

최신 네트워크 정렬 방법에 대한 서베이 및 평가

임재환¹, 서동혁², 김상욱³

¹한양대학교 컴퓨터소프트웨어학과 석사과정

²한양대학교 컴퓨터소프트웨어학과 박사과정

³한양대학교 컴퓨터소프트웨어학과 교수

gowjr235@hanyang.ac.kr, hyuk125@hanyang.ac.kr, wook@hanyang.ac.kr

Surveys and Evaluation of Recent Network Alignment Methods

Jae-Hwan Lim¹, Dong-Hyuk Seo², Sang-Wook Kim³

^{1,2,3}Dept. of Computer Science, Hanyang University

요 약

최근 온라인 소셜 네트워크 플랫폼의 증가에 따라 사용자들은 다양한 서비스를 제공받기 위해 여러 소셜 네트워크 플랫폼에 가입하는 경향이 있다. 네트워크 정렬은 보안상의 문제로 사용자들의 개인정보가 제한된 상황에서 네트워크의 구조와 속성 정보를 이용하여 서로 다른 소셜 플랫폼에서 동일한 사용자를 찾는 것을 목표로 한다. 본 논문은 최근 몇 년간의 네트워크 정렬 연구들을 서베이 하고, 그들을 분류한 후, 그 중 대표적인 것들에 대한 성능 평가를 수행한다.

1. 서론

소셜 네트워크의 관계 정보를 활용하여 맞춤형 개인화 추천을 수행하는 것은 추천 시스템 분야의 중요한 연구 분야이다. 이때 소셜 네트워크의 관계 정보는 주로 그래프로 표현되며, 이에 따른 다양한 그래프 알고리즘들이 활용된다. 최근에는 다양한 소셜 네트워크들이 등장함에 따라 각 소셜 네트워크에서 얻을 수 있는 관계 정보가 다양해져 추천의 범위도 다양해지고 있다.

그러나, 이러한 소셜 네트워크에서 얻을 수 있는 데이터는 보안상의 문제로 인해 개개인의 유저를 특정할 수 없는 경우가 많다. 이에 서로 다른 소셜 네트워크에서 같은 사람을 찾아 일관된 추천 등을 수행하는 데에 어려움이 발생한다.

네트워크 정렬은 이렇게 그래프로 표현된 서로 다른 소셜 네트워크 상에서 여러가지 부가정보를 활용하여 같은 사람을 찾아내는 것을 목표로 한다[1]. 이때 활용되는 부가정보로서 여러가지 속성 정보 (예: 이름, 나이, 성별)를 활용한다.

이와 같은 방향성으로서 최근 다양한 네트워크 정렬 방법들이 제안되고 있다. 제안되는 방법들을 살펴보면 여러가지 기준을 통해 그 방법들을 구별할 수 있다. (1) 사전 정렬 정보의 사용 유무: 일부 유저의 정렬 정보를 알고 있어 이를 활용하는 방법들과, 전혀 사전 정보가 없는 상황에서 정렬하려 하는 방법들이 있다. (2) 속성 정보 활용 유무: 유저에 대한 부가

적인 정보가 주어지 이를 활용하는 방법들과, 부가적인 정보가 없는 상황에서 정렬하려 하는 방법들이 있다. (3) 네트워크 임베딩 방법: 최근 네트워크 연구 분야에서 다양한 네트워크 임베딩 방법들이 연구되었으며(DeepWalk, GNN, Contrastive learning 등) 네트워크 정렬 방법 또한 어떠한 임베딩 방법을 선택했는지에 따라 구별을 할 수 있다. (4) 네트워크 간 비교 방법: 네트워크 정렬은 두 개의 다른 네트워크를 대상으로 노드를 비교하기 때문에 매핑 함수, 가중치를 공유하는 GNN, 대조 학습을 통해 두 개의 네트워크에 존재하는 노드를 효과적으로 비교한다.

본 논문은 기존의 네트워크 정렬 방법들을 종합적으로 분석하고 분류하여 정리 및 비교한 논문이다. <표 1>은 기존 네트워크 정렬 방법들을 시간 순으로 나열하여 다양한 분류 기준으로 분류한 결과를 보여주는 표이다. 우리는 여섯 개의 대표적인 네트워크 정렬 방법(FINAL[1], PALE[2], DeepLink[3], CENALP[4], GAlign[5], Grad-Align[9])을 선택하여 세 가지의 실세계 네트워크에 대한 실험을 통한 성능을 비교 분석하였다.

2. 네트워크 정렬 방법들

FINAL[1]은 초기 네트워크 정렬 방법들 중 하나로, 사전 정렬 정보를 사용하며 노드의 속성 정보 또한 사용하는 방법이다. 이웃 노드들의 속성 정보가 비슷

<표 1> 최근 네트워크 정렬 방법의 분류

방법	사전정렬정보	속성 정보	DeepWalk	GNN	매핑 함수	가중치 공유 GNN	대조 학습
FINAL'16 [1]	O	O					
PALE'16 [2]	O		O		O		
DeepLink'18 [3]	O		O		O		
CENALP'19 [4]	O		O				
GAlign'20 [5]		O		O		O	
WAlign'21 [6]		O		O	O		
HTC'22 [7]				O		O	
CPUM'23 [8]	O	Optional		O			O
Grad-Align'23 [9]	O	O		O		O	
CANA'23 [10]	O	O		O		O	

한 대상 노드는 같은 노드일 것 이라는 전제하에 네트워크 정렬 방법을 연구하였으며 대상 노드의 이웃 노드들의 속성 정보를 사용하여 서로 다른 네트워크에서 같은 노드를 찾는다.

PALE[2]은 사전 정렬 정보를 사용하지만, 노드 속성 정보를 사용하지 않은 방법으로, DeepWalk 를 기반으로 두 네트워크의 노드들을 각각 저차원의 벡터로 표현하여 네트워크 정렬 방법을 연구하였다. 두 네트워크 간의 비교를 위한 매핑 함수(Φ)를 고안해낸 방법이다 (예: $u_i \in G^s, u_j \in G^t$ 일때, $\Phi(u_i) = u_j$). 매핑 함수 학습을 위해 사전 정렬 정보를 사용하였으며, 학습된 매핑 함수를 정렬되어 있지 않은 노드에 적용시켜 네트워크 정렬 성능을 향상시켰다.

DeepLink[3] 또한 사전 정렬 정보를 사용하며 노드 속성 정보를 사용하지 않는 방법이다. DeepWalk 기반 네트워크 임베딩 방법을 택하였으며, PALE[2]과는 다르게 두 네트워크 간 두 개의 매핑 함수를 사용함으로써 (예: $u_i \in G^s, u_j \in G^t$ 일때, $\Phi(u_i) = u_j, \Phi^{-1}(u_j) = u_i$) Auto Encoder 방식의 학습을 고안해내어 저차원으로 표현된 두 네트워크 간의 관계를 밝혀냈다.

CENALP[4] 또한 사전 정렬 정보를 사용하며 노드 속성 정보는 사용하지 않는 방법이다. DeepWalk 기반 네트워크 임베딩 방법을 택하였지만, 앞서 언급한 PALE[2], DeepLink[3]와 다르게 두 네트워크 간의 연결성을 정의하여 두 네트워크를 각각의 저차원 벡터 공간에 표현하는 것이 아닌 하나의 저차원 벡터 공간에 표현하였다.

GAlign[5]은 사전 정렬 정보를 필요로 하지 않으며 노드 속성 정보를 사용하는 방법으로, 처음으로 GNN 기반의 네트워크 임베딩 방법을 사용하여 네트워크 정렬 문제를 해결한 방법이다. 두 네트워크를 임베딩 하는데 사용된 GNN 레이어의 가중치 행렬을 공유하여 두 네트워크를 하나의 임베딩 공간으로 표현하였으며 두 노드의 벡터간의 코사인 유사도를 계산하여 네트워크 정렬 문제를 해결하였다.

GradAlign[9]은 사전 정렬 정보를 사용하며 노드 속성 정보는 사용하지 않는 방법이다. 두 네트워크 간 가중치 행렬을 공유하는 GNN 기반의 임베딩을 사

용하여 네트워크를 저차원의 공간으로 표현하였으며, 두 개의 네트워크의 크기 차이로 인해 발생하는 오차를 줄이기 위해서 Tversky similarity 이론을 적용하여 네트워크 정렬 문제를 해결하였다.

3. 실험

실험 환경. 우리는 3 가지 실제 네트워크를 반영한 데이터 집합들을 사용하여 실험을 수행한다. Douban online-offline, Facebook-Twitter, Allmovie-IMDb. <표 2>은 본 논문에서 사용한 데이터 집합들의 통계를 보여준다.

<표 2> 데이터 집합에 대한 통계

데이터 집합	노드	간선	속성정보	사전 정렬 노드
Douban Online	3,906	8,164	538	1,118
Douban Offline	1,118	1,511	538	
Facebook	1,043	4,734	0	1,043
Twitter	1,043	4,860	0	
Allmovie	6,011	124,709	14	5,176
IMDb	5,713	119,073	14	

경쟁 방법. 우리는 공정한 비교를 위해, 노드 속성 정보가 없는 상황에서의 6 가지 네트워크 정렬 방법들의 성능을 분석한다. 그 다음으로는, 노드 속성 정보가 있는 환경에서의 FINAL, GAlign 그리고 Grad-Align 방법을 비교한다.

실험 방법. 우리는 최근 네트워크 정렬 방법들의 성능을 분석하기 위해 각 데이터 집합마다 존재하는 사전 정렬 노드의 수의 10%를 트레이닝 집합으로 사용하고, 나머지를 테스트 집합으로 사용하였다.

성능 측정 방법. 우리는 hit@k 평가지표를 사용하여 6 가지 네트워크 정렬 방법들을 평가한다. hit@k 는 테스트 집합으로 사용된 사전 정렬 노드들 중 실제로 대응되는 노드를 정확히 예측한 비율을 측정한다.

실험 결과. <표 3>는 기존 네트워크 정렬 방법들의 속성 정보가 주어지지 않은 상황에서의 hit@k 값을 보여준다. <표 3>를 통해 우리는 현재 GNN 기반 네트워크 임베딩을 활용하는 네트워크 정렬 방법들이[5, 9] 우수한 성능을 보여주고 있는 경향이 있다. 또한, Douban online-offline 과 같이 두 네트워크 간의 크기

차이가 많이 나는 네트워크에서의 모든 네트워크 정렬 방법의 성능이 facebook-twitter 와 같이 네트워크 크기 차이가 많이 나지 않는 그래프보다 전반적으로 떨어지는 경향을 보인다. <표 4>는 노드 속성 정보가 추가적으로 주어진 상황에서의 기존 네트워크 정렬 방법의 성능을 비교한 것으로, 직접적으로 노드 속성 정보를 비교하는 방법보다[1] GNN 기반 네트워크 임베딩을 활용한 네트워크 정렬 방법들이 더욱 우수한 성능을 보이는 경향을 확인할 수 있다.

4. 결론

본 논문에서, 우리는 기존 네트워크 정렬 연구들을 종합적으로 분석하고 분류하였다. 초기 단계의 노드 속성 정보와 주변 노드 정보를 직접적으로 사용하는 방법보다 DeepWalk 기반 노드 임베딩 방법을 사용하는 방법이 우수하며, 노드 속성 정보를 사용하는 GNN 기반 네트워크 정렬 방법들은 더욱 우수한 성능을 보이고 있다.

<표 3> 기존 네트워크 정렬 방법의 성능 비교

데이터	평가지표	FINAL	PALE	DeepLink
Douban	hit@1	0.0072	0.1377	0.0921
	hit@5	0.0358	0.3283	0.2290
	hit@10	0.0769	0.4338	0.2907
Facebook Twitter	hit@1	0.0019	0.5295	0.1151
	hit@5	0.0144	0.7854	0.2927
	hit@10	0.0259	0.8803	0.3350
Allmovie IMDb	hit@1	0.0019	0.5793	0.1361
	hit@5	0.0138	0.8242	0.3327
	hit@10	0.0193	0.8519	0.4650
데이터	평가지표	CENALP	GAlign	GradAlign
Douban	hit@1	0.0235	0.4395	0.3292
	hit@5	0.0571	0.6798	0.4902
	hit@10	0.1130	0.7862	0.5707
Facebook Twitter	hit@1	0.9069	0.1083	0.9501
	hit@5	0.9309	0.2464	0.9781
	hit@10	0.9357	0.3164	0.9862
Allmovie IMDb	hit@1	0.4238	0.5381	0.8316
	hit@5	0.5721	0.7246	0.9101
	hit@10	0.7154	0.7837	0.9308

<표 4>노드 속성 정보가 주어진 상황에서의 기존 네트워크 정렬 방법의 성능 비교

데이터	평가지표	FINAL	GAlign	GradAlign
Douban	hit@1	0.3479	0.3686	0.6682
	hit@5	0.5134	0.5233	0.8551
	hit@10	0.6764	0.6324	0.8980
Facebook Twitter	hit@1	0.6007	0.6306	0.9674
	hit@5	0.7864	0.8422	0.9856
	hit@10	0.7773	0.8612	0.9904
Allmovie IMDb	hit@1	0.6524	0.7251	0.9601
	hit@5	0.8594	0.8101	0.9744
	hit@10	0.8945	0.8749	0.9783

사사

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2022-0-00352 IITP 방송통신산업기술개발사업, No.2018R1A5A7059549 한국연구재단 CRC, No. RS-2022-00155586 실세계의 다양한 다운스트림 태스크를 위한 고성능 빅하이퍼그래프 마이닝 플랫폼 개발(SW 스타랩))

참고문헌

- [1] S. Zhang and H. Tong, "FINAL: Fast attributed network alignment," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining (KDD'16), San Francisco, CA, Aug. 2016, pp. 1345–1354.
- [2] T. Man, H. Shen, S. Liu, X. Jin, and X. Cheng, "Predict anchor links across social networks via an embedding approach." in Proc. 25th Int. Joint Conf. Artif. Intell. (IJCAI'16), vol. 16, New York City, NY, Jul. 2016, pp. 1823–1829
- [3] F. Zhou, L. Liu, K. Zhang, G. Trajcevski, J. Wu, and T. Zhong, "DeepLink: A deep learning approach for user identity linkage," in Proc. 37th IEEE Conf. Comput. Commun. (INFOCOM'18), Honolulu, HI, Apr. 2018, pp. 1313–1321.
- [4] X. Du, J. Yan, and H. Zha, "Joint link prediction and network alignment via cross-graph embedding." in Proc. 28th Int. Joint Conf. Artif. Intell. (IJCAI'19), Macao, China, Aug. 2019, pp. 2251–2257
- [5] H. T. Trung, T. Van Vinh, N. T. Tam, H. Yin, M. Weidlich, and N. Q. V. Hung, "Adaptive network alignment with unsupervised and multi-order convolutional networks," in Proc. 36th Int. Conf. Data Eng. (ICDE'20), Dallas, TX, Apr. 2020, pp. 85–96.
- [6] J. Gao, X. Huang, and J. Li, "Unsupervised graph alignment with wasserstein distance discriminator," in The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Singapore, August 14-18, 2021, 2021, pp. 426–435.
- [7] Q. Sun, X. Lin, Y. Zhang, W. Zhang, and C. Chen, "Towards higherorder topological consistency for unsupervised network alignment," arXiv preprint arXiv:2208.12463, 2022
- [8] W. Tang, H. Sun, J. Wang, C. Liu, Q. Qi, J. Wang, and J. Liao, "Identifying users across social media networks for interpretable fine-grained neighborhood matching by adaptive gat," IEEE Transactions on Services Computing, to appear. DOI: 10.1109/TSC.2023.3288872
- [9] Jin-Duk Park, Cong Tran, Won-Yong Shin, and Xin Cao. 2022. Grad-Align: Gradual Network Alignment via Graph Neural Networks (Student Abstract). In AAAI. AAAI Press, 13027–13028.
- [10] Shao J, Wang Y, Guo F, Shi B, Shen H, Cheng X. CANA: Causal-enhanced Social Network Alignment. InProceedings of the 32nd ACM International Conference on Information and Knowledge Management 2023 Oct 21 (pp. 2219-2228).