

8. REFERENCES

- [1] N. Bansal, F. Chiang, N. Koudas, and F. W. Tompa. Seeking stable clusters in the blogosphere. In *VLDB*, pages 806–817, 2007.
- [2] H. Bruhn, R. Diestel, and M. Stein. Menger’s theorem for infinite graphs with ends. *J. Graph Theory*, 50:199–211, November 2005.
- [3] C. Chambers, A. Raniwala, F. Perry, S. Adams, R. R. Henry, R. Bradshaw, and N. Weizenbaum. Flumejava: easy, efficient data-parallel pipelines. In *PLDI*, pages 363–375, 2010.
- [4] Y.-H. Chiang, A. Doan, and J. F. Naughton. Tracking entities in the dynamic world: A fast algorithm for matching temporal records. *PVLDB*, 7(6):469–480, 2014.
- [5] W. W. Cohen, P. Ravikumar, and S. E. Fienberg. A comparison of string distance metrics for name-matching tasks. In *IIWEB*, pages 73–78, 2003.
- [6] D. Dey. Entity matching in heterogeneous databases: A logistic regression approach. *Decis. Support Syst.*, 44:740–747, 2008.
- [7] A. K. Elmagarmid, P. G. Ipeirotis, and V. S. Verykios. Duplicate record detection: A survey. *IEEE Trans. Knowl. Data Eng.*, 19(1):1–16, 2007.
- [8] S. Even and E. R. Tarjan. Network flow and testing graph connectivity. *SIAM Journal on Computing*, 4(4):507–518, 1975.
- [9] W. Fan, X. Jia, J. Li, and S. Ma. Reasoning about record matching rules. *PVLDB*, 2(1):407–418, 2009.
- [10] I. P. Fellegi and A. B. Sunter. A theory for record linkage. *Journal of the Americal Statistical Association*, 64(328):1183–1210, 1969.
- [11] L. R. Ford and D. R. Fulkerson. *Flows in networks*. Princeton University Press, 1962.
- [12] L. Getoor and C. P. Diehl. Link mining: A survey. *SIGKDD Explor. Newsl.*, 7(2):3–12, 2005.
- [13] L. Getoor and A. Machanavajjhala. Entity resolution: Theory, practice & open challenges. *PVLDB*, 5(12):2018–2019, 2012.
- [14] S. Guo, X. Dong, D. Srivastava, and R. Zajac. Record linkage with uniqueness constraints and erroneous values. *PVLDB*, 3(1):417–428, 2010.
- [15] O. Hassanzadeh, F. Chiang, H. C. Lee, and R. J. Miller. Framework for evaluating clustering algorithms in duplicate detection. *PVLDB*, pages 1282–1293, 2009.
- [16] M. A. Hernandez and S. J. Stolfo. Real-world data is dirty: Data cleansing and the merge/purge problem. *Data Mining and Knowledge Discovery*, 2:9–37, 1998.
- [17] S. Huang. Mixed group discovery: Incorporating group linkage with alternatively consistent social network analysis. *International Conference on Semantic Computing*, 0:369–376, 2010.
- [18] N. Koudas, S. Sarawagi, and D. Srivastava. Record linkage: similarity measures and algorithms. In *SIGMOD*, pages 802–803, 2006.
- [19] B. Larsen and C. Aone. Fast and effective text mining using linear-time document clustering. In *KDD*, pages 16–22, 1999.
- [20] P. Li, X. L. Dong, A. Maurino, and D. Srivastava. Linking temporal records. *PVLDB*, 4(11):956–967, 2011.
- [21] P. Li, X. Luna Dong, S. Guo, A. Maurino, and D. Srivastava. Robust Group Linkage. <http://arxiv.org/abs/1503.00604>, Mar. 2015.
- [22] X. Liu, Y. Gong, W. Xu, and S. Zhu. Document clustering with cluster refinement and model selection capabilities. In *SIGIR*, pages 191–198, 2002.
- [23] B. W. On, N. Koudas, D. Lee, and D. Srivastava. Group linkage. In *ICDE*, pages 496–505, 2007.
- [24] B. Taskar, M. fai Wong, P. Abbeel, and D. Koller. Link prediction in relational data. In *Advances in Neural Information Processing Systems*, 2003.
- [25] S. E. Whang, D. Menestrina, G. Koutrika, M. Theobald, and H. Garcia-Molina. Entity resolution with iterative blocking. In *SIGMOD*, pages 219–232, 2009.
- [26] D. T. Wijaya and S. Bressan. Ricochet: A family of unconstrained algorithms for graph clustering. In *DASFAA*, pages 153–167, 2009.
- [27] W. E. Winkler. Methods for record linkage and bayesian networks. Technical report, U.S. Bureau of the Census, 2002.
- [28] M. Yoshida, M. Ikeda, S. Ono, I. Sato, and H. Nakagawa. Person name disambiguation by bootstrapping. In *SIGIR*, pages 10–17, 2010.