Wang, Chenxu; **Wang, Yang**; **Zhao, Zhiyuan**; **Qin, Dong**; **Luo, Xiapu**; **Qin, Tao**
Credible seed identification for large-scale structural network alignment. (English)
Zbl 1455.68149
Data Min. Knowl. Discov. 34, No. 6, 1744-1776 (2020).

Summary: Structural network alignment utilizes the topological structure information to find correspondences between nodes of two networks. Researchers have proposed a line of useful algorithms which usually require a prior mapping of seeds acting as landmark points to align the rest nodes. Several seed-free algorithms are developed to solve the cold-start problem. However, existing approaches suffer high computational cost and low reliability, limiting their applications to large-scale network alignment. Moreover, there is a lack of useful metrics to quantify the credibility of seed mappings. To address these issues, we propose a credible seed identification framework and develop a metric to assess the reliability of a mapping. To tackle the cold-start problem, we employ graph embedding techniques to represent nodes by structural feature vectors in a latent space. We then leverage point set registration algorithms to match nodes algebraically and obtain an initial mapping of nodes. Besides, we propose a heuristic algorithm to improve the credibility of the initial mapping by filtering out mismatched node pairs. To tackle the computational problem in large-scale network alignment, we propose a divide-and-conquer scheme to divide large networks into smaller ones and then match them individually. It significantly improves the recall of mapping results. Finally, we conduct extensive experiments to evaluate the effectiveness and efficiency of our new approach. The results illustrate that the proposed method outperforms the state-of-the-art approaches in terms of both effectiveness and efficiency.

MSC:

68R10 Graph theory (including graph drawing) in computer science
68T05 Learning and adaptive systems in artificial intelligence
68T20 Problem solving in the context of artificial intelligence (heuristics, search strategies, etc.)

Keywords:

network alignment; seed identification; edge consistency; mapping credibility

Software:

L-GRAAL; MAGNA++; node2vec; SANA; SecGraph; SPINAL; struc2vec

Full Text: DOI

References:

[1] Aladağ, AE; Erten, C., Spinal: scalable protein interaction network alignment, Bioinformatics, 29, 7, 917-924 (2013)

[2] Backstrom L, Dwork C, Kleinberg J (2007) Wherefore art thou r3579x? Anonymized social networks, hidden patterns, and structural steganography. In: International world wide web conference. ACM, pp 181-190

[3] Bayati, M.; Gleich, DF; Saberi, A.; Wang, Y., Message-passing algorithms for sparse network alignment, ACM TKDD, 7, 1, 1-31 (2013)

[4] Bayati M, Gerritsen M, Gleich DF, Saberi A, Wang Y (2009) Algorithms for large, sparse network alignment problems. In: International conference on data mining (ICDM). IEEE, pp 705-710

[5] Blondel, VD; Guillaume, JL; Lambiotte, R.; Lefebvre, E., Fast unfolding of communities in large networks, J Stat Mech Theory Exp, 2008, 10, P10008 (2008) · Zbl 1459.91130

[6] Box, GE; Hunter, WG; Hunter, JS, Statistics for experimenters (1978), New York: Wiley, New York

[7] Chen Z, Yu X, Song B, Gao J, Hu X, Yang WS (2017) Community-based network alignment for large attributed network. In: Conference on information and knowledge management (CIKM). ACM, pp 587-596

[8] Chui H, Rangarajan A (2000) A feature registration framework using mixture models. In: IEEE workshop on mathematical methods in biomedical image analysis. IEEE, pp 190-197

[9] Collins, SR; Kemmeren, P.; Zhao, XC; Greenblatt, JF; Spencer, F.; Holstege, FC; Weissman, JS; Krogan, NJ, Toward a comprehensive atlas of the physical interactome of saccharomyces cerevisiae, Mol Cell Proteom, 6, 3, 439-450 (2007)

[10] Conte, D.; Foggia, P.; Sansone, C.; Vento, M., Thirty years of graph matching in pattern recognition, Int J Pattern Recognit Artif Intell, 18, 3, 265-298 (2004)

[11] Du B, Tong H (2018) Fasten: fast Sylvester equation solver for graph mining. In: SIGKDD conference on knowledge discovery and data mining. ACM, pp 1339-1347

[12] Emms, D.; Hancock, ER; Wilson, RC, A correspondence measure for graph matching using the discrete quantum walk, International workshop on graph-based representations in pattern recognition, 81-91 (2007), Berlin: Springer, Berlin · Zbl 1182.68146

[13] Fabiana C, Garetto M, Leonardi E (2015) De-anonymizing scale-free social networks by percolation graph matching. In: IEEE conference on computer communications. IEEE, pp 1571-1579

[14] Fitzgibbon, AW, Robust registration of 2d and 3d point sets, Image Vis Comput, 21, 13, 1145-1153 (2003)

[15] Garey, MR; Johnson, DS, Computers and intractability (2002), New York: WH Freeman, New York

[16] Gold, S.; Rangarajan, A., A graduated assignment algorithm for graph matching, IEEE Trans Pattern Anal Mach Intell, 18, 4, 377-388 (1996)

[17] Gold S, Lu CP, Rangarajan A, Pappu S, Mjolsness E (1995) New algorithms for 2d and 3d point matching: pose estimation and correspondence. In: Annual conference on neural information processing systems, pp 957-964

[18] Grover A, Leskovec J (2016) node2vec: scalable feature learning for networks. In: SIGKDD conference on knowledge discovery and data mining. ACM, pp 855-864

[19] Hashemifar, S.; Ma, J.; Naveed, H.; Canzar, S.; Xu, J., Modulealign: module-based global alignment of protein-protein interaction networks, Bioinformatics, 32, 17, i658-i664 (2016)

[20] Heimann M, Shen H, Koutra D (2018) Node representation learning for multiple networks: the case of graph alignment. arXiv preprint arXiv:1802.06257

[21] Hu, W.; Qu, Y.; Cheng, G., Matching large ontologies: a divide-and-conquer approach, Data Knowl Eng, 67, 1, 140-160 (2008)

[22] Ji S, Li W, Srivatsa M, Beyah R (2014a) Structural data de-anonymization. In: ACM conference on computer and communications security (CCS). ACM, pp 1040-1053

[23] Ji S, Li W, Srivatsa M, He JS, Beyah R (2014b) Structure based data de-anonymization of social networks and mobility traces. In: Information security. Springer, pp 237-254

[24] Ji S, Li W, Mittal P, Hu X, Beyah R (2015) Secgraph: a uniform and open-source evaluation system for graph data anonymization and de-anonymization. In: USENIX security, pp 303-318

[25] Ji S, Li W, Yang S, Mittal P, Beyah R (2016) On the relative de-anonymizability of graph data: quantification and evaluation. In: IEEE conference on computer communications. IEEE, pp 1-9

[26] Klau, GW, A new graph-based method for pairwise global network alignment, Bioinformatics, 10, 1, S59 (2009)

[27] Kollias, G.; Mohammadi, S.; Grama, A., Network similarity decomposition (nsd): a fast and scalable approach to network alignment, IEEE Trans Knowl Data Eng, 24, 12, 2232-2243 (2012)

[28] Korula N, Lattanzi S (2014) An efficient reconciliation algorithm for social networks. In: International conference on very large data bases, VLDB endowment, pp 377-388

[29] Koutra D, Tong H, Lubensky D (2013) Big-align: fast bipartite graph alignment. In: International conference on data mining (ICDM). IEEE, pp 389-398

[30] Kuchaiev, O.; Milenković, T.; Memišević, V.; Hayes, W.; Pržulj, N., Topological network alignment uncovers biological function and phylogeny, J R Soc Interface, 7, 50, 1341-1354 (2010)

[31] Leskovec J, Krevl A (2014) Snap datasets: Stanford large network dataset collection. http://snap.stanford.edu/data. Accessed 10 Apr 2020

[32] Liu L, Cheung WK, Li X, Liao L (2016) Aligning users across social networks using network embedding. In: International joint conferences on artificial intelligence (IJCAI), pp 1774-1780

[33] Luo, B.; Hancock, ER, Structural graph matching using the em algorithm and singular value decomposition, IEEE Trans Pattern Anal Mach Intell, 23, 10, 1120-1136 (2001)

[34] Malod-Dognin, N.; Pržulj, N., L-graal: Lagrangian graphlet-based network aligner, Bioinformatics, 31, 13, 2182-2189 (2015)

[35] Mamano N, Hayes W (2016) Sana: simulated annealing network alignment applied to biological networks. arXiv preprint arXiv:1607.02642

[36] Melnik S, Garcia-Molina H, Rahm E (2002) Similarity flooding: a versatile graph matching algorithm and its application to schema matching. In: International conference on data engineering, vol 2002. IEEE Computer Society, pp 117-128

[37] Myronenko, A.; Song, X., Point set registration: coherent point drift, IEEE Trans Pattern Anal Mach Intell, 32, 12, 2262-2275 (2010)

[38] Narayanan A, Shmatikov V (2009) De-anonymizing social networks. In: IEEE symposium on security and privacy. IEEE, pp 173-187

[39] Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. In: International conference on very large data bases. VLDB endowment, pp 102-114

[40] Nilizadeh S, Kapadia A, Ahn YY (2014) Community-enhanced de-anonymization of online social networks. In: ACM conference on computer and communications security (CCS). ACM, pp 537-548

[41] Pedarsani P, Figueiredo DR, Grossglauser M (2013) A Bayesian method for matching two similar graphs without seeds. In:

Allerton. IEEE, pp 1598-1607

[42] Peng, W.; Li, F.; Zou, X.; Wu, J., A two-stage deanonymization attack against anonymized social networks, IEEE Trans Comput, 63, 2, 290-303 (2014) · Zbl 1364.68075

[43] Perozzi B, Al-Rfou R, Skiena S (2014) Deepwalk: online learning of social representations. In: SIGKDD conference on knowledge discovery and data mining. ACM, pp 701-710

[44] Pfeiffer JJ, Neville J (2011) Methods to determine node centrality and clustering in graphs with uncertain structure. In: International AAAI conference on web and social media (ICWSM), pp 590-593

[45] Pothen A, Simon HD, Liou K (2000) Partitioning sparse matrices with eigenvectors of graphs. In: International conference on very large data bases. VLDB endowment, pp 102-114 · Zbl 0711.65034

[46] Ribeiro LF, Saverese PH, Figueiredo DR (2017) struc2vec: learning node representations from structural identity. In: SIGKDD conference on knowledge discovery and data mining. ACM, pp 385-394

[47] Robles-Kelly A, Hancock ER (2003) Graph matching using spectral seriation. In: International workshop on energy minimization methods in computer vision and pattern recognition. Springer, pp 517-532 · Zbl 1040.68562

[48] Rubinstein, RY; Kroese, DP, Simulation and the Monte Carlo method (2016), New York: Wiley, New York

[49] Rusinkiewicz S, Levoy M (2001) Efficient variants of the ICP algorithm. In: International conference on 3-D digital imaging and modeling. IEEE, pp 145-152

[50] Scott J (2000) Social network analysis: a handbook. In: International conference on very large data bases. VLDB endowment, pp 102-114

[51] Singh R, Xu J, Berger B (2007) Pairwise global alignment of protein interaction networks by matching neighborhood topology. In: Annual international conference on research in computational molecular biology. Springer, pp 16-31

[52] Srivatsa M, Hicks M (2012) Deanonymizing mobility traces: using social network as a side-channel. In: ACM conference on computer and communications security (CCS). ACM, pp 628-637

[53] Tang J, Qu M, Wang M, Zhang M, Yan J, Mei Q (2015) Line: large-scale information network embedding. In: International world wide web conference. ACM, pp 1067-1077

[54] Vijayan, V.; Saraph, V.; Milenković, T., Magna++: maximizing accuracy in global network alignment via both node and edge conservation, Bioinformatics, 31, 14, 2409-2411 (2015)

[55] Wang C, Zhao Z, Wang Y, Qin D, Luo X, Qin T (2018) Deepmatching: a structural seed identification framework for social network alignment. In: International conference on distributed computing systems, pp 459-470

[56] Wondracek G, Holz T, Kirda E, Kruegel C (2010) A practical attack to de-anonymize social network users. In: IEEE symposium on security and privacy. IEEE, pp 223-238

[57] Yang, Q.; Sze, SH, Path matching and graph matching in biological networks, J Comput Biol, 14, 1, 56-67 (2007)

[58] Yartseva L, Grossglauser M (2013) On the performance of percolation graph matching. In: ACM conference on Online social networks. ACM, pp 119-130

[59] Zaslavskiy, M.; Bach, F.; Vert, JP, Global alignment of protein-protein interaction networks by graph matching methods, Bioinformatics, 25, 12, i259-1267 (2009)

[60] Zhang, Z., Iterative point matching for registration of free-form curves and surfaces, Int J Comput Vis, 13, 2, 119-152 (1994)

[61] Zhang J, Philip SY (2015) Integrated anchor and social link predictions across social networks. In: Twenty-fourth international joint conference on artificial intelligence

[62] Zhang J, Yu PS (2015) Multiple anonymized social network alignment. In: International conference on data mining (ICDM). IEEE, pp 599-608

[63] Zhang S, Tong H (2016) Final: fast attributed network alignment. In: SIGKDD conference on knowledge discovery and data mining. ACM, pp 1345-1354

[64] Zhang Y, Tang J, Yang Z, Pei J, Yu PS (2015) Cosnet: connecting heterogeneous social networks with local and global consistency. In: SIGKDD conference on knowledge discovery and data mining. ACM, pp 1485-1494

[65] Zhang S, Tong H, Tang J, Xu J, Fan W (2017) ineat: incomplete network alignment. In: International conference on data mining (ICDM). IEEE, pp 1189-1194

[66] Zhang S, Tong H, Maciejewski R, Eliassi-Rad T (2019) Multilevel network alignment. In: International world wide web conference. ACM, pp 2344-2354

[67] Zhou, X.; Liang, X.; Zhang, H.; Ma, Y., Cross-platform identification of anonymous identical users in multiple social media networks, IEEE Trans Knowl Data Eng, 28, 2, 411-424 (2016)