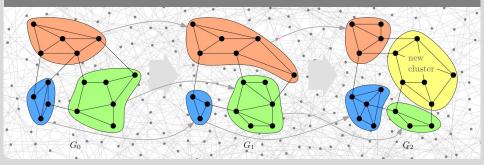


Exercises AE for Graph Clustering

School on Graph Theory, Algorithms and Applications

Dorothea Wagner | Erice, Italy, 25. September - 3. October, 2011

KARLSRUHE INSTITUTE OF TECHNOLOGY – INSTITUTE OF THEORETICAL INFORMATICS



St	ructure	
1	Introduction	Karlsruher Institut für Technologie
	Scenario: Network Analysis	
	Paradigm of Clustering	
	Example Applications	
2	Formalization of Aims and Objectives	
-	 Objective Functions 	
3	Algorithmic Approaches	
-	Greedy Merge	
	Local Moving and Multilevel	
	Clustering with Minimum-Cut Tree	
	Integer Linear Programs	
	 Other Algorithmic Approaches 	
4	Experimental Evaluation	
	The Role of Test Data in Algorithm Engineering	
	Comparing Clusterings	
5	Dynamic Graph Clustering	1 D >
	Online Dynamic Clustering	< ∰ >
	 Offline Dynamic Clustering 	< 문 → 2 일 ↓
6	Appendix	 IE
		500

Experimental Evaluation Dynamic Graph Clustering Appendix

Erice, Italy, 25. September - 3. October, 2011



	cture			Karlsruher I	KIT Istitut für Technologie
0	Scenario: Network Anal Paradigm of Clustering Example Applications	ysis			
	ormalization of Aims and Objective Functions	Objectives			
4 E>	gorithmic Approaches Greedy Merge Local Moving and Multil Clustering with Minimur Integer Linear Programs Other Algorithmic Appro- operimental Evaluation The Role of Test Data in	n-Cut Tree s baches	ineering		
5 Dy	Comparing Clusterings ynamic Graph Clustering Online Dynamic Cluster Offline Dynamic Cluster	ring			
	opendix	AL			うくぐ
Introduction	Formalization of Aims and Objectives	Algorithmic Approaches	Experimental Evaluation	Dynamic Graph Clustering	Appendix

Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

Erice, Italy, 25. September - 3. October, 2011 3/28



Introduction
Scenario: Network Analysis
Paradigm of Clustering
Example Applications
Formalization of Aims and Objectives
 Objective Functions
Algorithmic Approaches
Greedy Merge
Local Moving and Multilevel
Clustering with Minimum-Cut Tree
Integer Linear Programs
 Other Algorithmic Approaches
Experimental Evaluation
The Role of Test Data in Algorithm Engineering
Comparing Clusterings
Dynamic Graph Clustering
 Online Dynamic Clustering
 Offline Dynamic Clustering
Appendix

Experimental Evaluation Dynamic Graph Clustering

∽ ९ (~ Appendix

Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

Erice, Italy, 25. September - 3. October, 2011 4/28

Str	Introduction	Kardsruher ins	KIT titut für Technologie
	 Scenario: Network Analysis Paradigm of Clustering Example Applications 		
	 Example Applications Formalization of Aims and Objectives Objective Functions 		
	Algorithmic Approaches Greedy Merge Local Moving and Multilevel Clustering with Minimum-Cut Tree Integer Linear Programs Other Algorithmic Approaches		
4	Experimental Evaluation • The Role of Test Data in Algorithm Engineering • Comparing Clusterings		
	 Dynamic Graph Clustering Online Dynamic Clustering Offline Dynamic Clustering Appendix 		
Introduct		Dynamic Graph Clustering	୬ ୯.୧ Appendix

Requirements for Clusterings



Consider reasonable desiderata for a quality measure for clusterings and try to design a measure that fulfills them.

Algorithmic Approaches

Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

Formalization of Aims and Objectives

Introduction

Erice, Italy, 25. September – 3. October, 2011

Dynamic Graph Clustering

Experimental Evaluation

Appendix 6/28



Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

Erice, Italy, 25. September – 3. October, 2011 7/28



roduct	ion Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	Appendix
	Appendix	しゃい
	Offline Dynamic Clustering	< <u>₹</u> >
		- ₹ ≣ >
	Online Dynamic Clustering	
	Dynamic Graph Clustering	< • • •
	Comparing Clusterings	
	The Role of Test Data in Algorithm Engineering	
	Experimental Evaluation	
	Other Algorithmic Approaches	
	Integer Linear Programs Other Algorithmic Approaches	
	Clustering with Minimum-Cut Tree	
	Local Moving and Multilevel	
	Greedy Merge	
	Algorithmic Approaches	
	Objective Functions	
2	Formalization of Aims and Objectives	
_	Example Applications	
	Paradigm of Clustering	
	• Scenario. Network Analysis	

Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

Sooparia: Natwork Analysis

Erice, Italy, 25. September - 3. October, 2011 8/28



	 Paradigm of Clustering 	
	Example Applications	
2	Formalization of Aims and ObjectivesObjective Functions	
	Algorithmic Approaches Greedy Merge Local Moving and Multilevel Clustering with Minimum-Cut Tree Integer Linear Programs Other Algorithmic Approaches 	
	Experimental EvaluationThe Role of Test Data in Algorithm EngineeringComparing Clusterings	
	 Dynamic Graph Clustering Online Dynamic Clustering Offline Dynamic Clustering Appendix 	
roduc	tion Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	∽ ९ (~ Appendix

Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

Sooparia: Natwork Analysis

Erice, Italy, 25. September - 3. October, 2011 9/28

Properties of Objective Functions

Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)



- Consider the desiderata and the objective functions mentioned in the lecture. Try to discover violations.
- How large is the best cluster editing set you can find for Zachary's Karate Club? At most 54 should be doable ...
- For some family of sparse graphs, prove that the modularity of the modularity-optimal clustering approaches 1.
- Prove that plugging performance instead of coverage into the concept of modularity yields an equivalent measure.
- Prove that Density-Constrained Clustering combining maximum, global or average intra-cluster density as a constraint with the number of inter-cluster edges as the objective function is NP-hard (i.e., finding the optimal clustering).

Erice, Italy, 25. September - 3. October, 2011

In

	T
Karlsruher Institut für Teo	hnologie

 Online Dynamic Clustering Offline Dynamic Clustering Oppendix 	
	< ₹≣≯ ₹≣≯
	< ≣ >
Online Dynamic Clustering	
Dynamic Graph Clustering	• • • •
xperimental Evaluation	
Other Algorithmic Approaches	
Integer Linear Programs	
Clustering with Minimum-Cut Tree	
Local Moving and Multilevel	
formalization of Aims and Objectives	
Example Applications	
Paradigm of Clustering	
Scenario: Network Analysis	
	 Paradigm of Clustering Example Applications ormalization of Aims and Objectives Objective Functions Idgorithmic Approaches Greedy Merge Local Moving and Multilevel Clustering with Minimum-Cut Tree Integer Linear Programs

Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

Erice, Italy, 25. September – 3. October, 2011 11/28

	T
Karlsruher Institut für Teo	hnologie

	Scenario: Network Analysis	
	Paradigm of Clustering	
	• Example Applications	
	Formalization of Aims and Objectives	
	Objective Functions	
3	Algorithmic Approaches	
	Greedy Merge	
	 Local Moving and Multilevel 	
	 Clustering with Minimum-Cut Tree 	
	Integer Linear Programs	
	Other Algorithmic Approaches	
	Experimental Evaluation	
	 The Role of Test Data in Algorithm Engineering 	
	 Comparing Clusterings 	
	Dynamic Graph Clustering	< • • •
	 Online Dynamic Clustering 	
	Offline Dynamic Clustering	 ₹ Ξ + 1 = 1
	Appendix	< 문 → 문
	Арреник	500
troduc	tion Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	Appendix

Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

Erice, Italy, 25. September – 3. October, 2011 12/28

Greedy Merge Modularity



- Consider the global greedy agglomerative algorithm, using modularity as the objective function. Prove that modularity attains a single peak during the process, i.e., once it deteriorates, it never again improves.
- Prove that there is always a modularity-optimal clustering where each degree-1 vertex is in the same cluster as its neighbor.

Greedy Merge General



 Prove the following reachability result: Given density α, clustering C and a subset D ⊆ C of dense clusters: ∀D ∈ D : density(D) ≥ α. If D's union U = U_{D∈D} D is dense, i.e., density(U) ≥ α, then there exist A, B ∈ D such that density(A ∪ B) ≥ α.

Prove that maximum isolated (and pairwise) inter-cluster density have unbounded merge behavior (actually you can try to prove the whole list!).



troduc	tion Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	Appendix
	Appendix	しゃい
	• Online Dynamic Gustering	< ≞ >
	Offline Dynamic Clustering	< ≣⇒
	Online Dynamic Clustering	
	Dynamic Graph Clustering	< • • •
	Comparing Clusterings	
	The Role of Test Data in Algorithm Engineering	
	Experimental Evaluation	
	 Other Algorithmic Approaches 	
	Integer Linear Programs	
	Clustering with Minimum-Cut Tree	
	Local Moving and Multilevel	
	Greedy Merge	
	Algorithmic Approaches	
3		
	Objective Functions	
	Formalization of Aims and Objectives	
	• Example Applications	
	Paradigm of Clustering	
	Scenario: Network Analysis	

Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

Erice, Italy, 25. September - 3. October, 2011 15/28

Data Structures



Local movements usually repeatedly require the following information: What is the sum of the weights of all edges between a given node and a given cluster? Avoiding hashmaps where possible, think about how to implement this efficiently and take into account the time required for array initialization.

Appendix

Introduction Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation

Dynamic Graph Clustering



 2 Formalization of Aims and Objectives Objective Functions 3 Algorithmic Approaches Greedy Merge Local Moving and Multilevel Clustering with Minimum-Cut Tree Integer Linear Programs Other Algorithmic Approaches 4 Experimental Evaluation The Role of Test Data in Algorithm Engineering Comparing Clusterings 5 Dynamic Graph Clustering Online Dynamic Clustering Online Dynamic Clustering 	↓ □ ○
Offline Dynamic Clustering Appendix Appendix Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	ৰ ≣ ⊧ ≣ প্©্

Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

Erice, Italy, 25. September – 3. October, 2011 17/28

Bounds Guaranteed by Min-Cuts



Preliminaries

Let G = (V, E, c) denote the input graph to the Cut-Clustering Algorithm and let $G_{\alpha} = (V \cup \{t\}, E \cup E', c_{\alpha})$ result from *G* by adding the artificial vertex *t* together with the artificial edges weighted by α . A cut in G_{α} is denotes by (X, \overline{X}) with $X \subseteq V \cup \{t\}$. A cut in *G* is denoted by $(X, V \setminus X)$ with $X \subseteq V$. The notation (P, Q) with $P, Q \subseteq V, P \cap Q = \emptyset$ describes a cut in the subgraph of *G* induced by $P \uplus Q$.

Problem 1: Let C denote the clustering returned by the Cut-Clustering Algorithm with respect to α and let $C \in C$ denote a cluster. Show that the intra-cluster expansion

$$\Phi(C) := \min_{P, Q \neq \emptyset, P \uplus Q = C} \left\{ \frac{c(P, Q)}{\min |P|, |Q|} \right\}$$

of C is at least α .

Problem 2: Let C denote the clustering returned by the Cut-Clustering Algorithm with respect to α and let $C \in C$ denote a cluster. Show that the inter-cluster expansion*

$$\Phi^*(C) := \left\{ \frac{c(C, V \setminus C)}{|V \setminus C|} \right\}$$

of C is at most α .

duction Formalization of Aims and Objectives Algorithmic Approaches

es Experimental Evaluation

Dynamic Graph Clustering Appendix

↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓
 ↓

Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

Erice, Italy, 25. September - 3. October, 2011 18/28



troducti	on Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	Annendix
	Appendix	しゃい
	 Offline Dynamic Clustering 	< ≣ >
	Online Dynamic Clustering	
	Dynamic Graph Clustering	• • • •
	Comparing Clusterings	
	The Role of Test Data in Algorithm Engineering	
	Experimental Evaluation	
	Other Algorithmic Approaches	
	 Integer Linear Programs 	
	 Clustering with Minimum-Cut Tree 	
	Local Moving and Multilevel	
	Greedy Merge	
3	Algorithmic Approaches	
	Objective Functions	
	Formalization of Aims and Objectives	
	Example Applications	
	Paradigm of Clustering	
	Scenario: Network Analysis	

Variables and Constraints

Kalsruher Institut für Technologie

20/28

Starting with the *X*-variables from the lecture try to build the following tools:

- cluster-leaders: a binary variable stating whether a vertex is the unique leader of its cluster
- $|\mathcal{C}| \leq \text{given value}$
- **(3)** containment: binary variables stating whether vertex v is in cluster C_i

Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

- Iuster-size variables: binary variable stating whether $|C_i| = \ell$
- $\forall C \in C$: density $(C) \ge$ given value

Erice, Italy, 25. September - 3. October, 2011

In



troduct	ion Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	Appendix
	Appendix	きょう
		< ≣ >
	Offline Dynamic Clustering	< ≣ >
	Online Dynamic Clustering	< @ >
	Dynamic Graph Clustering	< • • •
	Comparing Clusterings	
	The Role of Test Data in Algorithm Engineering	
	Experimental Evaluation	
	 Other Algorithmic Approaches 	
	Integer Linear Programs	
	Clustering with Minimum-Cut Tree	
	 Local Moving and Multilevel 	
	 Greedy Merge 	
3	Algorithmic Approaches	
	Objective Functions	
	Formalization of Aims and Objectives	
	• Example Applications	
	Paradigm of Clustering	
	Scenario: Network Analysis	

Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

Erice, Italy, 25. September - 3. October, 2011 21/28

Introductio

on	Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	Appendix
A[opendix	50
		~ 문
	Offline Dynamic Clustering	< ह < ह
	Online Dynamic Clustering	4 🗗
Dy	ynamic Graph Clustering	< □
0	Comparing Clusterings	
	The Role of Test Data in Algorithm Engineering	
	perimental Evaluation	
	Other Algorithmic Approaches	
	Integer Linear Programs	
	Clustering with Minimum-Cut Tree	
	Local Moving and Multilevel	
	Greedy Merge	
	gorithmic Approaches	
	Objective Functions	
Fc	ormalization of Aims and Objectives	
•	Example Applications	
•	Paradigm of Clustering	

Scenario: Network Analysis

	Scenario: Network Analysis	
	Paradigm of Clustering	
	• Example Applications	
	Formalization of Aims and Objectives	
	Objective Functions	
	Algorithmic Approaches	
	Greedy Merge	
	 Local Moving and Multilevel 	
	 Clustering with Minimum-Cut Tree 	
	Integer Linear Programs	
	Other Algorithmic Approaches	
4	Experimental Evaluation	
	The Role of Test Data in Algorithm Engineering	
	Comparing Clusterings	
	Dynamic Graph Clustering	< • • •
	Online Dynamic Clustering	< @ >
	 Offline Dynamic Clustering 	≪ 분 ►
	Appendix	- - = → - ==
	Appendix	500
oduc	tion Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	Appendix



Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

Erice, Italy, 25. September - 3. October, 2011 23/28

Intr

oduct	tion Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	Appendix
	Appendix	き
	Offline Dynamic Clustering	< ≣
	Online Dynamic Clustering	< ₽ < Ξ
	Dynamic Graph Clustering	4 🗆
	Comparing Clusterings	
	The Role of Test Data in Algorithm Engineering	
4	Experimental Evaluation	
_	 Other Algorithmic Approaches 	
	Integer Linear Programs	
	Clustering with Minimum-Cut Tree	
	Local Moving and Multilevel	
	 Greedy Merge 	
	Algorithmic Approaches	
_	 Objective Functions 	
	Formalization of Aims and Objectives	
	 Example Applications 	
	Paradigm of Clustering	
	C. Devedient of Chustering	



Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

Scenario: Network Analysis

Erice, Italy, 25. September - 3. October, 2011 24/28

In

troduc	tion Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	Appendix
	Appendix	しょつ
		<
	Offline Dynamic Clustering	≥ →
	 Online Dynamic Clustering 	
5	Dynamic Graph Clustering	< • • •
_	Comparing Clusterings	
	The Role of Test Data in Algorithm Engineering	
	Experimental Evaluation	
	 Other Algorithmic Approaches 	
	Integer Linear Programs	
	Clustering with Minimum-Cut Tree	
	Local Moving and Multilevel	
	Greedy Merge	
	Algorithmic Approaches	
	Objective Functions	
	Formalization of Aims and Objectives	
	• Example Applications	
	 Paradigm of Clustering 	
	Scenario: Network Analysis	

troduct	ion Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	Appendix
	Appendix	しょう
		< 문 → 문
	Offline Dynamic Clustering	< ≞ >
	Online Dynamic Clustering	< 🗗 >
5	Dynamic Graph Clustering	< • • •
	Comparing Clusterings	
	 The Role of Test Data in Algorithm Engineering 	
	Experimental Evaluation	
	Other Algorithmic Approaches	
	Integer Linear Programs Others Algorithmic Augustations	
	Clustering with Minimum-Cut Tree	
	Local Moving and Multilevel	
	Greedy Merge	
	Algorithmic Approaches	
	Objective Functions	
	Formalization of Aims and Objectives	
	 Example Applications 	
	Paradigm of Clustering	
	Scenario: Network Analysis	



Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

Erice, Italy, 25. September - 3. October, 2011 26/28

ntroduc	ction Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	Appendix
	Appendix	しょう
	 Offline Dynamic Clustering 	< ≣ >
	 Online Dynamic Clustering 	< @ > < ≥ >
5	Dynamic Graph Clustering	${}^{+}\Box \rightarrow$
	Comparing Clusterings	
	The Role of Test Data in Algorithm Engineering	
	Experimental Evaluation	
	Other Algorithmic Approaches	
	Integer Linear Programs Other Algorithmic Approaches	
	Clustering with Minimum-Cut Tree	
	 Greedy Merge Local Moving and Multilevel 	
	Algorithmic Approaches	
	Objective Functions	
	Formalization of Aims and Objectives	
	• Example Applications	
	Paradigm of Clustering	
	Scenario: Network Analysis	



Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

Erice, Italy, 25. September - 3. October, 2011 27/28

troducti	tion Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering A	nnendix	
6	Appendix	きょう	
		< ≞ >	
	Offline Dynamic Clustering	< ≡ >	
	Online Dynamic Clustering	< 67 ▶	
	Dynamic Graph Clustering	< D >	
_	Comparing Clusterings		
	The Role of Test Data in Algorithm Engineering		
	Experimental Evaluation		
	Other Algorithmic Approaches		
	Integer Linear Programs Other Algorithmic Approaches		
	 Clustering with Minimum-Cut Tree 		
	 Local Moving and Multilevel 		
	Greedy Merge		
	Algorithmic Approaches		
	 Objective Functions 		
	Formalization of Aims and Objectives		
	 Example Applications 		
	Paradigm of Clustering		
	Scenario: Network Analysis		
	Scenario: Netw	ork Analysis	ork Analysis



Dorothea Wagner - Exercises (Algorithm Engineering for Graph Clustering)

Erice, Italy, 25. September - 3. October, 2011 28/28