



Universiteit
Leiden
The Netherlands

Exploring graph-based clustering and outlier detection algorithms

Li, J.

Citation

Li, J. (2025, November 12). *Exploring graph-based clustering and outlier detection algorithms*. Retrieved from <https://hdl.handle.net/1887/4282945>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4282945>

Note: To cite this publication please use the final published version (if applicable).

References

- [1] Charu C. Aggarwal and Saket Sathe. Theoretical Foundations and Algorithms for Outlier Ensembles. *ACM SIGKDD Explorations Newsletter*, 17(1):24–47, 2015.
- [2] Charu C Aggarwal and Philip S Yu. Outlier detection for high dimensional data. In *Proceedings of the 2001 ACM SIGMOD international conference on Management of data*, pages 37–46, 2001.
- [3] Zeeshan Ahmad, Adnan Shahid Khan, Cheah Wai Shiang, Johari Abdullah, and Farhan Ahmad. Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies*, 32(1):e4150, 2021.
- [4] Daichi Amagata, Makoto Onizuka, and Takahiro Hara. Fast and exact outlier detection in metric spaces: a proximity graph-based approach. In *Proceedings of the 2021 International Conference on Management of Data*, pages 36–48, 2021.
- [5] Daichi Amagata, Makoto Onizuka, and Takahiro Hara. Fast, exact, and parallel-friendly outlier detection algorithms with proximity graph in metric spaces. *The VLDB Journal*, 31(4):797–821, 2022.
- [6] L. An, Q. S. Xiang, and S Chavez. A fast implementation of the minimum spanning tree method for phase unwrapping. *IEEE Transactions on Medical Imaging*, 19(8):805–8, 2000.
- [7] Fabrizio Angiulli and Fabio Fassetti. Distance-based outlier queries in data streams: the novel task and algorithms. *Data Mining and Knowledge Discovery*, 20(2):290–324, 2010.
- [8] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L Reyesortiz. A public domain dataset for human activity recognition using smartphones. In *21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, volume 3, pages 3–4, 2013.

References

- [9] Sunil Arya and David M. Mount. A fast and simple algorithm for computing approximate euclidean minimum spanning trees. In *Twenty-Seventh Acm-Siam Symposium on Discrete Algorithms*, pages 1220–1233, 2016.
- [10] A. C. Atkinson and D. M. Hawkins. Identification of Outliers. 37(4):860, 1980.
- [11] Rashi Bansal, Nishant Gaur, and Shailendra Narayan Singh. Outlier detection: applications and techniques in data mining. In *2016 6th International conference-cloud system and big data engineering (Confluence)*, pages 373–377. IEEE, 2016.
- [12] Stephen D Bay and Mark Schwabacher. Mining distance-based outliers in near linear time with randomization and a simple pruning rule. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 29–38, 2003.
- [13] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Commun. ACM*, 18(9):509–517, September 1975.
- [14] Jon Louis Bentley and J. H Friedman. Fast algorithms for constructing minimal spanning trees in coordinate spaces. *Computers IEEE Transactions on*, C-27(2):97–105, 1975.
- [15] Alina Beygelzimer, Sham Kakade, and John Langford. Cover trees for nearest neighbor. In *Proceedings of the 23rd international conference on Machine learning*, pages 97–104, 2006.
- [16] Brian G. Booth, Jan Sijbers, and Noël L. W. Keijsers. Outlier detection for foot complaint diagnosis: Modeling confounding factors using metric learning. *IEEE Intelligent Systems*, 36(3):41–49, 2021.
- [17] O. Borů vka. Über ein Minimalproblem. *Práce moravské přírodovědecké společnosti* 3, 1926.
- [18] Ismael Cabero, Irene Epifanio, Ana Piérola, and Alfredo Ballester. Archetype analysis: A new subspace outlier detection approach. *Knowledge-Based Systems*, 217:106830, 2021.
- [19] Fuyu Cai and Fons J Verbeek. Rough fuzzy c-means and particle swarm optimization hybridized method for information clustering problem. *J. Commun.*, 11(12):1106–1113, 2016.
- [20] Paul B Callahan and S Rao Kosaraju. Faster algorithms for some geometric graph problems in higher dimensions. In *SODA*, volume 93, pages 291–300, 1993.

-
- [21] Guilherme O. Campos, Arthur Zimek, Jörg Sander, Ricardo J. G. B. Campello, Barbora Micenkova, Erich Schubert, Ira Assent, and Michael E. Houle. On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study. *Data Mining and Knowledge Discovery*, 30(4):891–927, Jul 2016.
- [22] Lei Cao, Yizhou Yan, Caitlin Kuhlman, Qingyang Wang, Elke A Rundensteiner, and Mohamed Eltabakh. Multi-tactic distance-based outlier detection. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pages 959–970. IEEE, 2017.
- [23] Hong Chang and Dit Yan Yeung. Robust path-based spectral clustering. *Pattern Recognition*, 41(1):191–203, 2008.
- [24] Sanjay Chawla and Aristides Gionis. k-means-: A unified approach to clustering and outlier detection. In *Proceedings of the 2013 SIAM international conference on data mining*, pages 189–197. SIAM, 2013.
- [25] Chuan Chen, Yu Wang, Weibo Hu, and Zibin Zheng. Robust multi-view k-means clustering with outlier removal. *Knowledge-Based Systems*, 210:106518, 2020.
- [26] Baolei Cheng, Dajin Wang, and Jianxi Fan. Independent spanning trees in networks: a survey. *ACM Computing Surveys*, 55(14s):1–29, 2023.
- [27] Dongdong Cheng, Qingsheng Zhu, Jinlong Huang, Quanwang Wu, and Lijun Yang. Clustering with local density peaks-based minimum spanning tree. *IEEE Transactions on Knowledge and Data Engineering*, 33(2):374–387, 2019.
- [28] Li Cheng, Yijie Wang, and Xinwang Liu. Neighborhood consensus networks for unsupervised multi-view outlier detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 7099–7106, 2021.
- [29] Ka Wong Chong and Christos Zaroliagis. An optimal parallel algorithm for minimum spanning trees in planar graphs. In *Algorithms, Probability, Networks, and Games: Scientific Papers and Essays Dedicated to Paul G. Spirakis on the Occasion of His 60th Birthday*, pages 169–182. Springer, 2015.
- [30] Matteo Corain, Paolo Garza, and Abolfazl Asudeh. DBSCOUT: A Density-based Method for Scalable Outlier Detection in Very Large Datasets. In *ICDE*, pages 37–48. IEEE, 2021.
- [31] Thomas T. Cormen, Charles E. Leiserson, and Ronald L. Rivest. Introduction to algorithms. *Resonance*, 1(9):14–24, 2009.

References

- [32] Xuan Hong Dang, Barbora Micenková, Ira Assent, and Raymond T. Ng. Outlier detection with space transformation and spectral analysis. *Proceedings of the 2013 SIAM International Conference on Data Mining, SDM 2013*, (i):225–233, 2013.
- [33] Roy De Maesschalck, Delphine Jouan-Rimbaud, and Désiré L Massart. The mahalanobis distance. *Chemometrics and intelligent laboratory systems*, 50(1):1–18, 2000.
- [34] Ali Degirmenci and Omer Karal. Efficient density and cluster based incremental outlier detection in data streams. *Information Sciences*, 607:901–920, 2022.
- [35] Arnab Dey, B Rushi Kumar, Bishakha Das, and Arnab Kumar Ghoshal. Outlier detection in social networks leveraging community structure. *Information Sciences*, 634:578–586, 2023.
- [36] Nameirakpam Dhanachandra, Khumanthem Manglem, and Yambem Jina Chanu. Image segmentation using k -means clustering algorithm and subtractive clustering algorithm. *Procedia Computer Science*, 54:764 – 771, 2015.
- [37] Ling Ding, Chao Li, Di Jin, and Shifei Ding. Survey of spectral clustering based on graph theory. *Pattern Recognition*, page 110366, 2024.
- [38] Rémi Domingues, Maurizio Filippone, Pietro Michiardi, and Jihane Zouaoui. A comparative evaluation of outlier detection algorithms: Experiments and analyses. *Pattern recognition*, 74:406–421, 2018.
- [39] Xuefeng Du, Yiyou Sun, Jerry Zhu, and Yixuan Li. Dream the impossible: Outlier imagination with diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, pages 60878–60901, 2023.
- [40] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017.
- [41] George Economou, Vassilios Pothos, and Apostolos Ifantis. Geodesic distance and mst based image segmentation. In *2004 12th European Signal Processing Conference*, pages 941–944. IEEE, Sept 2004.
- [42] Sarah M. Erfani, Sutharshan Rajasegarar, Shanika Karunasekera, and Christopher Leckie. High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning. *Pattern Recognition*, 58:121 – 134, 2016.
- [43] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.

-
- [44] O Utlier D Etection. SUOD : Accelerating Large-Scale Unsupervised Heterogeneous Outlier Detection. In *MLSys Conference*, pages 463–478, 2021.
- [45] Hadi Fanaee-T and João Gama. Tensor-based anomaly detection: An interdisciplinary survey. *Knowledge-Based Systems*, 98:130–147, 2016.
- [46] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.
- [47] Ugo Fiore, Alfredo De Santis, Francesca Perla, Paolo Zanetti, and Francesco Palmieri. Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences*, 479:448–455, 2019.
- [48] Ugo Fiore, Alfredo De Santis, Francesca Perla, Paolo Zanetti, and Francesco Palmieri. Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences*, 479:448–455, 2019.
- [49] Duarte Folgado, Marília Barandas, Margarida Antunes, Maria Lua Nunes, Hui Liu, Yale Hartmann, Tanja Schultz, and Hugo Gamboa. TSSEARCH: Time series subsequence search library. *SoftwareX*, 18:101049, 2022.
- [50] Marek Gagolewski, Anna Cena, Maciej Bartoszek, and Łukasz Brzozowski. Clustering with minimum spanning trees: How good can it be? *Journal of Classification*, pages 1–23, 2024.
- [51] Jing Gao, Wei Fan, Deepak Turaga, Srinivasan Parthasarathy, and Jiawei Han. A spectral framework for detecting inconsistency across multi-source object relationships. In *2011 IEEE 11th International Conference on Data Mining*, pages 1050–1055. IEEE, 2011.
- [52] David Gil, Jose Luis Girela, Joaquin De Juan, M. Jose Gomez-Torres, and Magnus Johnsson. Predicting seminal quality with artificial intelligence methods. *Expert Systems with Applications*, 39(16):12564–12573, 2012.
- [53] Keyvan Golalipour, Ebrahim Akbari, Seyed Saeed Hamidi, Malrey Lee, and Rasul Enayatifar. From clustering to clustering ensemble selection: A review. *Engineering Applications of Artificial Intelligence*, 104:104388, 2021.
- [54] Markus Goldstein and Andreas Dengel. Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm. volume 1, pages 59–63. Citeseer, 2012.
- [55] Emre Güngör and Ahmet Özmen. Distance and density based clustering algorithm using gaussian kernel. *Expert Systems with Applications*, 69:10–20, 2016.

References

- [56] Maria Halkidi, Yannis Batistakis, and Michalis Vazirgiannis. On clustering validation techniques. *Journal of Intelligent Information Systems*, 17(2-3):107–145, 2001.
- [57] Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 3 edition, 2011.
- [58] V. Hautamaki, I. Karkkainen, and P. Franti. Outlier detection using k-nearest neighbour graph. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 3, pages 430–433 Vol.3, Aug 2004.
- [59] Yu He and Lihui Chen. Minclue: a mst-based clustering method with auto-threshold-detection. In *IEEE Conference on Cybernetics and Intelligent Systems, 2004.*, volume 1, pages 229–233. IEEE, Dec 2004.
- [60] Zengyou He, Xiaofei Xu, and Shengchun Deng. Discovering cluster-based local outliers. *Pattern recognition letters*, 24(9-10):1641–1650, 2003.
- [61] Zengyou He, Xiaofei Xu, and Shengchun Deng. Discovering cluster-based local outliers. *Pattern Recognition Letters*, 24(9-10):1641–1650, 2003.
- [62] Waleed Hilal, S Andrew Gadsden, and John Yawney. Financial fraud: a review of anomaly detection techniques and recent advances. *Expert systems With applications*, 193:116429, 2022.
- [63] Jinlong Huang, Qingsheng Zhu, Lijun Yang, DongDong Cheng, and Quanwang Wu. A novel outlier cluster detection algorithm without top-n parameter. *Knowledge-Based Systems*, 121:32–40, 2017.
- [64] Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of Classification*, 2(1):193–218, 1985.
- [65] Ramin Jahanbegloo and Ramin Jahanbegloo. LOF: Identifying Density-Based Local Outliers. In *SIGMOD*, number October 2017, pages 1–22, 2000.
- [66] Yu-Xuan Ji, Ling Huang, Heng-Ping He, Chang-Dong Wang, Guangqiang Xie, Wei Shi, and Kun-Yu Lin. Multi-view outlier detection in deep intact space. In *2019 IEEE International Conference on Data Mining (ICDM)*, pages 1132–1137. IEEE, 2019.
- [67] Mon-Fong Jiang, Shian-Shyong Tseng, and Chih-Ming Su. Two-phase clustering process for outliers detection. *Pattern Recognition Letters*, 22(6-7):691–700, 2001.

-
- [68] Wen Jin, Anthony K. H. Tung, and Jiawei Han. Mining top-n local outliers in large databases. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '01, pages 293–298, New York, NY, USA, 2001. ACM.
- [69] R. Jothi, Sraban Kumar Mohanty, and Aparajita Ojha. Fast minimum spanning tree based clustering algorithms on local neighborhood graph. In *International workshop on graph-based representations in pattern recognition*, pages 292–301. Springer, Springer International Publishing, 2015.
- [70] R. Jothi, Sraban Kumar Mohanty, and Aparajita Ojha. Fast approximate minimum spanning tree based clustering algorithm. *Neurocomputing*, 272:542–557, 2018.
- [71] Piotr Juszczak, David M. J. Tax, Elbieta PeÿKalska, and Robert P. W. Duin. Minimum spanning tree based one-class classifier. *Neurocomputing*, 72(7–9):1859–1869, 2009.
- [72] Myeongsu Kang, Md Rashedul Islam, Jaeyoung Kim, Jongmyon Kim, and Michael Pecht. A hybrid feature selection scheme for reducing diagnostic performance deterioration caused by outliers in data-driven diagnostics. *IEEE Transactions on Industrial Electronics*, 63(5):3299–3310, 2016.
- [73] Somasundaram Kannan and K Somasundaram. Autoregressive-based outlier algorithm to detect money laundering activities. *Journal of Money Laundering Control*, 20(2):190–202, 2017.
- [74] George Karypis, Eui Hong Han, and Vipin Kumar. Chameleon a hierarchical clustering algorithm using dynamic modeling. *Computer*, 32(8):68–75, 2008.
- [75] Antanas Kascenas, Pedro Sanchez, Patrick Schrempf, Chaoyang Wang, William Clackett, Shadia S Mikhael, Jeremy P Voisey, Keith Goatman, Alexander Weir, Nicolas Pugeault, et al. The role of noise in denoising models for anomaly detection in medical images. *Medical Image Analysis*, 90:102963, 2023.
- [76] Leonard Kaufman and Peter J Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, 2009.
- [77] Fabian Keller, Emmanuel Müller, and Klemens Böhm. HiCS: High contrast subspaces for density-based outlier ranking. In *2012 IEEE 28th international conference on data engineering*, pages 1037–1048. IEEE, 2012.
- [78] Abdul Atif Khan and Sraban Kumar Mohanty. A fast spectral clustering technique using mst based proximity graph for diversified datasets. *Information Sciences*, 609:1113–1131, 2022.

References

- [79] Edwin M Knox and Raymond T Ng. Algorithms for mining distance-based outliers in large datasets. In *Proceedings of the international conference on very large data bases*, pages 392–403. Citeseer, 1998.
- [80] Trupti M Kodinariya, Prashant R Makwana, et al. Review on determining number of cluster in k-means clustering. *International Journal*, 1(6):90–95, 2013.
- [81] Maria Kontaki, Anastasios Gounaris, Apostolos N. Papadopoulos, Kostas Tsihclas, and Yannis Manolopoulos. Efficient and flexible algorithms for monitoring distance-based outliers over data streams. *Information Systems*, 2016.
- [82] Hans-Peter Kriegel, Peer Kröger, Erich Schubert, and Arthur Zimek. Outlier detection in axis-parallel subspaces of high dimensional data. In *Pacific-asia conference on knowledge discovery and data mining*, pages 831–838. Springer, 2009.
- [83] Joseph B Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1):48–50, 1956.
- [84] Neeraj Kumar and Upendra Kumar. Anomaly-based network intrusion detection: An outlier detection techniques. In *Proceedings of the Eighth International Conference on Soft Computing and Pattern Recognition (SoCPaR 2016)*, pages 262–269. Springer, 2018.
- [85] Bjornar Larsen and Chinatsu Aone. Fast and effective text mining using linear-time document clustering. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 16–22, 1999.
- [86] Longin Jan Latecki, Aleksandar Lazarevic, and Dragoljub Pokrajac. Outlier detection with kernel density functions. In *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pages 61–75. Springer, 2007.
- [87] Chi-Kwong Li and Fuzhen Zhang. Eigenvalue continuity and gersgorin’s theorem. *The Electronic Journal of Linear Algebra*, 35:619–625, 2019.
- [88] Jia Li, Xiaochun Wang, and Xiali Wang. A scaled-MST-based clustering algorithm and application on image segmentation. *Journal of Intelligent Information Systems*, 2019.
- [89] Weizhi Li, Weirong Mo, Xu Zhang, John J Squiers, Yang Lu, Eric W Sellke, Wensheng Fan, J Michael DiMaio, and Jeffrey E Thatcher. Outlier detection

- and removal improves accuracy of machine learning approach to multispectral burn diagnostic imaging. *Journal of biomedical optics*, 20(12):121305, 2015.
- [90] Xiaojie Li, Jiancheng Lv, and Zhang Yi. An efficient representation-based method for boundary point and outlier detection. *IEEE transactions on neural networks and learning systems*, 29(1):51–62, 2016.
- [91] Xiaojie Li, Xi Wu, Jiancheng Lv, Jia He, Jianping Gou, and Mao Li. Automatic detection of boundary points based on local geometrical measures. *Soft Computing*, 22(11):3663–3674, 2018.
- [92] Zechao Li and Jinhui Tang. Weakly supervised deep matrix factorization for social image understanding. *IEEE Transactions on Image Processing*, 26(1):276–288, Jan 2017.
- [93] Zheng Li, Yue Zhao, Nicola Botta, Cezar Ionescu, and Xiyang Hu. Copod: copula-based outlier detection. In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 1118–1123. IEEE, 2020.
- [94] Zheng Li, Yue Zhao, Xiyang Hu, Nicola Botta, Cezar Ionescu, and George Chen. Ecod: Unsupervised outlier detection using empirical cumulative distribution functions. *IEEE Transactions on Knowledge and Data Engineering*, 35(12):12181–12193, 2022.
- [95] Jiaxiang Lin, Dongyi Ye, Chongcheng Chen, and Miaoxian Gao. Minimum spanning tree based spatial outlier mining and its applications. In *Rough Sets and Knowledge Technology: Third International Conference, RSKT 2008, Chengdu, China, May 17-19, 2008. Proceedings 3*, pages 508–515. Springer, 2008.
- [96] Fang Liu, Yanwei Yu, Peng Song, Yangyang Fan, and Xiangrong Tong. Scalable KDE-based top-n local outlier detection over large-scale data streams. *Knowledge-Based Systems*, 204:106186, 2020.
- [97] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 eighth ieee international conference on data mining*, pages 413–422. IEEE, 2008.
- [98] Hongfu Liu, Jun Li, Yue Wu, and Yun Fu. Clustering with outlier removal. *IEEE transactions on knowledge and data engineering*, 33(6):2369–2379, 2019.
- [99] Hui Liu and Tanja Schultz. How long are various types of daily activities? statistical analysis of a multimodal wearable sensor-based human activity dataset. In *Proceedings of the 15th International Joint Conference*

References

- on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 5: HEALTHINF*, pages 680–688, 2022.
- [100] Yezheng Liu, Zhe Li, Chong Zhou, Yuanchun Jiang, Jianshan Sun, Meng Wang, and Xiangnan He. Generative adversarial active learning for unsupervised outlier detection. *IEEE Transactions on Knowledge and Data Engineering*, 32(8):1517–1528, 2019.
- [101] Yezheng Liu, Zhe Li, Chong Zhou, Yuanchun Jiang, Jianshan Sun, Meng Wang, and Xiangnan He. Generative Adversarial Active Learning for Unsupervised Outlier Detection. *IEEE Transactions on Knowledge and Data Engineering*, 32(8):1517–1528, 2019.
- [102] Ting Luo and Caiming Zhong. A neighborhood density estimation clustering algorithm based on minimum spanning tree. pages 557–565, 2010.
- [103] Xiaobo Lv, Yan Ma, Xiaofu He, Hui Huang, and Jie Yang. CciMST: A Clustering Algorithm Based on Minimum Spanning Tree and Cluster Centers. *Mathematical Problems in Engineering*, 2018(1):8451796, 2018.
- [104] Emaad Manzoor, Sadegh M. Milajerdi, and Leman Akoglu. Fast memory-efficient anomaly detection in streaming heterogeneous graphs. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1035–1044, 2016.
- [105] William B. March, Parikshit Ram, and Alexander G. Gray. Fast euclidean minimum spanning tree:algorithm, analysis, and applications. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 603–612, 2010.
- [106] Fokrul Alom Mazarbhuiya, Mohammed Y AlZahrani, and Lilia Georgieva. Anomaly detection using agglomerative hierarchical clustering algorithm. In *International Conference on Information Science and Applications*, pages 475–484. Springer, 2018.
- [107] Jogamohan Medak. Review and analysis of minimum spanning tree using prim’s algorithm. *International Journal of Computer Science Trends and Technology (IJCT)*, 6(2), 2018.
- [108] Gaurav Mishra and Sraban Kumar Mohanty. A fast hybrid clustering technique based on local nearest neighbor using minimum spanning tree. *Expert Syst. Appl.*, 132:28–43, 2019.
- [109] Gaurav Mishra and Sraban Kumar Mohanty. A minimum spanning tree based partitioning and merging technique for clustering heterogeneous data sets. *Journal of Intelligent Information Systems*, 55(3):587–606, 2020.

-
- [110] Sayyed Ahmad Naghavi Nozad, Maryam Amir Haeri, and Gianluigi Folino. SDCOR: Scalable density-based clustering for local outlier detection in massive-scale datasets. *Knowledge-Based Systems*, 228:107256, 2021.
- [111] Arpita Nagpal, Arnan Jatain, and Deepti Gaur. Review based on data clustering algorithms. In *2013 IEEE conference on information & communication technologies*, pages 298–303. IEEE, 2013.
- [112] Giri Narasimhan, Jianlin Zhu, and Martin Zachariassen. Experiments with computing geometric minimum spanning trees. In *In Proceedings of ALENEX'00, Lecture Notes in Computer Science*, pages 183–196. SpringerVerlag, 2000.
- [113] Hoang Vu Nguyen, Hock Hee Ang, and Vivekanand Gopalkrishnan. Mining outliers with ensemble of heterogeneous detectors on random subspaces. In *Database Systems for Advanced Applications: 15th International Conference, DASFAA 2010, Tsukuba, Japan, April 1-4, 2010, Proceedings, Part I 15*, pages 368–383. Springer, 2010.
- [114] Hoang Vu Nguyen and Vivekanand Gopalkrishnan. Feature extraction for outlier detection in high-dimensional spaces. In *Feature Selection in Data Mining*, pages 66–75. PMLR, 2010.
- [115] Boya Ouyang, Yu Song, Yuhai Li, Gaurav Sant, and Mathieu Bauchy. EBOD: An ensemble-based outlier detection algorithm for noisy datasets. *Knowledge-Based Systems*, 231:107400, 2021.
- [116] Spiros Papadimitriou, Hiroyuki Kitagawa, Phillip B. Gibbons, and Christos Faloutsos. Loci: Fast outlier detection using the local correlation integral. In *International Conference on Data Engineering*, pages 315–326, 2002.
- [117] Srinivasan Parthasarathy, Matthew Eric Otey, and Amol Ghoting. Fast mining of distance-based outliers in high-dimensional datasets. *Data Mining and Knowledge Discovery*, 16(3):349–364, 2008.
- [118] Bo Peng, Lei Zhang, and David Zhang. A survey of graph theoretical approaches to image segmentation. *Pattern Recognition*, 46(3):1020 – 1038, 2013.
- [119] Yang Peng and Huang Biao. A spectral clustering algorithm for outlier detection. *Proceedings - 2008 International Seminar on Future Information Technology and Management Engineering, FITME 2008*, X:33–36, 2008.
- [120] Tomas Pevný. Loda: Lightweight on-line detector of anomalies. *Machine Learning*, 102(2):275–304, 2016.

References

- [121] Ninh Pham and Rasmus Pagh. A near-linear time approximation algorithm for angle-based outlier detection in high-dimensional data. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 877–885, 2012.
- [122] Harun Pirim, Burak Ekiolu, and Andy D. Perkins. Clustering high throughput biological data with B-MST, a minimum spanning tree based heuristic. *Computers in Biology and Medicine*, 62:94–102, 2015.
- [123] Franco P. Preparata and Michael Ian Shamos. *Computational Geometry*. Springer-Verlag, 1985.
- [124] R. C. Prim. Shortest connection networks and some generalizations. *Bell Labs Technical Journal*, 36(6):1389–1401, 1957.
- [125] Miloš Radovanović, Alexandros Nanopoulos, and Mirjana Ivanović. Reverse nearest neighbors in unsupervised distance-based outlier detection. *IEEE Transactions on Knowledge and Data Engineering*, 27(5):1369–1382, 2014.
- [126] Sridhar Ramaswamy, Rajeev Rastogi, Kyuseok Shim, and Taejon Korea. Efficient algorithms for mining outliers from large data sets. In *Acm Sigmod International Conference on Management of Data*, pages 427–438, 2000.
- [127] William M. Rand. Objective criteria for the evaluation of clustering methods. *Publications of the American Statistical Association*, 66(336):846–850, 1971.
- [128] B Joyce Beula Rani et al. Survey on applying GAN for anomaly detection. In *2020 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–5. IEEE, 2020.
- [129] Shebuti Rayana. ODDS library, 2016.
- [130] Vikas Chandrakant Raykar and Ramani Duraiswami. Fast optimal bandwidth selection for kernel density estimation. In *Proceedings of the 2006 SIAM International Conference on Data Mining*, pages 524–528. SIAM, 2006.
- [131] Mahboobeh Riahi-Madvar, Ahmad Akbari Azirani, Babak Nasersharif, and Bijan Raahemi. A new density-based subspace selection method using mutual information for high dimensional outlier detection. *Knowledge-Based Systems*, 216:106733, 2021.
- [132] João Rodrigues, Hui Liu, Duarte Folgado, David Belo, Tanja Schultz, and Hugo Gamboa. Feature-based information retrieval of multimodal biosignals with a self-similarity matrix: Focus on automatic segmentation. *Biosensors*, 12(12):1182, 2022.

-
- [133] Lior Rokach and Oded Maimon. Clustering methods. *Data mining and knowledge discovery handbook*, pages 321–352, 2005.
- [134] Ali Saglam and Nurdan Akhan Baykan. Sequential image segmentation based on minimum spanning tree representation. *Pattern Recognition Letters*, 87:155 – 162, 2017.
- [135] Saket Sathe and Charu C Aggarwal. Lodes: Local density meets spectral outlier detection. pages 171–179, 2016.
- [136] Amit Saxena, Mukesh Prasad, Akshansh Gupta, Neha Bharill, Om Prakash Patel, Aruna Tiwari, Meng Joo Er, Weiping Ding, and Chin-Teng Lin. A review of clustering techniques and developments. *Neurocomputing*, 267:664–681, 2017.
- [137] Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *Information Processing in Medical Imaging*, pages 146–157, Cham, 2017. Springer International Publishing.
- [138] Erich Schubert, Remigius Wojdanowski, Arthur Zimek, and Hanspeter Kriegel. On evaluation of outlier rankings and outlier scores. pages 1047–1058, 2012.
- [139] Erich Schubert, Arthur Zimek, and Hans Peter Kriegel. Generalized outlier detection with flexible kernel density estimates. *SIAM International Conference on Data Mining 2014, SDM 2014*, 2:542–550, 2014.
- [140] Erich Schubert, Arthur Zimek, and Hans-Peter Kriegel. Local outlier detection reconsidered: a generalized view on locality with applications to spatial, video, and network outlier detection. *Data mining and knowledge discovery*, 28(1):190–237, 2014.
- [141] Abir Smiti. A critical overview of outlier detection methods. *Computer Science Review*, 38:100306, 2020.
- [142] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. CSI: Novelty detection via contrastive learning on distributionally shifted instances. *Advances in neural information processing systems*, 33:11839–11852, 2020.
- [143] Bo Tang and Haibo He. A local density-based approach for outlier detection. *Neurocomputing*, 241:171–180, 2017.
- [144] Athanasios Tsanas and Angeliki Xifara. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings*, 49(49):560–567, 2012.

References

- [145] Maximilian E Tschuchnig and Michael Gadermayr. Anomaly detection in medical imaging—a mini review. *Data Science—Analytics and Applications*, pages 33–38, 2022.
- [146] Vincent S Tseng, Jiaching Ying, Chewei Huang, Yimin Kao, and Kuanta Chen. Fraudetector: A graph-mining-based framework for fraudulent phone call detection. pages 2157–2166, 2015.
- [147] Jake VanderPlas. `mst_clustering`: Clustering via euclidean minimum spanning trees. *Journal of Open Source Software*, 1(1):12, 2016.
- [148] Filippo Vella, Ignazio Infantino, Salvatore Gaglio, and Gaspare Vetrano. Image segmentation through a hierarchy of minimum spanning trees. In *2012 Eighth International Conference on Signal Image Technology and Internet Based Systems*, pages 381–388. IEEE, Nov 2012.
- [149] Shiping Wang, Jinbin Yang, Jie Yao, Yang Bai, and William Zhu. An overview of advanced deep graph node clustering. *IEEE Transactions on Computational Social Systems*, 11(1):1302–1314, 2023.
- [150] Wei Wang, Jiong Yang, and Richard Muntz. STING: A statistical information grid approach to spatial data mining. In *VLDB*, volume 97, pages 186–195, 1997.
- [151] Xia Li Wang and Xiaochun Wang. An efficient approximate emst algorithm for color image segmentation. In *Machine Learning and Data Mining in Pattern Recognition*, pages 147–159. Springer International Publishing, 2018.
- [152] Xia Li Wang, Xiaochun Wang, and Xiaqiong Li. A fast two-level approximate euclidean minimum spanning tree algorithm for high-dimensional data. In Petra Perner, editor, *Machine Learning and Data Mining in Pattern Recognition*, pages 273–287, Cham, 2018. Springer International Publishing.
- [153] Xiaochun Wang, Xia Li Wang, Cong Chen, and D. Mitchell Wilkes. Enhancing minimum spanning tree-based clustering by removing density-based outliers. *Digital Signal Processing*, 23(5):1523–1538, 2013.
- [154] Xiaochun Wang, Xia Li Wang, and D. Mitch Wilkes. A minimum spanning tree-inspired clustering-based outlier detection technique. In *Industrial Conference on Advances in Data Mining: Applications and Theoretical Aspects*, pages 209–223, 2012.
- [155] Xiaochun Wang, Xiali Wang, and D. Mitchell Wilkes. A divide-and-conquer approach for minimum spanning tree-based clustering. *IEEE Transactions on Knowledge and Data Engineering*, 21(7):945–958, 2009.

-
- [156] Xuhong Wang, Ying Du, Shijie Lin, Ping Cui, Yuntian Shen, and Yupu Yang. advae: A self-adversarial variational autoencoder with gaussian anomaly prior knowledge for anomaly detection. *Knowledge-Based Systems*, 190:105187, 2020.
- [157] Yu Wang, Chuan Chen, Jinrong Lai, Lele Fu, Yuren Zhou, and Zibin Zheng. A self-representation method with local similarity preserving for fast multi-view outlier detection. *ACM Transactions on Knowledge Discovery from Data*, 17(1):1–20, 2023.
- [158] Yuan Wang, Xiaochun Wang, and Xia Li Wang. A Spectral Clustering Based Outlier Detection Technique. In *MLDM*, volume 3587, pages 15–27, 2016.
- [159] Yue Fei Wang, Yu Jiong, Guo Ping Su, and Yu Rong Qian. A new outlier detection method based on OPTICS. *Sustainable Cities and Society*, 45(September 2018):197–212, 2019.
- [160] Ying Xu, Victor Olman, and Dong Xu. Clustering gene expression data using a graph-theoretic approach: an application of minimum spanning trees. *Bioinformatics*, 18(4):536–545, 2002.
- [161] Ying Xu and Edward C. Uberbacher. 2d image segmentation using minimum spanning trees. *Image and Vision Computing*, 15(1):47–57, 1997.
- [162] C. T Zahn. Graph-theoretical methods for detecting and describing gestalt clusters. In *IEEE Trans. on Computers*, pages 68–86, 1971.
- [163] L Zelnik-Manor and P Perona. Self-tuning spectral clustering. *Advances in Neural Information Processing Systems*, 17:1601–1608, 2004.
- [164] Houssam Zenati, Manon Romain, Chuan-Sheng Foo, Bruno Lecouat, and Vijay Chandrasekhar. Adversarially learned anomaly detection. In *2018 IEEE International conference on data mining (ICDM)*, pages 727–736. IEEE, 2018.
- [165] Hui Zhang, Jason E. Fritts, and Sally A. Goldman. Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding*, 110(2):260 – 280, 2008.
- [166] Ke Zhang, Marcus Hutter, and Huidong Jin. A new local distance-based outlier detection approach for scattered real-world data. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 813–822. Springer, 2009.
- [167] Liangwei Zhang, Jing Lin, and Ramin Karim. Adaptive kernel density-based anomaly detection for nonlinear systems. *Knowledge-Based Systems*, 139:50–63, 2018.

References

- [168] Lingxiao Zhang, Xiang Li, Haifeng Liu, Jing Mei, Gang Hu, Junfeng Zhao, Yanzhen Zou, Bing Xie, and Guotong Xie. Probabilistic-mismatch anomaly detection: Do one’ s medications match with the diagnoses. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pages 659–668, 2016.
- [169] Yang Zhang, Nirvana Meratnia, and Paul Havinga. Outlier detection techniques for wireless sensor networks: A survey. *IEEE communications surveys & tutorials*, 12(2):159–170, 2010.
- [170] Handong Zhao and Yun Fu. Dual-Regularized Multi-View Outlier Detection. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, number Ijcai, pages 4077–4083, 2015.
- [171] Handong Zhao, Hongfu Liu, Zhengming Ding, and Yun Fu. Consensus regularized multi-view outlier detection. *IEEE Transactions on Image Processing*, 27(1):236–248, 2017.
- [172] Yue Zhao, Maciej K Hryniewicki, Zain Nasrullah, and Zheng Li. Lscp: Locally selective combination in parallel outlier ensembles. In *Proceedings of the 2019 SIAM international conference on data mining*, pages 585–593. SIAM, 2019.
- [173] Yue Zhao, Zain Nasrullah, and Zheng Li. Pyod: A python toolbox for scalable outlier detection. *Journal of Machine Learning Research*, 20(96):1–7, 2019.
- [174] Caiming Zhong, Mikko Malinen, Duoqian Miao, and Pasi Fränti. A fast minimum spanning tree algorithm based on k-means. *Information Sciences*, 295:1–17, 2015.
- [175] Caiming Zhong, Duoqian Miao, and Pasi Nti. Minimum spanning tree based split-and-merge: A hierarchical clustering method. *Information Sciences*, 181(16):3397–3410, 2011.
- [176] Caiming Zhong, Duoqian Miao, and Ruizhi Wang. A graph-theoretical clustering method based on two rounds of minimum spanning trees. *Pattern Recognition*, 43(3):752–766, 2010.
- [177] Xiaofeng Zhu, Shichao Zhang, Yonggang Li, Jilian Zhang, Lifeng Yang, and Yue Fang. Low-Rank Sparse Subspace for Spectral Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 31(8):1532–1543, 2019.
- [178] Arthur Zimek, Matthew Gaudet, Ricardo JGB Campello, and Jörg Sander. Subsampling for efficient and effective unsupervised outlier detection ensembles. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 428–436, 2013.

- [179] Bo Zong, Qi Song, Martin Renqiang Min, Wei Cheng, Cristian Lumezanu, Daeki Cho, and Haifeng Chen. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *International Conference on Learning Representations*, 2018.

References
