

Tosyali, Ali; Kim, Jinho; Choi, Jeongsub; Kang, Yunyi; Jeong, Myong K.

New node anomaly detection algorithm based on nonnegative matrix factorization for directed citation networks. (English) Zbl 1436.90031

Ann. Oper. Res. 288, No. 1, 457-474 (2020).

Summary: Outlier detection is a crucial task for network data analysis, which identifies abnormal entities that deviate from the rest of the dataset. Ranking in outlierness is often used for identifying abnormal nodes in directed citation networks containing citation relationship among nodes. A challenging issue in outlier ranking is how to leverage the rich graph data of complex citation networks. In this paper, we propose a cluster-based outlier score function to identify outliers in citation networks based on nonnegative matrix factorization (NMF). We first represent the citation data as a directed graph, and cluster the directed graph into logical groupings of nodes using NMF. Based on the clustering results, we obtain the outlier score and ranking for each node using the proposed outlier scoring function. The proposed method leverages the direct and indirect citation links between nodes to measure the graph-based outlierness. We validate the proposed outlier ranking method using small artificial dataset and the real-world U.S. patent data.

MSC:

90B10 Deterministic network models in operations research

Keywords:

[anomaly detection](#); [directed graph](#); [nonnegative matrix factorization](#); [patent citation network](#)

Software:

[gSkeletonClu](#)

Full Text: [DOI](#)

References:

- [1] Agreste, S.; De Meo, P.; Ferrara, E.; Piccolo, S.; Proveti, A., Analysis of a heterogeneous social network of humans and cultural objects, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45, 4, 559-570 (2015)
- [2] Akoglu, L., McGlohon, M., & Faloutsos, C. (2010). Oddball: Spotting anomalies in weighted graphs. In *Pacific-Asia conference on knowledge discovery and data Mining* (pp. 410-421). Berlin: Springer
- [3] Banker, RD; Chang, H.; Zheng, Z., On the use of super-efficiency procedures for ranking efficient units and identifying outliers, *Annals of Operations Research*, 250, 1, 21-35 (2017) · [Zbl 1364.90133](#)
- [4] Boutsidis, C.; Gallopoulos, E., SVD based initialization: A head start for nonnegative matrix factorization, *Pattern Recognition*, 41, 4, 1350-1362 (2008) · [Zbl 1131.68084](#)
- [5] Cao, X.; Wang, X.; Jin, D.; Cao, Y.; He, D., Identifying overlapping communities as well as hubs and outliers via nonnegative matrix factorization, *Scientific Reports*, 3, 2993 (2013)
- [6] Codetta-Raiteri, D.; Portinale, L., Dynamic bayesian networks for fault detection, identification, and recovery in autonomous spacecraft, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45, 1, 13-24 (2015)
- [7] Ding, CH; He, X.; Simon, HD, On the equivalence of nonnegative matrix factorization and spectral clustering, *SDM, SIAM*, 5, 606-610 (2005)
- [8] Duan, L.; Xu, L.; Liu, Y.; Lee, J., Cluster-based outlier detection, *Annals of Operations Research*, 168, 1, 151-168 (2009) · [Zbl 1183.68212](#)
- [9] Džamić, D.; Aloise, D.; Mladenović, N., Ascent-descent variable neighborhood decomposition search for community detection by modularity maximization, *Annals of Operations Research*, 272, 273-287 (2017) · [Zbl 1411.90272](#)
- [10] Holder, LB; Cook, DJ, Graph-based data mining, *Encyclopedia of data warehousing and mining*, 2, 943-949 (2009)
- [11] Kaffash, S.; Marra, M., Data envelopment analysis in financial services: A citations network analysis of banks, insurance companies and money market funds, *Annals of Operations Research*, 253, 1, 307-344 (2017) · [Zbl 1406.91001](#)
- [12] Kang, U.; Akoglu, L.; Chau, DHP, Big graph mining: Algorithms, anomaly detection, and applications, *Proceedings of the ACM ASONAM*, 13, 25-28 (2013)

- [13] Lee, D. D., \& Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems* (pp. 556-562).
- [14] Lu, N.; Li, T.; Pan, J.; Ren, X.; Feng, Z.; Miao, H., Structure constrained semi-nonnegative matrix factorization for EEG-based motor imagery classification, *Computers in Biology and Medicine*, 60, 32-39 (2015)
- [15] Ma, Y.; Hu, X.; He, T.; Jiang, X., Hessian regularization based symmetric nonnegative matrix factorization for clustering gene expression and microbiome data, *Methods*, 111, 80-84 (2016)
- [16] Michel, J.; Bettels, B., Patent citation analysis. A closer look at the basic input data from patent search reports, *Scientometrics*, 51, 1, 185-201 (2001)
- [17] Moonesinghe, H., \& Tan, P. N. (2006). Outlier detection using random walks. In *2006 18th IEEE international conference on tools with artificial intelligence (ICTAI'06)*, IEEE (pp. 532-539).
- [18] Newman, M., *Networks: An introduction* (2010), Oxford: Oxford University Press, Oxford
- [19] Sun, H., Huang, J., Han, J., Deng, H., Zhao, P., \& Feng, B. (2010). gskeletonclu: Density-based network clustering via structure-connected tree division or agglomeration. In *2010 IEEE International Conference on Data Mining, IEEE* (pp. 481-490).
- [20] Tong, H., \& Lin, C. Y. (2011). Non-negative residual matrix factorization with application to graph anomaly detection. In *SDM, SIAM* (pp. 143-153).
- [21] Tosyali, A.; Kim, J.; Choi, J.; Jeong, MK, Regularized asymmetric nonnegative matrix factorization for clustering in directed networks, *Pattern Recognition Letters*, 125, 750-757 (2019)
- [22] Wang, F.; Li, T.; Wang, X.; Zhu, S.; Ding, C., Community discovery using nonnegative matrix factorization, *Data Mining and Knowledge Discovery*, 22, 3, 493-521 (2011) · [Zbl 1235.68034](#)
- [23] Xu, X., Yuruk, N., Feng, Z., \& Schweiger, T. A. (2007). Scan: A structural clustering algorithm for networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM* (pp. 824-833).
- [24] Yoon, J.; Kim, K., Detecting signals of new technological opportunities using semantic patent analysis and outlier detection, *Scientometrics*, 90, 2, 445-461 (2011)
- [25] Yuan, X.; Guo, J.; Hao, X.; Chen, H., Traffic sign detection via graph-based ranking and segmentation algorithms, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45, 12, 1509-1521 (2015)
- [26] Zhi, R.; Flierl, M.; Ruan, Q.; Kleijn, WB, Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41, 1, 38-52 (2011)
- [27] Zou, Z.; Li, J.; Gao, H.; Zhang, S., Mining frequent subgraph patterns from uncertain graph data, *IEEE Transactions on Knowledge and Data Engineering*, 22, 9, 1203-1218 (2010)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.