MORGAN & CLAYPOOL PUBLISHERS

Community Search over Big Graphs

Xin Huang
Laks V.S. Lakshmanan
Jianliang Xu

SYNTHESIS LECTURES ON DATA MANAGEMENT

H.V. Jagadish, Series Editor

Contents

Acknowledgments . . . . . **xvii**

1 Introduction **1**

1.1 Graphs and Communities 1
1.1.1 Graphs 1
1.1.2 Communities 1
1.2 Community Search 2
1.2.1 Community Search Problem 2
1.2.2 A Comparison with Community Detection 4
1.2.3 Applications 5
1.2.4 Datasets and Tools 6
1.3 Prerequisite and Target Reader 6
1.4 Outline of the Book 6

2 Cohesive Subgraphs **9**

2.1 Community Search and Cohesive Subgraphs 9
2.2 Notations and Notions 10
2.2.1 Graphs and Subgraphs 10
2.2.2 Degree and Neighbors 11
2.2.3 Path, Cycle, Connectivity, and Diameter 11
2.3 Classical Dense Subgraphs 12
2.3.1 Clique and Quasi-Clique 13
2.3.2 k-DBDSG, k-clan, k-club, and k-plex 13
2.4 k-core and k-truss 16
2.4.1 k-core 16
2.4.2 k-truss 18
2.5 More Dense Subgraphs 20
2.5.1 Densest Subgraphs 20
2.5.2 k-ecc and k-vcc 22
2.5.3 Other Dense Subgraphs 23
2.6 Summary 24

3 Cohesive Community Search 27
3.1 Quasi-Clique Community Models 27
3.1.1 Clique-Based Community Detection 28
3.1.2 Quasi-Clique-Based Community Search 30
3.2 Core-Based Community Models 34
3.2.1 Maximum-Core Community Search 35
3.2.2 Minimum-Sized k-Core Community Search 39
3.2.3 Influential Community Search 43
3.2.4 Comparison of Various k-core Community Models 51
3.3 Truss-Based Community Models 52
3.3.1 Triangle-Connected Truss Community Search 52
3.3.2 Closest Truss Community Search 65
3.4 Query-Biased Densest Community Model 73
3.4.1 Notions and Notations 73
3.4.2 Problem Formulation 75
3.4.3 Algorithms 75
3.5 Summary 78

4 Attributed Community Search 83
4.1 Introduction 83
4.1.1 Attributed Networks 83
4.1.2 Limitations of Cohesive Community Search without Query Attributes 85
4.1.3 Desiderata of Good Attributed Communities 86
4.2 K-Core-Based Attribute Community Model 87
4.2.1 Problem Formulation 87
4.2.2 Basic Query Processing Algorithm 89
4.2.3 CLTree-Index-Based Query Processing Algorithms 89
4.3 K-Truss-Based Attribute Community Model 94
4.3.1 (k, d)-Truss 94
4.3.2 Attribute Score Function 95
4.3.3 Attributed Truss Community Model 97
4.3.4 ATindex-Based Greedy Algorithm 98
4.4 Summary 102
4.4.1 Case Study on the DBLP Network 102
4.4.2 Case Study on the PPI Network with Ground-Truth Communities 103
4.4.3 Comparison between ACC and ATC Models 104
4.4.4 Comparison with Other Related Works 105

5 Social Circle Analysis . . . 109
5.1 Ego-Networks . . . 109
5.2 Structural Diversity Search . . . 110
5.2.1 Motivations . . . 110
5.2.2 Problem Formulation . . . 111
5.2.3 A Simple Degree-Based Approach . . . 112
5.2.4 A Novel Top-k Search Framework . . . 114
5.2.5 Case Studies . . . 122
5.3 Learning to Discover Social Circles . . . 126
5.3.1 Attributed Community Search and Social Circle Discovery . . . 127
5.3.2 Problem Formulation . . . 127
5.3.3 A Generative Model for Social Circle Discovery . . . 128

6 Geo-Social Group Search . . . 131
6.1 Geo-Social Group Search . . . 131
6.2 Proximity-Based Geo-Social Group Search . . . 133
6.2.1 Problem Statement . . . 133
6.2.2 R-Tree-Based Query Processing . . . 135
6.2.3 Social-Aware R-Tree . . . 136
6.2.4 SaR-Tree-Based Query Processing . . . 140
6.3 Geo-Social k-Cover Group Search . . . 143
6.3.1 Problem Statement . . . 143
6.3.2 Algorithms . . . 145
6.4 Geo-Social Group Search Based on Minimum Covering Circle . . . 147
6.4.1 Problem Statement . . . 147
6.4.2 Algorithms . . . 149

7 Datasets and Tools . . . 153
7.1 Real-World Datasets . . . 153
7.1.1 Networks with Ground-Truth Communities . . . 153
7.1.2 Attributed Graphs with Ground-Truth Communities . . . 154
7.1.3 Ego-Networks with Ground-Truth Social Circles . . . 155
7.1.4 Geo-Social Networks . . . 156
7.1.5 Public-Private Collaboration Networks . . . 157
7.2 Query Generation and Evaluation . . . 158
7.2.1 Query Generation . . . 158
7.2.2 Evaluation Metrics . . . 158

7.3 Software and Demo Systems 159
7.4 Suggestions on Dense Subgraph Selection for Community Models 160

8 Further Readings and Future Directions 163

8.1 Further Readings 163
8.1.1 Clique-Based Community Search 163
8.1.2 Core-Based Community Search 163
8.1.3 Truss-Based Community Search 164
8.1.4 Plex-Based Community Search 164
8.1.5 Others 164
8.2 Future Directions and Open Problems 165
8.2.1 Querying Communities on Heterogeneous Information Networks . 165
8.2.2 Scalable Algorithms for Big Graphs 166
8.2.3 Public-Private Social Networks 166
8.2.4 Community Search on Probabilistic Graphs 166
8.2.5 Applications and Case Studies 167
8.3 Conclusions 167

Bibliography 169

Authors' Biographies 187