Contents

1 Formal Concept Analysis and Extensions for Complex Data Analytics 1
Léonard Kwuida and Rokia Missaoui
1.1 Introduction 1
1.2 Background 2
1.2.1 Formal Concepts and Line Diagrams 2
1.2.2 Non Binary Data 4
1.2.3 Implication Computation 6
1.3 Extensions to FCA 7
1.3.1 Logical FCA 7
1.3.2 Fuzzy FCA 7
1.3.3 Relational Concept Analysis 7
1.3.4 Triadic Concept Analysis 8
1.3.5 Approximation 9
1.4 Complex Data Analytics 10
1.5 Contributions 11
References 13

2 Conceptual Navigation in Large Knowledge Graphs 17
Sébastien Ferré
2.1 Introduction 17
2.2 Graph-FCA: Extending FCA to Knowledge Graphs 19
2.2.1 Graph Context 19
2.2.2 Graph Patterns 21
2.2.3 Graph Concepts 21
2.2.4 Graph Concept Lattice 23
2.3 Conceptual Navigation in Graph-FCA Lattices 23
2.3.1 Abstract Conceptual Navigation (ACN) 25
2.3.2 Graph-ACN: Instantiating ACN to Knowledge Graphs . . . 27
2.4 Scaling to Large RDF Graphs with SPARQL Endpoints 30
2.4.1 From Graph-FCA to RDF and SPARQL 31

2.4.2 Computing the Result, Index, and Links 32
2.4.3 Living with Partial Results 35
2.5 Rising in Expressivity 35
2.5.1 An Algebraic Form of Queries 36
2.5.2 Extensions of the Query Algebra 37
2.6 The Sparklis Tool and Application Cases 38
2.6.1 Sparklis 38
2.6.2 Application Cases 40
2.7 Conclusion and Perspectives 42
References 43

3 FCA2VEC: Embedding Techniques for Formal Concept Analysis . . . 47
Dominik Dürrschnabel, Tom Hanika, and Maximilian Stubbemann
3.1 Introduction 47
3.2 Related Work 48
3.3 Foundations 49
3.3.1 Formal Concept Analysis 49
3.3.2 Word2Vec 50
3.4 Modeling 52
3.4.1 Retrieving FCA Features Through Closure2Vec 52
3.4.2 Object2Vec and Attribute2Vec 57
3.5 Experiments 60
3.5.1 Object2Vec and Attribute2Vec 61
3.5.2 FCA Features Through Closure2Vec 66
3.6 Conclusion 70
References 71

4 Analysis of Complex and Heterogeneous Data Using FCA and Monadic Predicates 75
Karell Bertet, Christophe Demko, Salah Boukhetta, Jérémy Richard, and Cyril Faucher
4.1 Introduction 75
4.2 The NextPriorityConcept Algorithm 76
4.2.1 Formal Concept Analysis 76
4.2.2 NextPriorityConcept 77
4.3 Use Cases 88
4.3.1 Binary and Categorical Characteristics with the `Lenses` Dataset 88
4.3.2 Numerical Characteristics with the `Iris` Dataset 91
4.3.3 Sequential Characteristics with the `Daily-actions` Dataset 96
4.3.4 Sequential Characteristics with the `Wine City` Dataset . . 98
4.4 Conclusion 101
References 102

5 Dealing with Large Volumes of Complex Relational Data Using RCA 105
Agnès Braud, Xavier Dolques, Alain Gutierrez, Marianne Huchard, Priscilla Keip, Florence Le Ber, Pierre Martin, Cristina Nica, and Pierre Silvie
5.1 Introduction 105
5.2 Background 107
5.3 Related Work 111
5.4 RCA for Environmental Data 114
5.4.1 Two Complex Datasets from the Environmental Domain 114
5.4.2 Experimenting RCA Algorithms 115
5.4.3 Discussion 121
5.5 Analysing Sequences from Water Quality Monitoring Using RCA 122
5.5.1 RCA-Seq 123
5.5.2 Experiments 126
5.5.3 Navigating the Resulting Hierarchy of Graphs 127
5.6 Conclusion 129
References 130

6 Computing Dependencies Using FCA 135
Jaume Baixeries, Victor Codocedo, Mehdi Kaytoue, and Amedeo Napoli
6.1 Introduction 135
6.2 Notation 137
6.2.1 Equivalence Relation 137
6.2.2 Tolerance Relations 138
6.3 FCA and Database Dependencies 140
6.3.1 Functional Dependencies 140
6.3.2 Similarity Dependencies 141
6.3.3 Formal Concept Analysis 141
6.3.4 Functional Dependencies as Implications 142
6.3.5 Pattern Structures 143
6.4 Results 144
6.4.1 Characterization of Functional Dependencies with Pattern Structures 144
6.4.2 Similarity Dependencies 146
6.5 Discussion 146
6.6 Conclusions 148
References 148

7 Leveraging Closed Patterns and Formal Concept Analysis for Enhanced Microblogs Retrieval 151
Meryem Bendella and Mohamed Quafafou
7.1 Introduction 151
7.2 Related Work 152

7.3 FCA-Based Query Expansion 154
7.3.1 Patterns Discovery 154
7.4 Patterns and Word Embeddings Based Query Expansion 155
7.4.1 Word Embeddings: Word2Vec Model 156
7.4.2 Expansion Terms Selection 157
7.5 Experiments 158
7.5.1 Dataset Description 158
7.5.2 Retrieval Model 159
7.5.3 Experimental Protocol 159
7.5.4 Experimental Results 160
7.6 Conclusion 163
References 164

8 Scalable Visual Analytics in FCA 167
Tim Pattison, Manuel Enciso, Ángel Mora, Pablo Cordero, Derek Weber, and Michael Broughton
8.1 Introduction 167
8.1.1 Scalable Visual Analytics in FCA 168
8.1.2 Organisation 169
8.2 Graph-Theoretic Introduction to FCA 169
8.2.1 Formal Context 170
8.2.2 Formal Concepts 170
8.2.3 Concept Lattice Digraph 171
8.2.4 Line Diagram 172
8.2.5 Simplifying Implications 172
8.2.6 Visualising Implications 173
8.2.7 Coordinating Views of Implications and Concepts 174
8.3 Introduction to Visual Analytics 175
8.3.1 Algorithmic Analysis 176
8.3.2 Graph Drawing 177
8.3.3 Information Visualisation 177
8.3.4 Multiple Coordinated Views 178
8.3.5 Tight Coupling 178
8.4 Layout, Visualisation and Interaction 179
8.4.1 Reducing Digraph Size 179
8.4.2 Layout of Line Diagram 179
8.4.3 Interactive Visualisation 180
8.4.4 Discovering or Imposing Tree Structure 181
8.4.5 Demand for Enhanced Tool Support 181
8.4.6 Implications 182
8.5 Three FCA Prototypes 182
8.5.1 Hierarchical Parallel Decomposition 182
8.5.2 User-Guided FCA 183
8.5.3 Structural Navigation 186

8.6 Discovering Insightful Implications 188
8.6.1 Visualisation of Implications 188
8.6.2 Our Data Visualisation Approach 192
8.7 Conclusions and Future Work 196
References 196

9 Formal Methods in FCA and Big Data 201
Domingo López-Rodríguez, Emilio Muñoz-Velasco, and Manuel Ojeda-Aciego
9.1 Introduction 201
9.2 Context and Concept Lattice Reduction Methods 204
9.3 Improved Management of Implications 209
9.4 Minimal Generators to Represent Knowledge 214
9.5 Probably Approximately Correct Implication Bases 216
9.6 Summary and Possible Future Trends 219
References 221

10 Towards Distributivity in FCA for Phylogenetic Data 225
Alain Gély, Miguel Couceiro, and Amedeo Napoli
10.1 Motivation 225
10.2 Models: Lattices, Semilattices, Median Algebras and Median Graphs 227
10.2.1 Lattices and FCA 227
10.2.2 Distributive Lattices 230
10.2.3 Median Graphs 232
10.3 Algorithm to Produce a Distributive $\vee$-Semilattice 233
10.4 A Counter-Example for the Existence of a Minimum Distributive $\vee$-Semilattice 235
10.5 Discussion and Perspectives 236
References 236

11 Triclustering in Big Data Setting 239
Dmitry Egurnov, Dmitry I. Ignatov, and Dmitry Tochilkin
11.1 Introduction 239
11.2 Prime Object-Attribute-Condition Triclustering 241
11.3 Triclustering Extensions 244
11.3.1 Multimodal Clustering 244
11.3.2 Many-Valued Triclustering 245
11.4 Implementations 245
11.4.1 Map-Reduce-Based Multimodal Clustering 245
11.4.2 Implementation Aspects and Used Technologies 248
11.4.3 Parallel Many-Valued Triclustering 249
11.5 Experiments 249
11.5.1 Datasets 250
11.5.2 Results 251

11.6 Experiments with Parallelisation 253
11.7 Conclusion 254
References 256

Index 259