



Universiteit
Leiden

The Netherlands

Metrics and visualisation for crime analysis and genomics

Laros, J.F.J.

Citation

Laros, J. F. J. (2009, December 21). *Metrics and visualisation for crime analysis and genomics*. *IPA Dissertation Series*. Retrieved from <https://hdl.handle.net/1887/14533>

Version: Corrected 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/1887/14533>

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

Bibliography

- [1] A. Aho and M. Corasick. Efficient string matching: An aid to bibliographic search. *Communications of the ACM*, 18:333–340, 1975.
- [2] A.N. Arslan. Efficient approximate dictionary look-up for long words over small alphabets. *Lecture Notes in Computer Science*, 3887:118–129, 2006. LATIN, Theoretical Informatics: 7th Latin American Symposium.
- [3] K. Bierkandt and J. Bierkandt. DNA rainbow. 2009. Website available via <http://www.dna-rainbow.org/contact.html>, Retrieved April 3, 2009.
- [4] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [5] I. Borg and P. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer, 1997.
- [6] J.R. Bray and J.T. Curtis. An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.*, 27:325–349, 1957.
- [7] K.J. Breslauer, R. Frank, H. Blöcker, and L.A. Marky. Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. USA*, 83:3746–3750, 1986.
- [8] J. Broekens, T.K. Cocx, and W.A. Kosters. Object-centered interactive multi-dimensional scaling: Ask the expert. In *Proceedings of the 18th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2006)*, pages 59–66, 2006.
- [9] J.S. de Bruin, T.K. Cocx, W.A. Kosters, J.F.J. Laros, and J.N. Kok. Data mining approaches to criminal career analysis. In *Proceedings of the Sixth IEEE International Conference on Data Mining (ICDM 2006)*, pages 171–177. IEEE, 2006.
- [10] J.S. de Bruin, T.K. Cocx, W.A. Kosters, J.F.J. Laros, and J.N. Kok. Onto clustering of criminal careers. In *Proceedings of the Workshop on Practical Data Mining: Applications, Experiences and Challenges (ECML/PKDD-2006)*, pages 90–93. IEEE, 2006.
- [11] P.Y. Chan, T.W. Lam, and S.M. Yiu. A more accurate and efficient whole genome phylogeny. *Asia-Pacific Bioinformatics*, pages 337–351, 2006.

- [12] T.K. Cocx and W.A. Kusters. A distance measure for determining similarity between criminal investigations. In *Advances in Data Mining, Proceedings of the Industrial Conference on Data Mining 2006 (ICDM2006)*, volume 4065 of *LNAI*, pages 511–525. Springer, 2006.
- [13] T.K. Cocx, W.A. Kusters, and J.F.J. Laros. Enhancing the automated analysis of criminal careers. In *SIAM Workshop on Link Analysis, Counterterrorism, and Security 2008 (LACTS2008)*. SIAM Press, 2008.
- [14] T.K. Cocx, W.A. Kusters, and J.F.J. Laros. Temporal extrapolation within a static clustering. In *Foundations of Intelligent Systems, Proceedings of ISMIS 2008*, volume 4994 of *LNAI*, pages 189–195. Springer, 2008.
- [15] T.K. Cocx, W.A. Kusters, and J.F.J. Laros. Temporal extrapolation within a static clustering. In *Proceedings of the 20th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2008)*, pages 295–296, 2008.
- [16] M.L. Davison. *Multidimensional Scaling*. John Wiley and Sons, New York, 1983.
- [17] C.W. Dieffenbach and G.S. Dveksler. *PCR Primer: A Laboratory Manual*. CSHL Press, Cold Spring Harbor, USA, 1995.
- [18] T. Duchamp and W. Stuetzle. Geometric properties of principal curves in the plane. *Lecture Notes in Statistics*, 109:135–152, 1996.
- [19] P. Eades. A heuristic for graph drawing. *Congressus Numerantium*, 42:149–160, 1984.
- [20] I. Fischer. Similarity-based neural networks for applications in computational molecular biology. *Lecture Notes in Computer Science*, 2810:208–218, 2003.
- [21] W.M. Fitch and E. Margoliash. Construction of phylogenetic trees. *Science*, 155:279–284, 1967.
- [22] I.K. Fodor. A survey of dimension reduction techniques. technical report. *Lawrence Livermore National Laboratory*, 2002.
<http://www.llnl.gov/tid/lof/documents/pdf/240921.pdf>.
- [23] G. Gibson and S. Muse. *A Primer of Genome Science*. Boyle Biochemistry and Molecular Biology Education, 2nd edition, 2005.
- [24] G. Glazko, A. Gordon, and A. Mushegian. The choice of optimal distance measure in genome-wide datasets. *Bioinformatics*, 21:iii3–iii11, 2005.
- [25] E.H. de Graaf, J.N. Kok, and W.A. Kusters. Clustering improves the exploration of graph mining results. In C. Boukis, A. Pnevmatikakis, and L. Polymenakos, editors, *Proceedings of the 4th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI07)*, pages 13–20. Springer, 2007.

- [26] D. Gusfield. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, 1997.
- [27] S. Haas, M. Vingron, A. Poustka, and S. Wiemann. Primer design for large scale sequencing. *Nucleic Acids Research*, 26(12):3006–3012, 1998.
- [28] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2001.
- [29] T.J. Hastie and W. Stuetzle. Principal curves. *J. Amer. Statist. Assoc.*, 84:502–516, 1989.
- [30] G. Hinton and S.T. Roweis. Stochastic neighbor embedding. *Advances in Neural Information Processing Systems*, 15:833–840, 2003.
- [31] D.S. Hirschberg. Algorithms for the longest common subsequence problem. *Journal of the ACM*, 24(4):664–675, 1977.
- [32] H.J. Hoogeboom, W.A. Kosters, and J.F.J. Laros. Selection of DNA markers. *IEEE Transactions on Systems, Man, and Cybernetics Part C*, 38:26–32, 2008.
- [33] S. Inenaga, H. Hoshino, A. Shinohara, M. Takeda, S. Arikawa, G. Mauri, and Pavesi G. On-line construction of compact directed acyclic word graphs. *Discrete Applied Mathematics*, 146(2):156–179, 2005.
- [34] P. Jaccard. Lois de distribution florale dans la zone alpine. *Bull. Soc. Vaud. Sci. Nat.*, 38:69–130, 1902.
- [35] I.T. Jolliffe. *Principal Component Analysis*. Springer, 2nd edition, 2002.
- [36] T. Kadota, M. Hirao, A. Ishino, M. Takeda, A. Shinohara, and F. Matsuo. Musical sequence comparison for melodic and rhythmic similarities. In *Proceedings of the Eighth International Symposium on String Processing and Information Retrieval*, pages 111–122. IEEE, 2001.
- [37] T. Kämpke, M. Kieninger, and M. Mecklenburg. Efficient primer design algorithms. *Bioinformatics*, 17(3):214–225, 2001.
- [38] D. Knuth, J. Morris, and V. Pratt. Fast pattern matching in strings. *SIAM J. Computing*, 6:323–350, 1977.
- [39] T. Kohonen. *Self-Organizing Maps*. New York: Springer-Verlag, 1997.
- [40] W.A. Kosters and J.F.J. Laros. Metrics for mining multisets. In *Research and Development in Intelligent Systems XXIV, Proceedings of AI-2007*, pages 293–303. Springer, 2007.
- [41] W.A. Kosters and J.F.J. Laros. Metrics for mining multisets. In *Proceedings of the 20th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2008)*, pages 329–330, 2007.

- [42] W.A. Kusters and J.F.J. Laros. Visualisation on a closed surface. In *Proceedings of the 19th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2007)*, pages 189–195, 2007.
- [43] W.A. Kusters and M.C. van Wezel. Competitive neural networks for customer choice models, in e-commerce and intelligent methods. *Studies in Fuzziness and Soft Computing*, 105:41–60, 2002.
- [44] J. F. J. Laros. Visualising DNA with Rauzy projections, 2009. Web application available via <http://www.liacs.nl/home/jlaros/dnavis/> [retrieved April 3, 2009].
- [45] V.I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Doklady Akademii Nauk SSSR*, 163(4):845–848, 1965. English translation in *Soviet Physics Doklady* 10(8):707–710, 1966.
- [46] Q. Ma, J.T.L. Wang, D. Shasha, and C.H. Wu. DNA sequence classification via an expectation maximization algorithm and neural networks: A case study. *IEEE Transactions on Systems, Man, and Cybernetics Part C — Applications and Reviews*, 31(4):468–475, 2001.
- [47] M.S. Madhusudhan, M.A. Marti-Renom, R. Sanchez, and A. Sali. Variable gap penalty for protein sequence-structure alignment. *Protein Engineering Design and Selection*, 19(3):129–133, 2006.
- [48] P.C. Mahalanobis. On the generalised distance in statistics. *Proceedings of the National Institute of Science of India*, 12:49–55, 1936.
- [49] H. Mannila and P. Ronkainen. Similarity of event sequences. In *Proceedings of the International Symposium on Temporal Representation and Reasoning*, pages 136–139. IEEE Computer Society, 1997.
- [50] C. McCue. *Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis*. Butterworth-Heinemann, 1st edition, 2007.
- [51] J. Mena. *Investigative Data Mining for Security and Criminal Detection*. Butterworth-Heinemann, 1st edition, 2003.
- [52] D.R. Morrison. PATRICIA — Practical algorithm to retrieve information coded in alphanumeric. *Journal of the ACM*, 15(4):514–534, 1968.
- [53] A. Mylläri, T. Salakoski, and A. Pasechnik. On the visualization of the DNA sequence and its nucleotide content. *SIGSAM Bull.*, 39(4):131–135, 2005.
- [54] S.B. Needleman and C.D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, 48:443–453, 1970.
- [55] A. Oliphant, D.L. Barker, J.R. Stuelpnagel, and M.S. Chee. SNPs: Discovery of markers for disease. *BioTechniques*, 32:S56–S61, 2002.

- [56] S.K. Pal, S. Bandyopadhyay, and S.S. Ray. Evolutionary computation in bioinformatics: A review. *IEEE Transactions on Systems, Man, and Cybernetics Part C — Applications and Reviews*, 36(5):601–615, 2006.
- [57] K. Pearson. On lines and planes of closest fit to systems of points in space. *Phil. Mag.*, 2(6):559–572, 1901.
- [58] G. Raddatz, M. Dehio, T.F. Meyer, and C. Dehio. Primearray: Genome-scale primer design for DNA-microarray construction. *Bioinformatics*, 17(1):98–99, 2001.
- [59] G. Rauzy. Nombres algébriques et substitutions. *Bull. Soc. Math. France*, 110:147–178, 1982.
- [60] S. Rozen and H.J. Skaletsky. Primer3, 1998.
http://www-genome.wi.mit.edu/genome_software/other/primer3.html.
- [61] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2nd edition, 2002.
- [62] P. Rutherford, C. Churcher, and J. McCallum. An interactive visualisation for investigating DNA sequence information. In *APVis '04: Proceedings of the 2004 Australasian Symposium on Information Visualisation*, pages 101–107. Australian Computer Society, Inc., 2004.
- [63] M. Schena, D. Shalon, R.W. Davis, and P.O. Brown. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270(5235):467–470, 1995.
- [64] J.P. Schouten, C.J. McElgunn, R. Waaijer, D. Zwijnenburg, F. Diepvens, and G. Pals. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acid Research*, 30(12, e57):1–13, 2002.
- [65] J. Setubal and J. Meidanis. *Introduction to Computational Molecular Biology*. Thomson Learning, 1997.
- [66] T. Smith and M. Waterman. Identification of common molecular subsequences. *J. Mol. Biol.*, 147:195–197, 1981.
- [67] P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Addison-Wesley, 2005.
- [68] Cox T.K., W.A. Kusters, and J.F.J. Laros. An early warning system for the prediction of criminal careers. In *MICAI 2008: Advances in Artificial Intelligence*, volume 5317 of *LNAI*, pages 77–89. Springer, 2008.
- [69] UCSC Genome Bioinformatics. <http://genome.ucsc.edu/>, version April 21, 2006 [retrieved April 3, 2009].

- [70] P.M. Vallone and J.M. Butler. Autodimer: A screening tool for primer-dimer and hairpin structures. *BioTechniques*, 37:226–231, 2004.
- [71] C. Varotto, E. Richly, F. Salamini, and D. Leister. GST-PRIME: A genome-wide primer design software for the generation of gene sequence tags. *Nucleic Acids Research*, 29(21):4373–4377, 2001.
- [72] M. Vingron and P.A. Pevzner. Multiple sequence comparison and consistency on multipartite graphs. *Adv. Appl. Math.*, 16:1–22, 1995.
- [73] A.A. Westen, A. S. Matai, J.F.J. Laros, H.C. Meiland, M. Jasper, W.J.F. de Leeuw, P. de Knijff, and T. Sijen. Tri-allelic SNP markers enable analysis of mixed and degraded DNA samples. *Forensic Science International: Genetics*, 3(4):233–241, 2009.