



















## 7. REFERENCES

- [1] C. C. Aggarwal and H. Wang. *Managing and Mining Graph Data*, chapter Graph Data Management and Mining: A Survey of Algorithms and Applications, pages 13–68. Springer, 2010.
- [2] A. Arora, M. Sachan, and A. Bhattacharya. Mining statistically significant connected subgraphs in vertex labeled graphs. In *SIGMOD*, pages 1003–1014, 2014.
- [3] L. Babai. Graph isomorphism in quasipolynomial time. In *STOC*, pages 684–697, 2016.
- [4] P. Barcelo, L. Libkin, and J. L. Reutter. Querying graph patterns. In *PODS*, pages 199–210, 2011.
- [5] G. Bejerano, N. Friedman, and N. Tishby. Efficient exact p-value computation for small sample, sparse and surprisingly categorical data. *JCB*, 11(5):867–886, 2004.
- [6] A. Bordes, S. Chopra, and J. Weston. Question answering with subgraph embeddings. In *EMNLP*, pages 615–620, 2014.
- [7] C. Chen, X. Yan, P. S. Yu, J. Han, D. qing Zhang, and X. Gu. Towards graph containment search and indexing. In *VLDB*, pages 926–937, 2007.
- [8] J. Cheng, Y. Ke, W. Ng, and A. Lu. FG-Index: Towards verification-free query processing on graph databases. In *SIGMOD*, pages 857–872, 2007.
- [9] D. Conte, P. Foggia, C. Sansone, and M. Vento. Thirty years of graph matching in pattern recognition. *IJPRAI*, 18(3):265–298, 2004.
- [10] S. A. Cook. The complexity of theorem-proving procedures. In *STOC*, pages 151–158, 1971.
- [11] L. P. Cordella, P. Foggia, C. Sansone, and M. Vento. A (sub)graph isomorphism algorithm for large graphs. *PAMI*, 26(10):1367–1372, 2004.
- [12] M. K. Das and H. K. Dai. A survey of DNA motif finding algorithms. *BMC Bioinf.*, 8(7):1–13, 2007.
- [13] S. Dutta. MIST: Top-k approximate sub-string mining using triplet statistical significance. In *ECIR*, pages 284–290, 2015.
- [14] S. Dutta and A. Bhattacharya. Most significant substring mining based on chi-square measure. In *PAKDD*, pages 319–327, 2010.
- [15] W. Fan, J. Li, J. Luo, Z. Tan, X. Wang, and Y. Wu. Incremental graph pattern matching. In *SIGMOD*, pages 925–936, 2011.
- [16] W. Fan, J. Li, S. Ma, N. Tang, Y. Wu, and Y. Wu. Graph pattern matching: From intractable to polynomial time. *PVLDB*, 3(1-2):264–275, 2010.
- [17] B. Gallagher. Matching structure and semantics: A survey on graph-based pattern matching. In *AAAI*, pages 45–53, 2006.
- [18] R. Giugno and D. Shasha. GraphGrep: A fast and universal method for querying graphs. *ICPR*, 2:201–212, 2002.
- [19] W.-S. Han, J. Lee, M.-D. Pham, and J. X. Yu. iGraph: A framework for comparisons of disk-based graph indexing techniques. *PVLDB*, 3(1-2):449–459, 2010.
- [20] B. Hixon, P. Clark, and H. Hajishirzi. Learning knowledge graphs for question answering through conversational dialog. In *NAACL*, pages 851–861, 2015.
- [21] Y. Jia, J. Zhang, and J. Huan. An efficient graph-mining method for complicated and noisy data with real-world applications. *Knowledge and Information Systems*, 28(2):423–447, 2011.
- [22] H. Jiang, H. Wang, P. S. Yu, and S. Zhou. GString: A novel approach for efficient search in graph dbs. In *ICDE*, pages 566–575, 2007.
- [23] B. P. Kelley, B. Yuan, F. Lewitter, R. Sharan, B. Stockwell, and T. Ideker. PathBLAST: A tool for alignment of protein interaction networks. *Nucleic Acids Res.*, 32:83–88, 2004.
- [24] A. Khan, Y. Wu, C. C. Aggarwal, and X. Yan. NeMa: Fast graph search with label similarity. *VLDB*, 6(3):181–192, 2013.
- [25] A. Khan, X. Yan, and K.-L. Wu. Towards proximity pattern mining in large graphs. *SIGMOD*, pages 867–878, 2010.
- [26] G. Klyne and J. J. Carroll. Resource description framework (RDF): Concepts and abstract syntax. <http://www.citeulike.org/group/2170/article/532408>, 2006.
- [27] S. Kpodjedo, P. Galinier, and G. Antoniol. Using local similarity measures to efficiently address approximate graph matching. *Disc. Appl. Math.*, 164:161–177, 2014.
- [28] M. Kuramochi and G. Karypis. Frequent subgraph discovery. *ICDM*, pages 313–320, 2001.
- [29] M. Kuramochi and G. Karypis. GREW – A scalable frequent subgraph discovery algorithm. *ICDM*, pages 439–442, 2004.
- [30] M. Kuramochi and G. Karypis. Finding frequent patterns in a large sparse graph. *Data Mining & Know. Disc.*, 11(3):243–271, 2005.
- [31] Z. Liang, M. Xu, M. Teng, and L. Niu. NetAlign: A web-based tool for comparison of protein interaction networks. *Bioinf.*, 22(17):2175–2177, 2006.
- [32] M. Mongiovi, R. Di Natale, R. Guigno, A. Pulvirenti, A. Ferro, and R. Sharan. SIGMA: A set-cover-based inexact graph matching algorithm. *JBCB*, 8(2):199–218, 2010.
- [33] T. Read and N. Cressie. Pearson’s  $\chi^2$  and the likelihood ratio statistic  $G^2$ : a comparative review. *Int. Statistical Review*, 57(1):19–43, 1989.
- [34] T. R. C. Read and N. A. C. Cressie. *Goodness-of-fit statistics for discrete multivariate data*. Springer Series in Statistics, 1988.
- [35] M. Sachan and A. Bhattacharya. Mining statistically significant substrings using the  $\chi^2$  statistic. *PVLDB*, 5(10):1052–1063, 2012.
- [36] H. Shang, Y. Zhang, X. Lin, and J. Yu. Taming verification hardness: An efficient algorithm for testing subgraph isomorphism. *PVLDB*, 1(1):364–375, 2008.
- [37] R. Shen and C. Guda. Applied graph-mining algorithms to study biomolecular interaction networks. *BioMed Research Int.*, 2014(article 439476):11, 2014.
- [38] R. Singh, J. Xu, and B. Berger. Global alignment of multiple protein interaction networks with application to functional orthology detection. *PNAS*, 105(35):12763–12768, 2008.
- [39] F. M. Suchanek, G. Kasneci, and G. Weikum. YAGO: A core of semantic knowledge. In *WWW*, pages 697–706, 2007.
- [40] B. Sun. *Mining, Indexing, and Search Approach to Entity and Graph Information Retrieval for Chemoinformatics*. PhD thesis, Pennsylvania State Univ., 2008.
- [41] Y. Tian, R. McEachin, C. Santos, D. States, and J. Patel. SAGA: A subgraph matching tool for biological graphs. *Bioinf.*, 23(2):232–239, 2006.
- [42] Y. Tian and J. Patel. TALE: A tool for approximate large graph matching. In *ICDE*, pages 963–972, 2008.
- [43] H. Tong and C. Faloutsos. Center-piece subgraphs: problem definition and fast solutions. In *KDD*, pages 404–413, 2006.
- [44] H. Tong, C. Faloutsos, B. Gallagher, and T. Eliassi-Rad. Fast best-effort pattern matching in large attributed graphs. In *KDD*, pages 737–746, 2007.
- [45] W. H. Tsai and K. Fu. Error-correcting isomorphisms of attributed relational graphs for pattern recognition. *IEEE Systems, Man, and Cybernetics*, 9(12):757–768, 1979.
- [46] J. R. Ullmann. An algorithm for subgraph isomorphism. *JACM*, 23(1):31–42, 1976.
- [47] F. Vandin, E. Uptal, and B. J. Raphael. Algorithms for detecting significantly mutated pathways in cancer. *JCB*, 18(3):507–522, 2011.
- [48] J. Wang, N. Ntarmos, and P. Triantafyllou. Indexing query graphs to speedup graph query processing. In *EDBT*, pages 41–52, 2016.
- [49] X. Yan and J. Han. gSpan: Graph-based substructure pattern mining. *ICDM*, pages 721–724, 2002.
- [50] X. Yan, P. S. Yu, and J. Han. Graph indexing based on discriminative frequent structure analysis. *TODS*, 30(4):960–993, 2005.
- [51] X. Yan, P. S. Yu, and J. Han. Substructure similarity search in graph databases. In *SIGMOD*, pages 766–777, 2005.
- [52] S. Zhang, J. Li, H. Gao, and Z. Zou. A novel approach for efficient supergraph query processing on graph databases. In *EDBT*, pages 204–215, 2009.
- [53] S. Zhang, S. Li, and J. Yang. GADDI: Distance index based subgraph matching in biological networks. In *EDBT*, pages 192–203, 2009.
- [54] S. Zhang, J. Yang, and W. Jin. SAPPER: Subgraph indexing and approximate matching in large graphs. *PVLDB*, 3:1185–1194, 2010.
- [55] L. Zou, L. Chen, J. X. Yu, and Y. Lu. A novel spectral coding in a large graph database. In *EDBT*, pages 181–192, 2008.
- [56] Q. Zou, S. Liu, and W. W. Chu. Ctree: A compact tree for indexing XML data. In *WIDM*, pages 39–46, 2004.