ULB UNIVERSITÉ LIBRE DE BRUXELLES

Apprentissage de similarités sur des grands graphes

Timothée Defoin

Université Libre de Bruxelles

4 Septembre 2019

Sommaire

Contexte

- Apprentissage semi-supervisé sur des graphes
- Complexité limitée par la taille des graphes
- Famille de méthodes basée sur les similarités
- Résultats peu consistants

Développement d'une méthode par Robin Devooght, Costas Bekas et Peter Staar

- Résultats robustes sur des datasets très divers
- Rapide

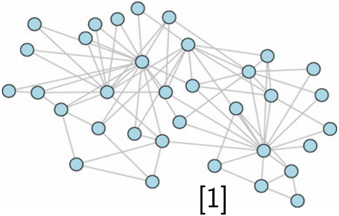
Objectifs

1) Implémenter l'algorithme dans un language open-source et reproduire les résultats obtenus sur de multiples datasets face à des benchmarks

2) Essayer des approches originales en vue d'améliorer les résultats

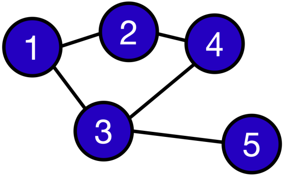
Qu'est-ce qu'un graphe?

- Ensemble de noeuds reliés par des arêtes
- Peut être orienté ou non
- Le degré d'un noeud est le nombre d'arêtes connecté à celui-ci



[1]

Matrices associées

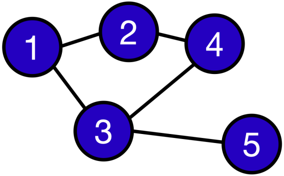


Matrice d'adjacence:

$$a_{ij} = \begin{cases} 1, & \text{si } (i,j) \text{ est une arête} \\ 0, & \text{sinon.} \end{cases}$$

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Matrices associées



Matrice des degrés:

$$d_{ij} = d(i)\delta_{ij}$$

$$\mathbf{D} = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Matrices associées

Matrice stochastique:

$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}$$

- Eléments donnés par

$$p_{ij} = \frac{a_{ij}}{d(i)}$$

- Probabilité de sauter dans un noeud voisin en supposant équiprobabilité

Matrice d'adjacence symétriquement normalisée:

$$\tilde{\mathbf{A}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$$

- Eléments maintenant symétriques

$$\tilde{a}_{ij} = \frac{a_{ij}}{\sqrt{d(i)d(j)}}$$

Similarités

- Assignent un scalaire réel aux paires de noeuds
- Forme matricielle $\mathbf{S}$ avec éléments s_{ij}
- Permettent de résoudre un grand nombre de problèmes
  - Clustering
  - Classification
  - ...

Similarités

Exemples:

- Personalized PageRank [2]:

$$\mathbf{S}_{\text{PPR}} = \left(\mathbf{I} - (1-\alpha)\mathbf{P}^T\right)^{-1}$$

- Exponential diffusion [3]:

$$\mathbf{S}_{\text{exp}} = \exp(-\alpha(\mathbf{I} - \tilde{\mathbf{A}}))$$

- Learning with local and global consistency (LLGC) [4]:

$$\mathbf{S}_{\text{LLGC}} = (\mathbf{I} - \alpha\tilde{\mathbf{A}})^{-1}$$

Fondement de la méthode [5]

Les mesures peuvent être exprimées comme une série de puissance:

$$\mathbf{S}_{\mathrm{PPR}} = \left(\mathbf{I} - (1-\alpha)\mathbf{P}^T\right)^{-1}$$

$$\mathbf{S}_{\mathrm{LLGC}} = (\mathbf{I} - \alpha\widetilde{\mathbf{A}})^{-1}$$

$$\mathbf{S}_{\mathrm{exp}} = \exp(-\alpha(\mathbf{I} - \widetilde{\mathbf{A}}))$$

Fondement de la méthode [5]

Les mesures peuvent être exprimées comme une série de puissance:

$$\mathbf{S}_{\text{PPR}} = \left(\mathbf{I} - (1-\alpha)\mathbf{P}^T\right)^{-1} = \sum_{i=0}^{\infty} ((1-\alpha)\mathbf{P}^T)^i$$

$$\mathbf{S}_{\text{LLGC}} = (\mathbf{I} - \alpha\tilde{\mathbf{A}})^{-1} = \sum_{i=0}^{\infty} (\alpha\tilde{\mathbf{A}})^i$$

$$\mathbf{S}_{\text{exp}} = \exp(-\alpha(\mathbf{I} - \tilde{\mathbf{A}})) = \sum_{i=0}^{\infty} \frac{(-\alpha(\mathbf{I} - \tilde{\mathbf{A}}))^i}{i!}$$

Fondement de la méthode [5]

Les mesures peuvent être exprimées comme une série de puissance:

$$\mathbf{S}_{\text{PPR}} = \left(\mathbf{I} - (1-\alpha)\mathbf{P}^T\right)^{-1} = \sum_{i=0}^{\infty} ((1-\alpha)\mathbf{P}^T)^i$$

$$\mathbf{S}_{\text{LLGC}} = (\mathbf{I} - \alpha\tilde{\mathbf{A}})^{-1} = \sum_{i=0}^{\infty} (\alpha\tilde{\mathbf{A}})^i$$

$$\mathbf{S}_{\text{exp}} = \exp(-\alpha(\mathbf{I} - \tilde{\mathbf{A}})) = \sum_{i=0}^{\infty} \frac{(-\alpha(\mathbf{I} - \tilde{\mathbf{A}}))^i}{i!}$$

et pour un k suffisamment grand $\rightarrow$

Fondement de la méthode [5]

Les mesures peuvent être exprimées comme une série de puissance:

$$\mathbf{S}_{\text{PPR}} = \left(\mathbf{I} - (1-\alpha)\mathbf{P}^T\right)^{-1} \approx \sum_{i=0}^{k} ((1-\alpha)\mathbf{P}^T)^i$$

$$\mathbf{S}_{\text{LLGC}} = (\mathbf{I} - \alpha\tilde{\mathbf{A}})^{-1} \approx \sum_{i=0}^{k} (\alpha\tilde{\mathbf{A}})^i$$

$$\mathbf{S}_{\text{exp}} = \exp(-\alpha(\mathbf{I} - \tilde{\mathbf{A}})) \approx \sum_{i=0}^{k} \frac{(-\alpha(\mathbf{I} - \tilde{\mathbf{A}}))^i}{i!}$$

Fondement de la méthode [5]

On cherche une similarité de la forme suivante:

$$\mathbf{S} = c_0\mathbf{I} + \sum_{i=1}^{k}(c_i(\mathbf{P}^T)^i + c_{k+i}\widetilde{\mathbf{A}}^i)$$

Quels coefficients prendre?

Calcul des coefficients [5]

- Problème de classification binaire avec communauté $\mathcal{A}$
- $\mathbf{y}^*$ le vecteur des labels complets

$$y_i^* = \begin{cases} 1 & \text{if } i \in \mathcal{A} \\ 0 & \text{if } i \in \bar{\mathcal{A}} \end{cases}$$

- Fraction $\mathcal{K}$ est connue
- Le vecteur des labels connus est

$$y_i = \begin{cases} 1 & \text{if } i \in \mathcal{A} \cap \mathcal{K} \\ 0 & \text{if } i \in \bar{\mathcal{A}} \cap \mathcal{K} \\ 0 & \text{otherwise.} \end{cases}$$

- Bonne similarité retrouve les labels dans $\bar{\mathcal{K}}$

Calcul des coefficients [5]

$$\hat{\mathbf{y}} = \mathbf{S}\mathbf{y} = \mathbf{S}\begin{pmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} . \\ . \\ . \\ \vdots \\ . \\ y^*_{n-x} \\ y^*_{n-x+1} \\ \vdots \\ y^*_n \end{pmatrix}$$

Calcul des coefficients [5]

Notation:

$$\mathbf{S}\mathbf{y} = \mathbf{F}(\mathbf{y})\mathbf{c}$$

avec

$$\mathbf{F}(\mathbf{y}) = \left(\mathbf{y},\ \mathbf{P}^T\mathbf{y},\ \cdots,\ (\mathbf{P}^T)^k\mathbf{y},\ \tilde{\mathbf{A}}\mathbf{y},\ \cdots,\ \tilde{\mathbf{A}}^k\mathbf{y}\right)$$

Calcul des coefficients [5]

- Vecteurs $\mathbf{y}^{\mathcal{S}_i}$ avec labels masqués

$$\mathbf{F} = \begin{pmatrix} \text{rows}_{\mathcal{S}_1}(\mathbf{F}(\mathbf{y}^{\mathcal{S}_1})) \\ \text{rows}_{\mathcal{S}_2}(\mathbf{F}(\mathbf{y}^{\mathcal{S}_2})) \\ \vdots \\ \text{rows}_{\mathcal{S}_\sigma}(\mathbf{F}(\mathbf{y}^{\mathcal{S}_\sigma})) \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} \text{rows}_{\mathcal{S}_1}(\mathbf{y}) \\ \text{rows}_{\mathcal{S}_2}(\mathbf{y}) \\ \vdots \\ \text{rows}_{\mathcal{S}_\sigma}(\mathbf{y}) \end{pmatrix}$$

- Problème de minimisation

$$\underset{\mathbf{c}}{\text{minimize}} \ ||\mathbf{Y} - \mathbf{F}\mathbf{c}||^2 + \lambda ||\mathbf{c}||^2$$

- Solution donnée par

$$\mathbf{c}^* = (\mathbf{F}^T\mathbf{F} + \lambda\mathbf{I})^{-1}\mathbf{F}^T\mathbf{Y}$$

Seed set expansion [5]

- Pas de négatifs connus
- Tous les inconnus sont considérés négatifs
- Ajout d'un terme pénalité

$$\underset{\mathbf{c}}{\text{minimize}}\ ||\mathbf{Y} - \mathbf{Fc}||^2 + ||\text{rows}_{\bar{\mathcal{K}}}(\mathbf{F}(\mathbf{y}))\mathbf{c}||^2 + \lambda||\mathbf{c}||^2.$$

Evaluation de la méthode [5]

Datasets utilisés:

	size	order	groups	type
BlogCatalog	10312	333983	29	Blogs
Flickr	80513	5899882	195	Image Hosting
DBLP	317080	1049866	149	Co-authorship
Amazon	334863	925872	295	Item sales
Youtube	1134890	2987624	136	Video Sharing

http://snap.stanford.edu/data
http://leitang.net/social_dimension.html

Evaluation de la méthode [5]

Métriques d'évaluation:

- Precision at 100 (prec100)
- Accuracy (acc)
- Normalized Discounted Cumulated Gain (NDCG)

$$\mathrm{NDCG} = \frac{1}{\mathrm{IDCG}} \sum_{i=1}^{n_{\text{test}}} \frac{2^{y_{\pi_i}} - 1}{\log(i+1)}$$

Evaluation de la méthode [5]

Méthode d'évaluation:

- 30 communautés tirées au hasard
- 10 seed sets générés par communauté (10%)
- Résultats calculés pour chaque benchmark et la méthode d'apprentissage
- Résultats standardisés avant d'être moyennés

Résultats

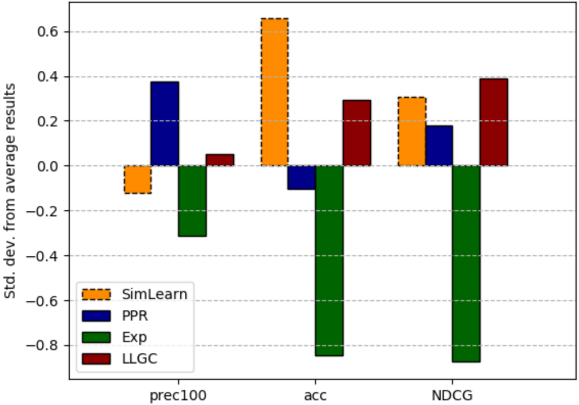


Figure: Amazon

Résultats

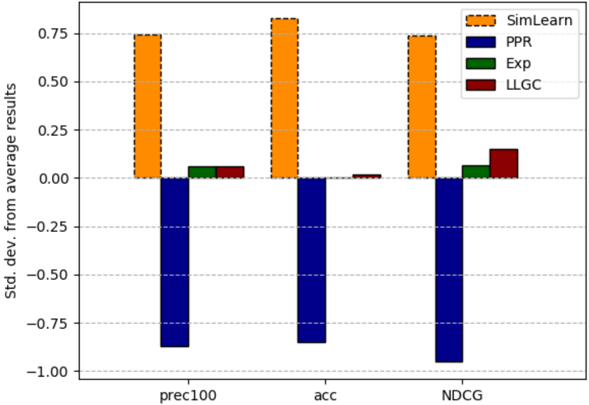


Figure: BlogCatalog

Résultats

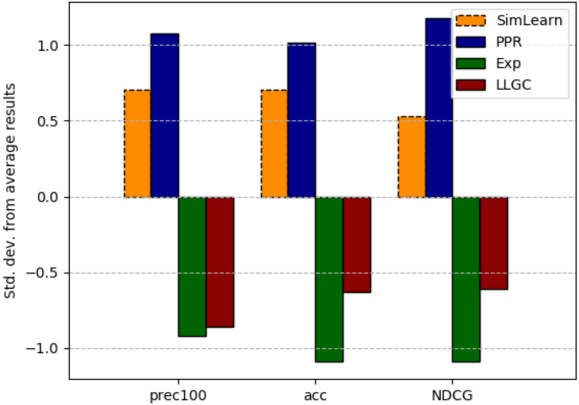


Figure: DBLP

Résultats

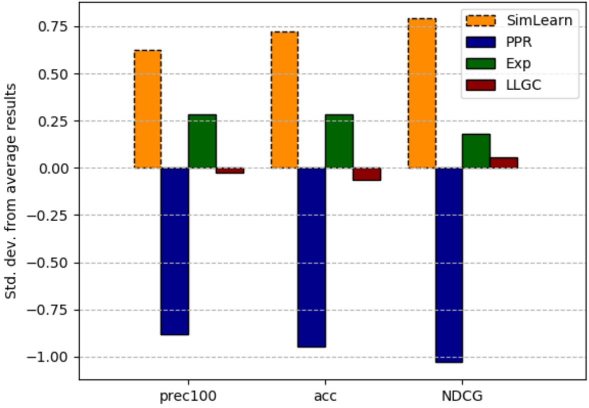


Figure: Flickr

Résultats

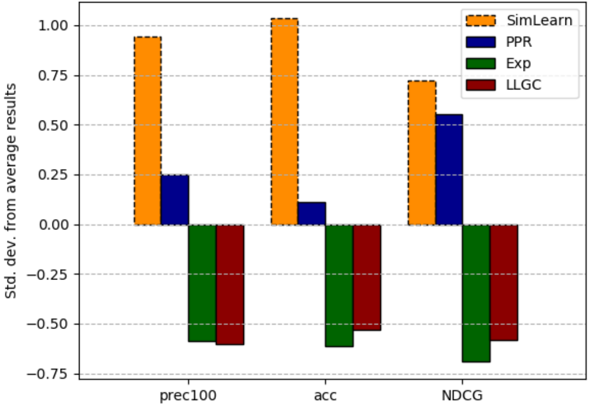
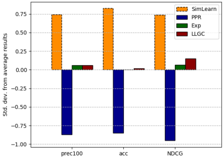
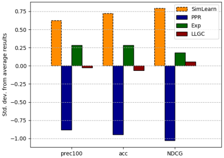


Figure: YouTube

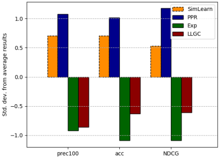
Résultats



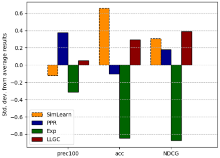
(a) BlogCatalog



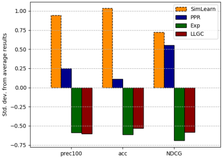
(b) Flickr



(c) DBLP



(d) Amazon



(e) YouTube

1. Introduction

2. Graphes et similarités

3. Apprentissage de similarités

4. Résultats de l'implémentation

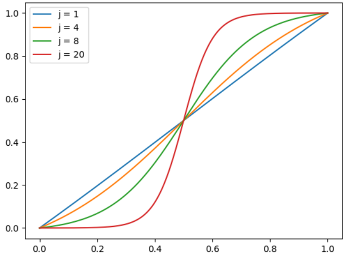
5. Perspectives d'améliorations
   - Renforcement
   - Non-linéarités
   - Propagation mixte

Renforcement

- Améliorer la pénalité $||\text{rows}_{\bar{\mathcal{K}}}(\mathbf{F}(\mathbf{y}))\mathbf{c}||^2$
- Retirer le top50 du ranking
- Résultats mitigés

Non-linéarités

$$\mathbf{F}_{\text{tot}}(\mathbf{y}) = (\mathbf{F}(\mathbf{y}), \mathbf{F}_{\text{nl}}(\mathbf{y}))$$



$$\mathbf{F}_{\text{nl}}(\mathbf{y}) = \Big((\mathbf{f} \circ \mathbf{P}^T)(\mathbf{y}),\ \cdots,\ (\mathbf{f} \circ \mathbf{P}^T)^k \mathbf{y},\ (\mathbf{f} \circ \tilde{\mathbf{A}})(\mathbf{y}),\ \cdots,\ (\mathbf{f} \circ \tilde{\mathbf{A}})^k(\mathbf{y})\Big)$$

- Régression de la précision sur tous les datasets

Propagation mixte

- Propagation des labels selon $\tilde{\mathbf{A}}$ et $\mathbf{P}^T$ séparée
- Ajout de termes croisés $\mathbf{P}^T\tilde{\mathbf{A}}^2$, $(\mathbf{P}^T)^2\tilde{\mathbf{A}}^2$, etc...
- Résultats encourageants

Propagation mixte

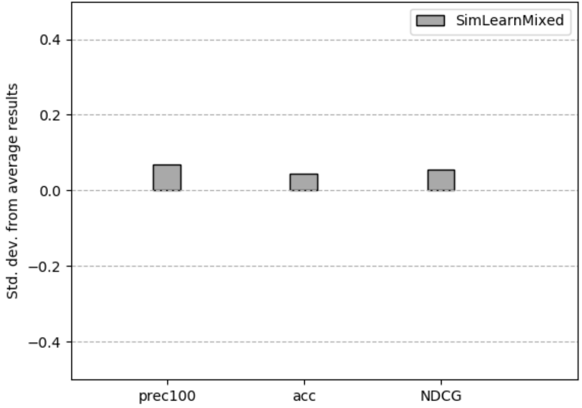


Figure: Amazon

Propagation mixte

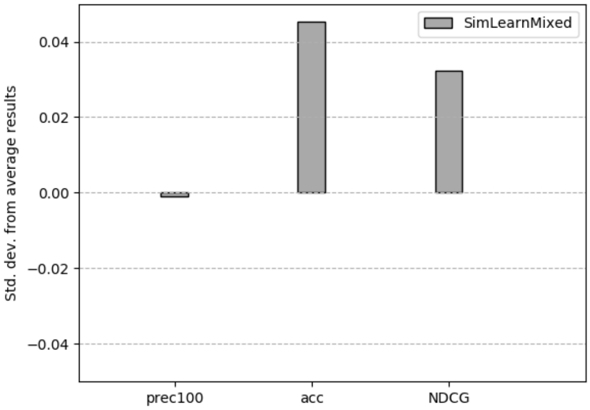


Figure: BlogCatalog

Propagation mixte

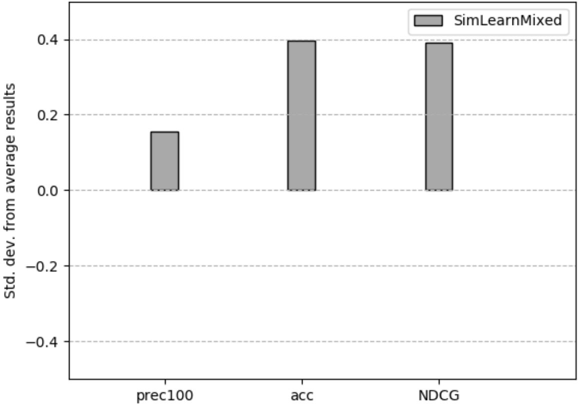


Figure: DBLP

Propagation mixte

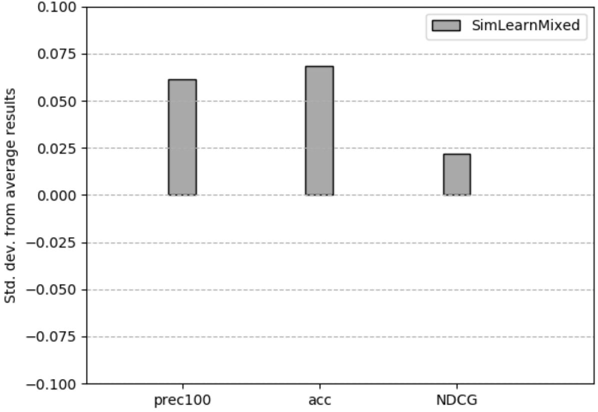


Figure: Flickr

Conclusion

- L'implémentation donne de très bons résultats
- L'addition de termes croisés améliore sensiblement les résultats sur tous les datasets testés
- Il serait intéressant de poursuivre cela

Merci pour votre attention!

References I

[1] https://skymind.ai/images/wiki/graph1.jpg.

[2] Jia-Yu Pan et al. "Automatic Multimedia Cross-modal Correlation Discovery". In: *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '04. Seattle, WA, USA: ACM, 2004, pp. 653–658. ISBN: 1-58113-888-1. DOI: 10.1145/1014052.1014135.

[3] Risi Imre Kondor and John D. Lafferty. "Diffusion Kernels on Graphs and Other Discrete Input Spaces". In: *Proceedings of the Nineteenth International Conference on Machine Learning*. ICML '02. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2002, pp. 315–322. ISBN: 1-55860-873-7.

[4] Dengyong Zhou et al. "Learning with Local and Global Consistency". In: *Proceedings of the 16th International Conference on Neural Information Processing Systems*. NIPS'03. Whistler, British Columbia, Canada: MIT Press, 2003, pp. 321–328.

[5] Robin Devooght. "Similarity measures on graphs and novel methods for collaborative filtering". PhD thesis. Université Libre de Bruxelles, 2017.

[6] François Fouss et al. "An Experimental Investigation of Kernels on Graphs for Collaborative Recommendation and Semisupervised Classification". In: *Neural Netw.* 31 (July 2012), pp. 53–72. ISSN: 0893-6080. DOI: 10.1016/j.neunet.2012.03.001.

Non-linéarités

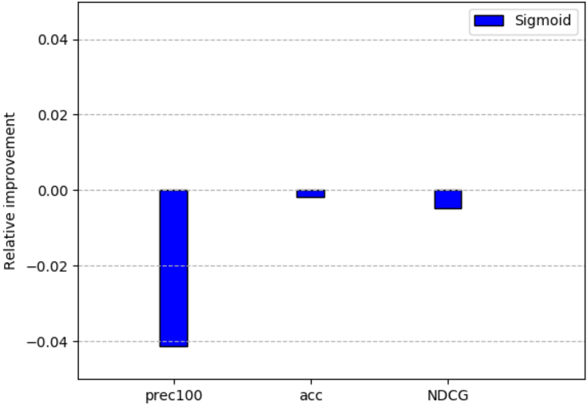


Figure: Amazon

Appendix A

Non-linéarités

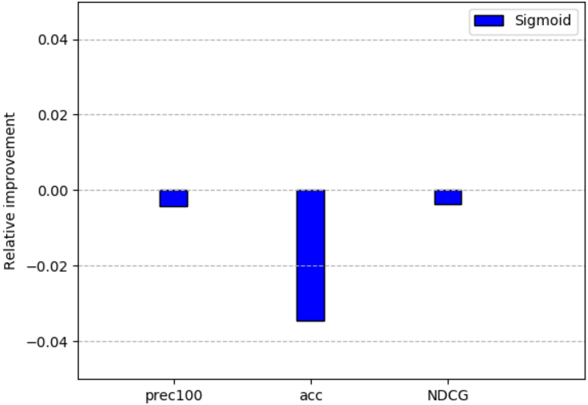


Figure: BlogCatalog

Appendix A

Non-linéarités

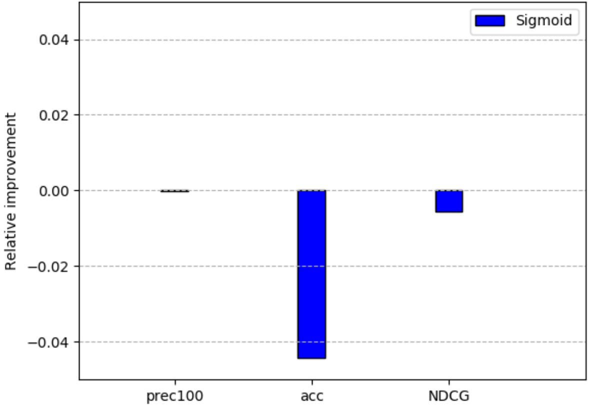


Figure: DBLP

Appendix A

Renf

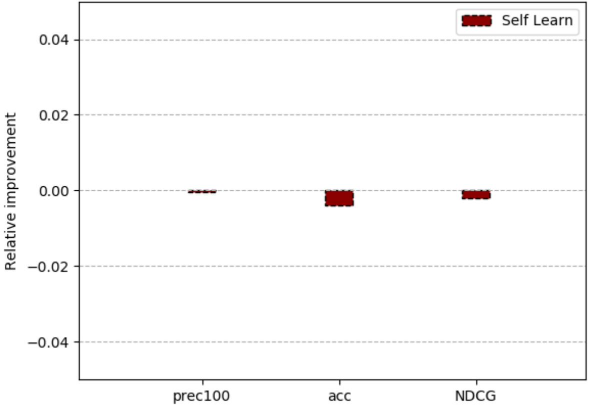


Figure: Amazon

Appendix A

Renf

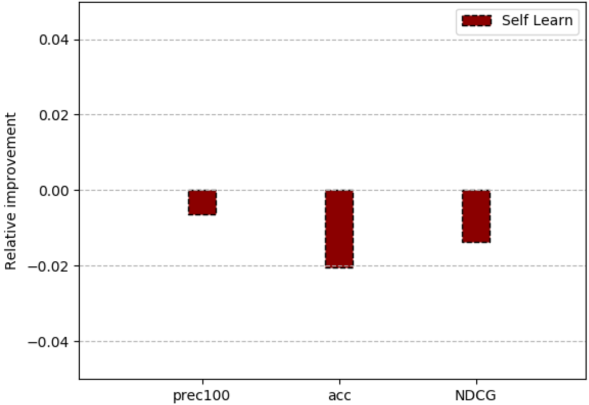


Figure: BlogCatalog

Appendix A

Renf

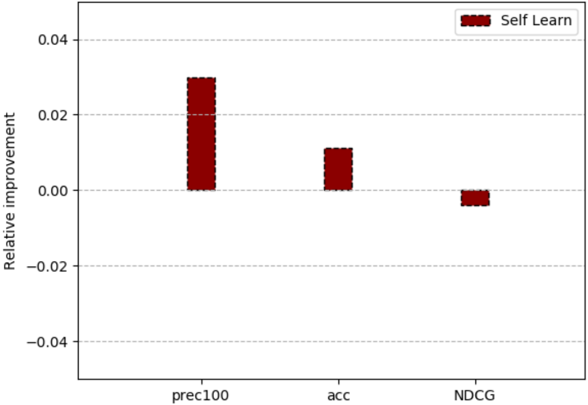


Figure: DBLP

Appendix A

Renf

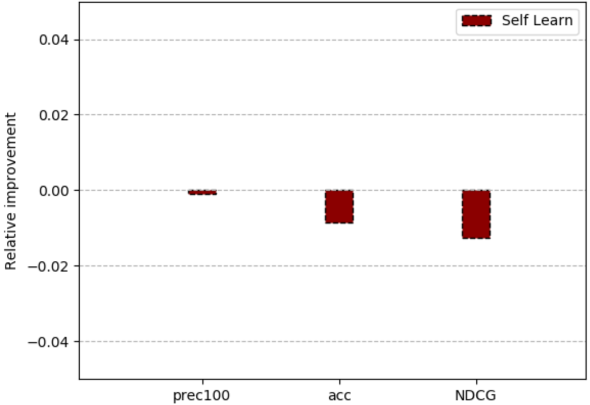


Figure: Flickr

Appendix A