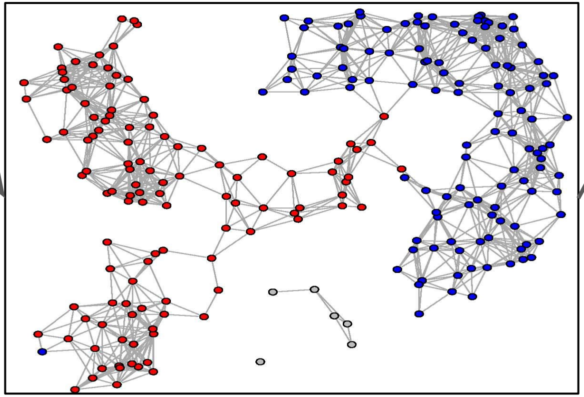
Skoltech, 2020



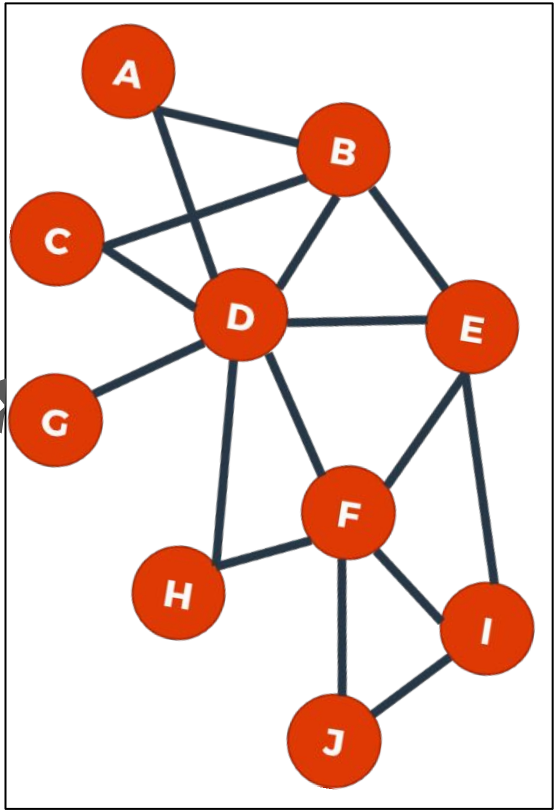
Initial graph



Local structures

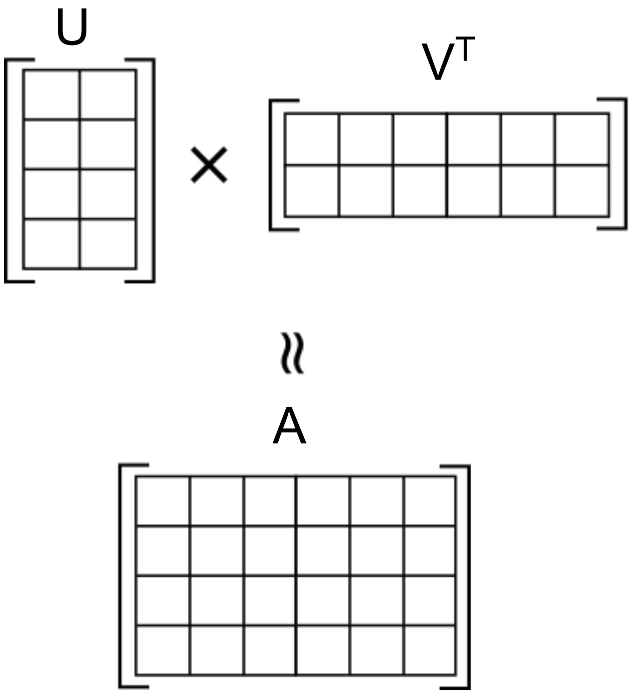


Global structures

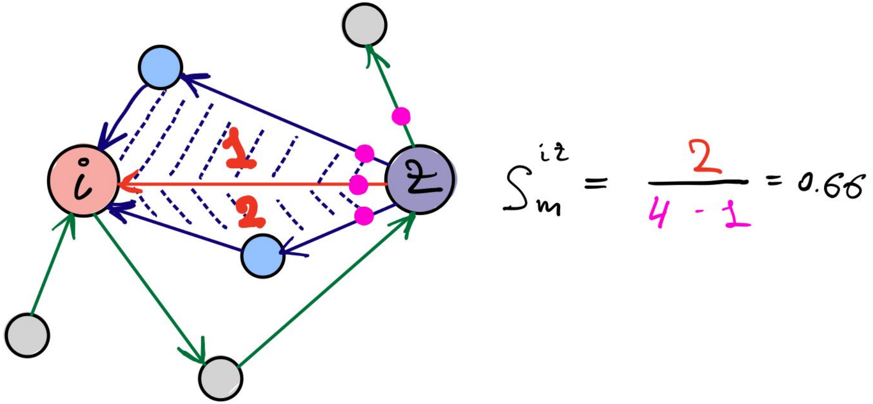


Final graph

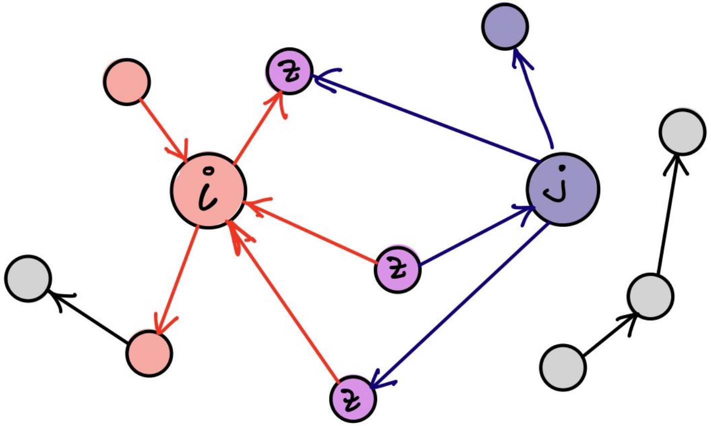
- $A \approx UV^T$
- $U \in \mathbb{R}_+^{n \times K},\ V \in \mathbb{R}_+^{n \times K}$
- $\mathcal{L}_{NMF} = \min_{U \geq 0, V \geq 0} \|A - UV^T\|_F^2$



$$S_m^{iz} = \frac{|\Gamma_{in}(i) \cap \Gamma_{out}(z)|}{k_{out}(z) - 1}$$



$$S_m^{ij} = \sum_{z \in \Gamma(i) \cap \Gamma(j)} S_m^{iz}$$



$$S_m^{ij} = \sum_{z \in \Gamma(i) \cap \Gamma(j)} S_m^{iz} \qquad S_m^{ji} = \sum_{z \in \Gamma(i) \cap \Gamma(j)} S_m^{jz}$$

$$S_m = max\{S_m^{ij}, S_m^{ji}\}$$

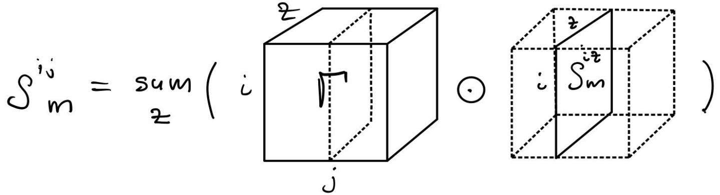
$$\mathcal{L}_{local} = \min_{U \geq 0, V \geq 0} \|(1 + \alpha S) \circ (A - UV^T)\|_F^2$$

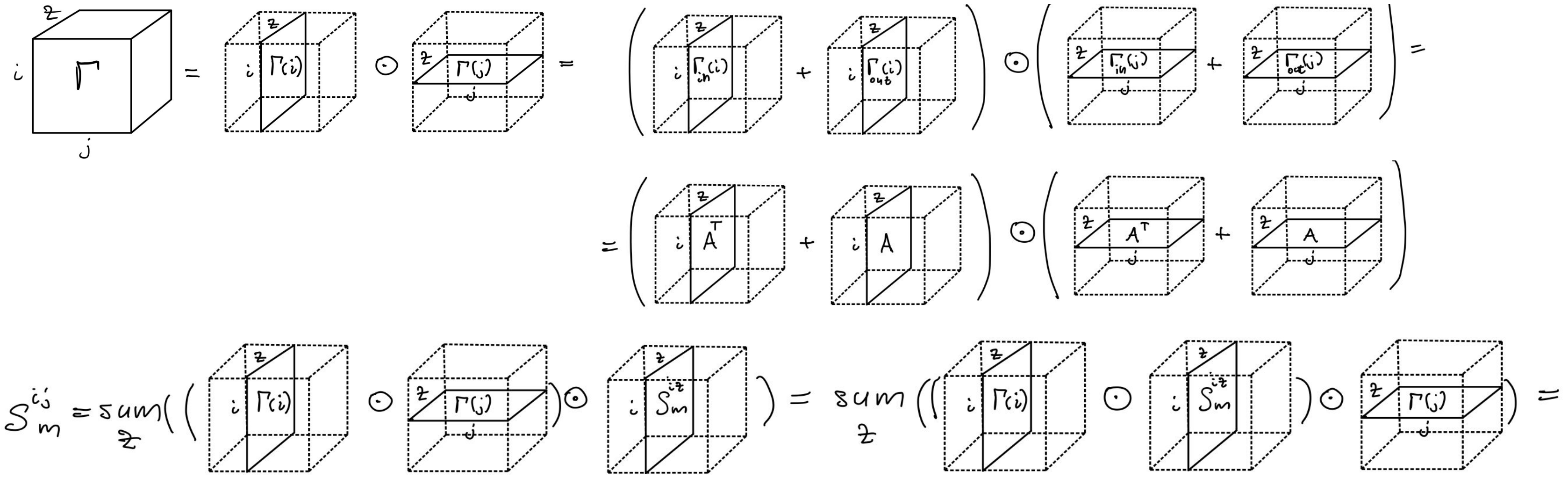
$$S_m^{iz} = \frac{|\Gamma_{in}(i) \cap \Gamma_{out}(z)|}{K_{out}(z) - 1}$$

$$S_m^{iz} = a_{zi} \frac{(A^2)_{iz}}{\sum_{k=1}^{n} a_{zk} - 1}$$

$$S_m^{iz} = A^T \odot \left(A^2 \cdot \mathrm{Diag}\left(\frac{1}{\sum_{k=1}^{n} a_{zk} - 1}\right)\right)$$

$$S_m^{ij} = \sum_{z \in \Gamma(i) \cap \Gamma(j)} S_m^{iz}$$





$$= \left(\boxed{\Gamma(i)} \odot \boxed{S_m^{iz}} \right) \cdot \boxed{\Gamma(i)} = ((A+A^T)\odot S_m^{iz})(A+A^T)$$

$$S_m^{ij} = ((A+A^T)\odot S_m^{iz})(A+A^T)$$

$$S_m^{ji} = (A+A^T)((A+A^T)\odot S_m^{iz\,T}) = S_m^{ij\,T}$$

Google

PageRank

$$C_{ij} = \begin{cases} c_i & \text{if } a_{ij} \neq 0 \\ 0 & \text{if } a_{ij} = 0 \end{cases}$$

$$C = c * A$$

$$+ \; W \in \mathbb{R}_+^{K \times K} \; = \; \mathcal{L}_{global} = \min_{U \geq 0, W \geq 0} \|C - UWU^T\|_F^2$$

$$\min_{U\geq 0, V\geq 0, W\geq 0} \mathcal{L} = \mathcal{L}_{local} + \gamma \mathcal{L}_{global} + \beta(\|U\|_F^2 + \|V\|_F^2)$$

$$\mathcal{L}_{local} = \min_{U\geq 0, V\geq 0} \|(1+\alpha S)\circ(A - UV^T)\|_F^2$$

$$\mathcal{L}_{global} = \min_{U\geq 0, W\geq 0} \|C - UWU^T\|_F^2$$

$$Y = 1 + \alpha S$$

$$\min_{U\geq 0, V\geq 0, W\geq 0} \mathcal{L} = \|Y\circ(A - UV^T)\|_F^2 + \gamma\|C - UWU^T\|_F^2 + \beta(\|U\|_F^2 + \|V\|_F^2)$$

$$\mathcal{L} = Tr\{[Y \circ (A - UV^T)][Y \circ (A - UV^T)]^T\} + \gamma Tr[(C - UWU^T)(C - UWU^T)^T] + \beta(Tr(U^TU) + Tr(V^TV))$$

$$\Phi = [\phi_{nk}] \in \mathbb{R}^{n\times k}, \Psi = [\psi_{nk}] \in \mathbb{R}^{n\times k} \ and \ \Theta = [\theta_{kk}] \in \mathbb{R}^{k\times k}$$

$$\mathcal{L} = Tr\{[Y \circ (A - UV^T)][Y \circ (A - UV^T)]^T\} + \gamma Tr[(C - UWU^T)(C - UWU^T)^T] + \beta Tr(U^TU) + \beta Tr(V^TV) + Tr(\Phi U^T) + Tr(\Omega V^T) + Tr(\Theta W^T)$$

Paper derivatives

$$\frac{\partial \mathcal{L}}{\partial V} = -(Y \circ A)^T U + (Y \circ (UV^T))^T U + \beta V + \Psi$$

$$\frac{\partial \mathcal{L}}{\partial W} = -U^T C U + U^T U W U^T U + \Theta$$

$$\frac{\partial \mathcal{L}}{\partial U} = -Y \circ AV + Y \circ \left(UV^T\right) V + \beta U - \gamma C U W^T + \gamma U W U^T U U W^T + \Phi$$

Our derivatives

$$\frac{\partial \mathcal{L}}{\partial V} = 2\beta V - 2(Y^T \circ (A^T - V U^T) \circ Y^T) U + \Psi$$

$$\frac{\partial \mathcal{L}}{\partial W} = -2\gamma U^T (C - U W U^T) U + \Theta$$

$$\frac{\partial \mathcal{L}}{\partial U} = 2\beta U - (2(Y \circ (A - U V^T) \circ Y) V + 2\gamma (C - U W U^T) U W^T + 2\gamma (C^T - U W^T U^T) U W) + \Phi$$

$$U_{nk} \leftarrow U_{nk} \frac{[(Y \circ A \circ Y)V + \gamma(CUW^T + C^T UW)]_{nk}}{[(Y \circ (UV^T) \circ Y) + \gamma(UWU^T UW^T + UW^T U^T UW) + \beta U]_{nk}}$$

$$V_{nk} \leftarrow V_{nk} \frac{[((Y \circ A)^T \circ Y^T)U]_{nk}}{[\beta V + ((Y \circ (UV^T))^T \circ Y^T)U]_{nk}}$$

$$W_{kk} \leftarrow W_{kk} \frac{[U^T CU]_{kk}}{[U^T UWU^T U]_{kk}}$$

Our updating rules

$$W_{kk} \leftarrow W_{kk} \frac{[U^T CU]_{kk}}{[U^T UWU^T U]_{kk}}$$

$$V_{nk} \leftarrow V_{nk} \frac{[(Y \circ A)^T U]_{nk}}{[(Y \circ (UV^T))^T U + \beta V]_{nk}}$$

$$U_{nk} \leftarrow U_{nk} \frac{[Y \circ AV + \gamma CUW^T]_{nk}}{[Y \circ (UV^T)V + \beta U + \gamma UWU^T UUW^T]_{nk}}$$

Paper updating rules

Algorithm 1 Algorithm NMF-AP.

Input:

A: adjacency matrix of directed network;
K: dimension of latent space;
N_{iter}: maximum number of iterations;
Paramenters: α, β, γ;

Output:

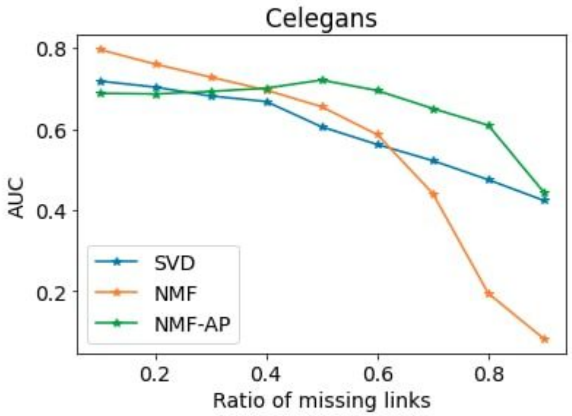
Similarity score matrix $\widehat{A}$

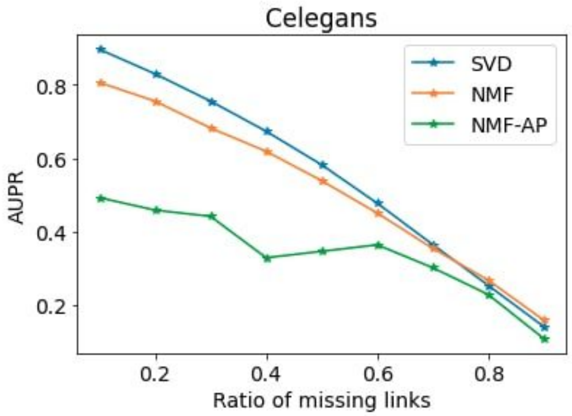
Divide A into training set E^T and probe set E^P
Randomly initialize U,V
Preserve the local information according to Eq. (7)
Preserve the global information according to Eq. (10)
For t=1:iter do
Update U according to Eq. (16)
Update V according to Eq. (18)
Update W according to Eq. (20)
Get U and V after convergence;
endfor
Compute probability matrix for link prediction $\widehat{A} = UV^T$

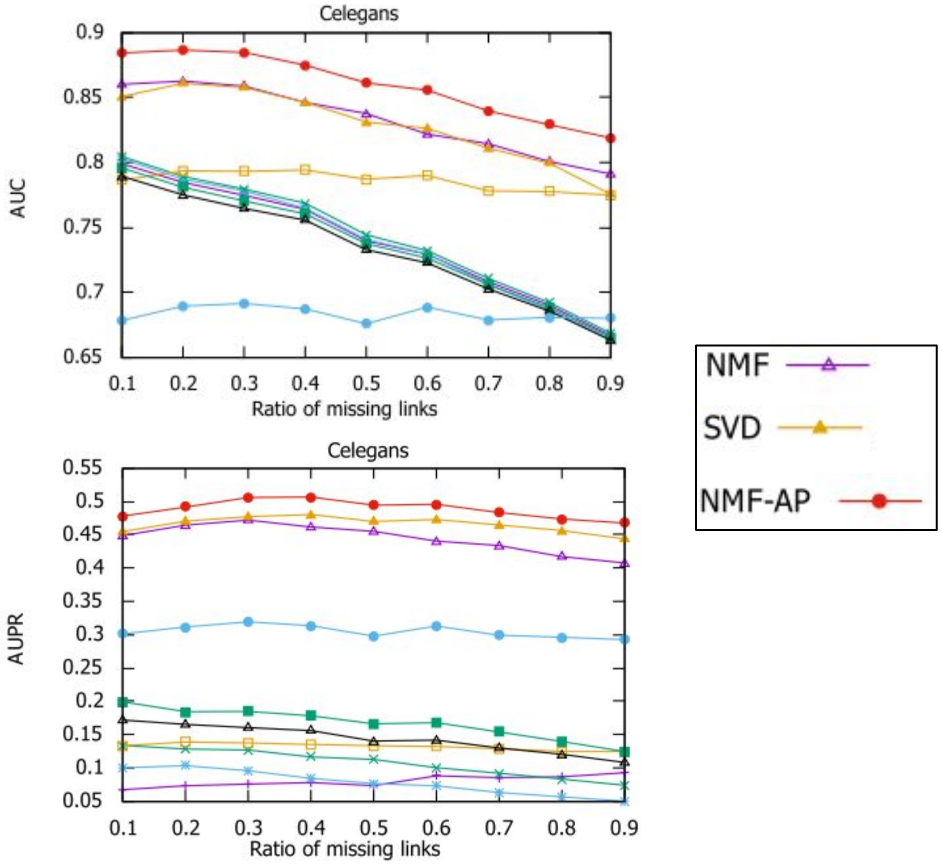
$$Y = 1 + \alpha S$$

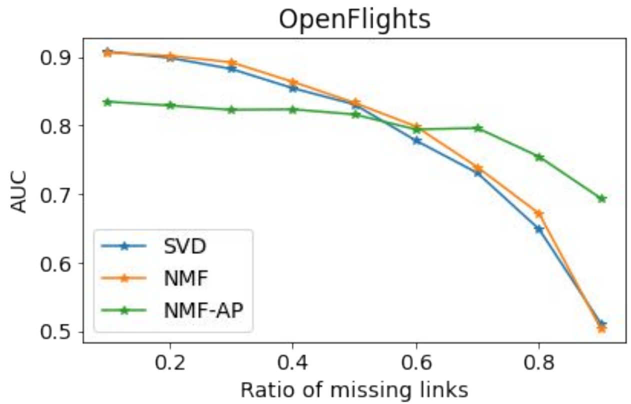
$$U_{n\times k}, W_{k\times k}, k < n$$
$$UW^TU^TUW - O(kn^2)$$
$$U(W^T(U^TU)W) - O(nk^2)$$

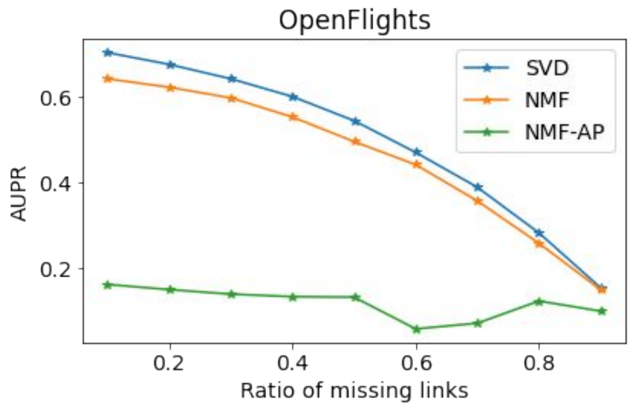
$$S_m^{ij} = ((A + A^T) \circ S_m^{iz})(A + A^T)$$
$$S_m^{ji} = (A + A^T)((A + A^T) \circ (S_m^{iz})^T) = (S_m^{ij})^T$$

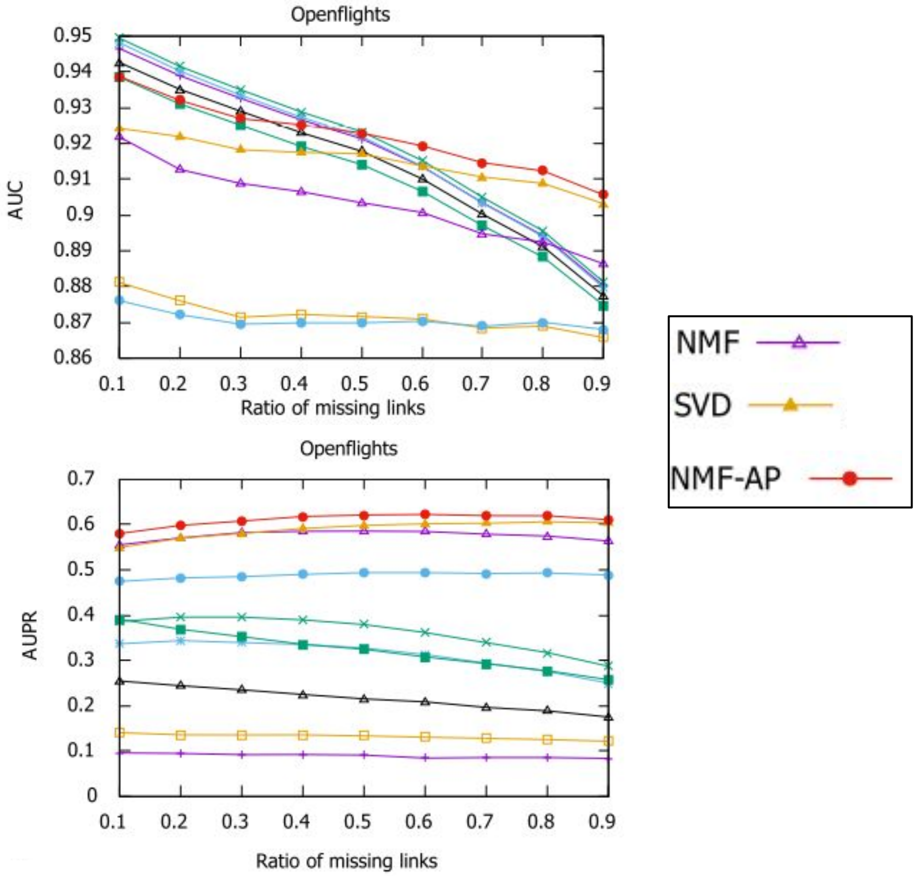

Celegans
AUC
SVD
NMF
NMF-AP
Ratio of missing links

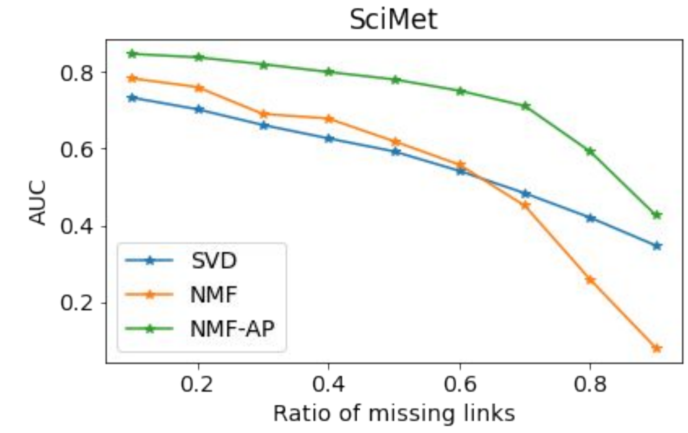

Celegans
AUPR
SVD
NMF
NMF-AP
Ratio of missing links

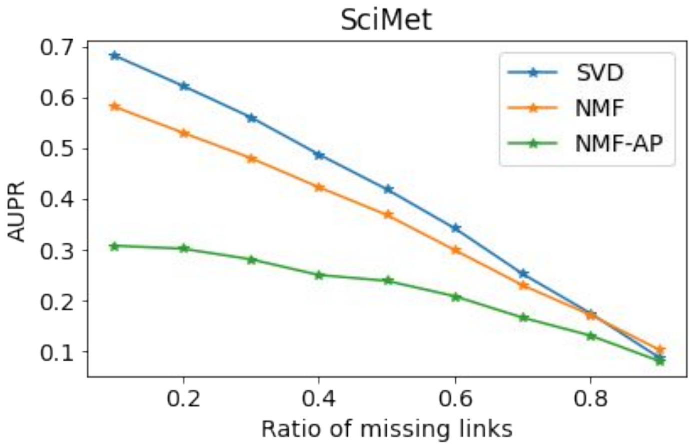

Celegans
AUC
Ratio of missing links
Celegans
AUPR
Ratio of missing links
NMF
SVD
NMF-AP

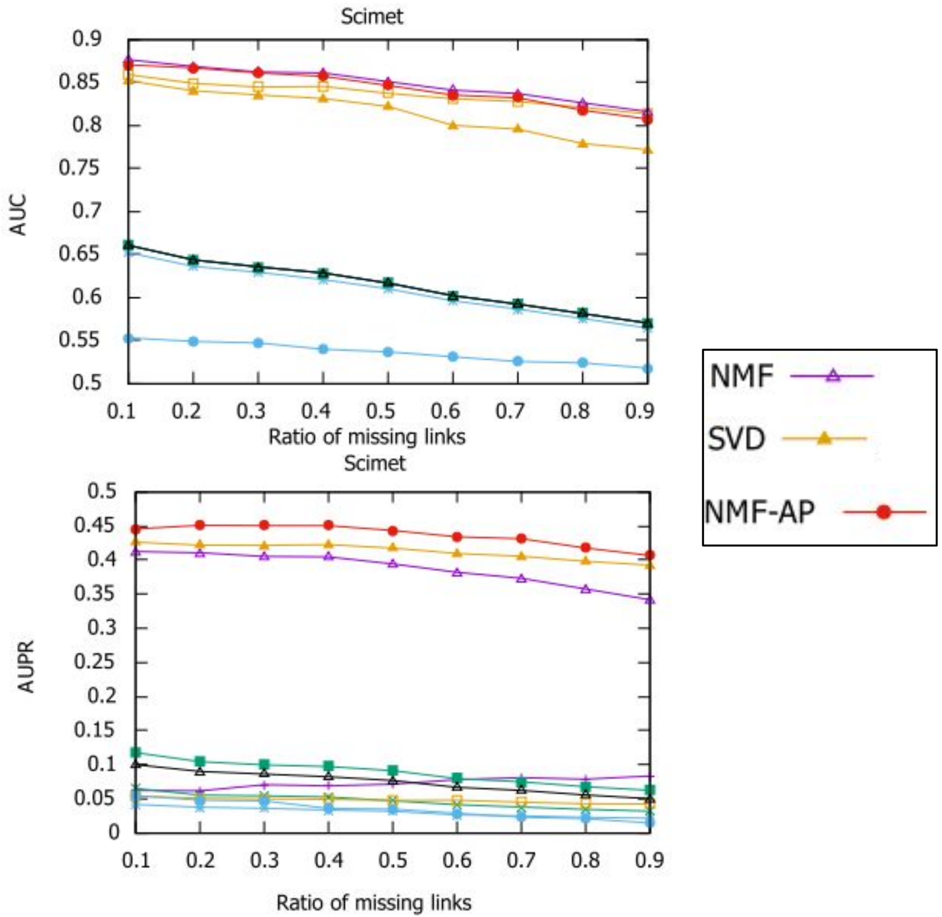

OpenFlights
AUC
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
Ratio of missing links
SVD
NMF
NMF-AP

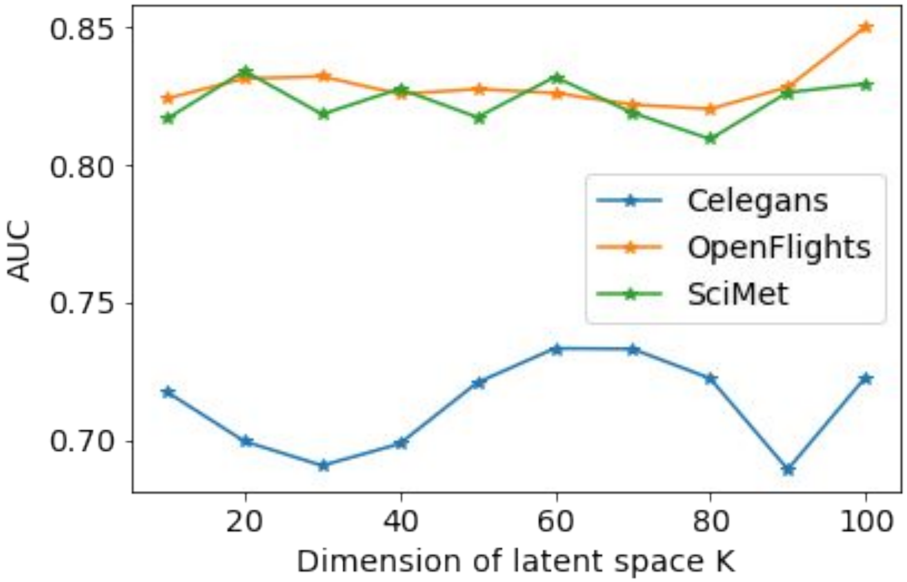

OpenFlights
AUPR
0.6
0.4
0.2
0.2
0.4
0.6
0.8
Ratio of missing links
SVD
NMF
NMF-AP

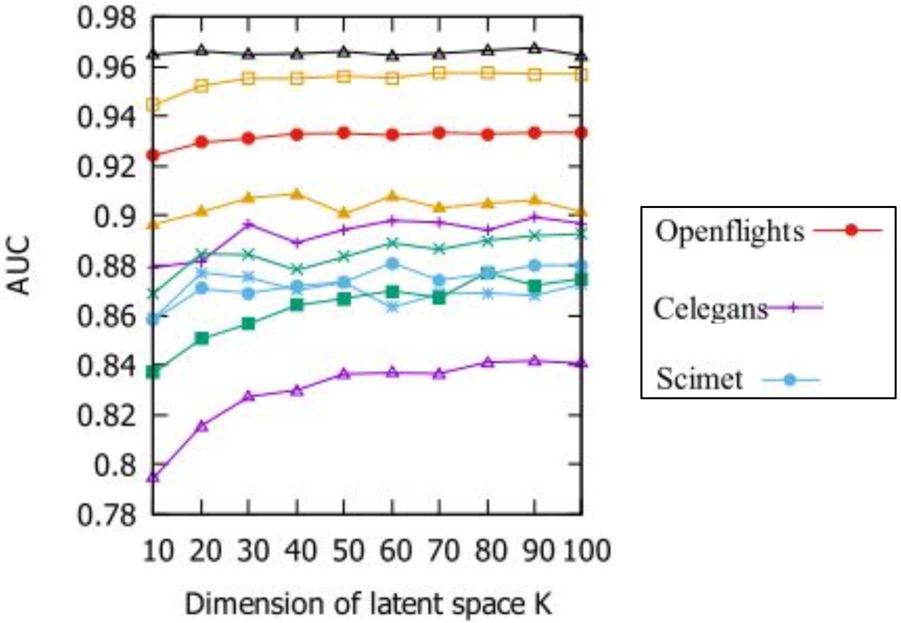

Openflights
AUC
0.95
0.94
0.93
0.92
0.91
0.9
0.89
0.88
0.87
0.86
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Ratio of missing links
Openflights
AUPR
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Ratio of missing links
NMF
SVD
NMF-AP


SciMet
AUC
0.2
0.4
0.6
0.8
Ratio of missing links
SVD
NMF
NMF-AP


SciMet
AUPR
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Ratio of missing links
SVD
NMF
NMF-AP


Scimet
AUC
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
Ratio of missing links
Scimet
AUPR
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
Ratio of missing links
NMF
SVD
NMF-AP


0.85
0.80
0.75
0.70
AUC
20
40
60
80
100
Dimension of latent space K
Celegans
OpenFlights
SciMet


0.98
0.96
0.94
0.92
0.9
0.88
0.86
0.84
0.82
0.8
0.78
AUC
10 20 30 40 50 60 70 80 90 100
Dimension of latent space K
Openflights
Celegans
Scimet

- Sparse and dense implementation for proposed algorithm
- Several NLA tricks for computational and storage saving
- Obtained AUC score is matched with the paper both for models and datasets
- Obtained AUPR score is mismatched with the paper

THANK YOU FOR YOUR ATTENTION!