



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).

**Metrics and Visualisation
for
Crime Analysis and Genomics**

Jeroen F. J. Laros



This research is part of the DALE (Data Assistance for Law Enforcement) project as financed in the ToKeN program from the Netherlands Organization for Scientific Research (NWO) under grant number 634.000.430.



The work in this thesis has been carried out under the auspices of the research school IPA (Institute for Programming Research and Algorithmics).

Cover: Stereogram of Figure 4.2.

ISBN: 978-90-9024936-0

**Metrics and Visualisation
for
Crime Analysis and Genomics**

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus prof. mr. P.F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op maandag 21 december 2009
klokke 15.00 uur

door

Jeroen Franciscus Jacobus Laros
geboren te Den Helder
in 1977

Promotiecommissie

Promotor: prof. dr. J.N. Kok
Co-promotor: dr. W.A. Kusters
Overige leden: prof. dr. Th. Bäck
prof. dr. J.T. den Dunnen (Leids Universitair Medisch Centrum)
dr. H.J. Hoogeboom
prof. dr. X. Liu (Brunel University)
dr. P.E.M. Taschner (Leids Universitair Medisch Centrum)

Contents

1	Introduction	1
1.1	Data Mining	1
1.2	DNA	2
1.3	Metrics	3
1.4	Overview	3
1.5	List of publications	7
I	The Push and Pull Model	
	with applications to criminal career analysis	9
2	Randomised Non-Linear Dimension Reduction	11
2.1	Introduction	11
2.2	The surface	12
2.3	Metric algorithms	12
2.3.1	Forces	14
2.4	Axes	18
2.5	The non-metric variant	20
2.6	Simulated annealing	22
2.7	Comparison with other methods	22
2.8	Conclusions and further research	23
3	Visualisation on a Closed Surface	25
3.1	Introduction	25
3.2	Background	26
3.3	Algorithm	27
3.4	Experiments	30
3.5	Conclusions and further research	33
4	Error Visualisation in the Particle Model	35
4.1	Introduction	35
4.2	Constructing the error map	36
4.2.1	Minimum correction	38
4.3	Experiments	38

4.4	Conclusions and further research	41
5	Temporal Extrapolation Using the Particle Model	43
5.1	Introduction	43
5.2	Parameters	45
5.3	Extrapolation method	46
5.4	Experiments	48
5.5	Conclusions and further research	48
II	Metrics	51
6	Metrics for Mining Multisets	53
6.1	Introduction	53
6.2	Background	54
6.3	The metric	55
6.4	Applications	59
6.5	Conclusions and further research	62
7	Alignment of Multiset Sequences	63
7.1	Introduction	63
7.2	Background	64
7.3	Alignment adaptation	66
7.4	Experiments	69
7.4.1	Criminal careers	69
7.4.2	Access logs	72
7.5	Conclusions and further research	73
III	DNA	77
8	Selection of DNA Markers	79
8.1	Introduction	79
8.2	Combinatorial background	81
8.3	Proximity search and distance selection	81
8.4	Applications	86
8.4.1	Primer pair selection	86
8.4.2	DNA marker selection	87
8.4.3	Other applications	87
8.5	Experiments	88
8.5.1	Finding markers: Determining unique substrings	89
8.5.2	Filtering out simple repeats	90
8.5.3	GC content and temperature	92
8.6	Conclusions and further research	92

9	Substring Differences in Genomes	95
9.1	Introduction	95
9.2	Determining rare factors	96
9.2.1	Conversion	96
9.2.2	Sliding window	97
9.2.3	Counting	97
9.3	Elementary statistics and visualisations	98
9.4	Distances and weights	100
9.5	Experiments and results	102
9.5.1	Raw data	102
9.5.2	Visualisation of the raw data	103
9.5.3	Comparison of many species	105
9.6	A multiset distance measure	107
9.7	Conclusions and further research	109
10	Visualising Genomes in 3D using Rauzy Projections	111
10.1	Introduction	111
10.2	Background	111
10.3	Application to DNA	113
10.4	A number of DNA sequence visualisations	114
10.4.1	Projections in three dimensions	114
10.4.2	Projections in two dimensions	116
10.5	Related work	118
10.6	Conclusions and further research	118
	Bibliography	121
	Nederlandse Samenvatting	127
	Curriculum Vitae	129

