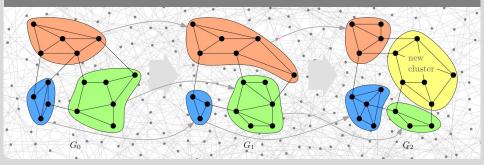


# Hints AE for Graph Clustering

School on Graph Theory, Algorithms and Applications

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

KARLSRUHE INSTITUTE OF TECHNOLOGY - INSTITUTE OF THEORETICAL INFORMATICS



Str	ucture	
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	
	<ul> <li>Clustering with Minimum-Cut Tree</li> </ul>	
	Integer Linear Programs	
_	<ul> <li>Other Algorithmic Approaches</li> </ul>	
4	Experimental Evaluation	
	The Role of Test Data in Algorithm Engineering	
_	Comparing Clusterings	
5	Dynamic Graph Clustering	< □ >
	Online Dynamic Clustering	< ⊕ >
_	<ul> <li>Offline Dynamic Clustering</li> </ul>	< 토 > < 토 >
6	Appendix	< 토 → 토

Experimental Evaluation Dynamic Graph Clustering



	icture			Kaitsruher ite	
	ntroduction Scenario: Network Anal	ysis			
0	<ul> <li>Paradigm of Clustering</li> <li>Example Applications</li> </ul>				
	ormalization of Aims and Objective Functions	Objectives			
3 A	Igorithmic Approaches				
9	<ul> <li>Greedy Merge</li> <li>Local Moving and Multil</li> </ul>	evel			
0	<ul> <li>Clustering with Minimur</li> <li>Integer Linear Programs</li> </ul>				
	Other Algorithmic Appro				
4 E	Experimental Evaluation The Role of Test Data in Comparing Clusterings	n Algorithm Eng	neering		
	ynamic Graph Clustering Online Dynamic Cluster				< □ > < @ >
•	Offline Dynamic Cluster	0			< 토 → < 토 →
6 A	ppendix				しょう
Introduction	Formalization of Aims and Objectives	Algorithmic Approaches	Experimental Evaluation	Dynamic Graph Clustering	Appendix

Dorothea Wagner - Hints (Algorithm Engineering for Graph Clustering)

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



< □ > < 🗇 > < ∃ > < ∃⇒ E 500

Introduction
Scenario: Network Analysis
Paradigm of Clustering
<ul> <li>Example Applications</li> </ul>
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

Algorithmic Approaches

Structure

Introduction

Str	ucture	Katisuher In	
	<ul><li>Scenario: Network Analysis</li><li>Paradigm of Clustering</li></ul>		
	<ul> <li>Example Applications</li> <li>Formalization of Aims and Objectives</li> <li>Objective Functions</li> </ul>		
	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		
	<ul><li>Dynamic Graph Clustering</li><li>Online Dynamic Clustering</li><li>Offline Dynamic Clustering</li><li>Appendix</li></ul>		- 4 4 4 - 4 4 - 4 4 - 4 4 4 4 - 4 4 4 4 - 4 4 4 - 4 4 4 - 4 - 4
Introducti	ion Formalization of Aims and Objectives Algorithmic Approaches Experimental Evalua	tion Dynamic Graph Clustering	Appendix

Dorothea Wagner - Hints (Algorithm Engineering for Graph Clustering)

Erice, Italy, 25. September - 3. October, 2011 5/28

#### **Requirements for Clusterings**



Start with a small set of desiderata, or try the converse: it might be more contradicting combinations of desiderata

Algorithmic Approaches

Appendix

Formalization of Aims and Objectives Dorothea Wagner - Hints (Algorithm Engineering for Graph Clustering)

Introduction

Erice, Italy, 25. September - 3. October, 2011 6/28

Dynamic Graph Clustering

Experimental Evaluation





roduct	tion Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	Appendix
	Appendix	三 う く (や
		< ≞ >
	Offline Dynamic Clustering	< ≣ >
	Online Dynamic Clustering	< 🗗 >
	Dynamic Graph Clustering	< • • •
	Comparing Clusterings	
	<ul> <li>The Role of Test Data in Algorithm Engineering</li> </ul>	
	Experimental Evaluation	
	0	
	<ul> <li>Other Algorithmic Approaches</li> </ul>	
	<ul> <li>Integer Linear Programs</li> </ul>	
	Clustering with Minimum-Cut Tree	
	Local Moving and Multilevel	
	Greedy Merge	
	Algorithmic Approaches	
	<ul> <li>Objective Functions</li> </ul>	
2	Formalization of Aims and Objectives	
	Example Applications	
	<ul> <li>Paradigm of Clustering</li> </ul>	
	Scenario: Network Analysis	

Dorothea Wagner - Hints (Algorithm Engineering for Graph Clustering)

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



	• Scenario. Network Analysis	
	Paradigm of Clustering	
	Example Applications	
2	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	
	<ul> <li>The Role of Test Data in Algorithm Engineering</li> </ul>	
	<ul> <li>Comparing Clusterings</li> </ul>	
	Dynamic Graph Clustering	
	Online Dynamic Clustering	
	Offline Dynamic Clustering	< E > < E >
	Appendix	~ 분기
	hbellery	500
roduct	ion Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	Appendix

Dorothea Wagner - Hints (Algorithm Engineering for Graph Clustering)

Sooparia: Natwork Analysis

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

## **Properties of Objective Functions**



10/28

- Consider the desiderata and the objective functions mentioned in the lecture. Try to discover violations.
- Don't attempt to be optimal, unless you want to spend a lot of time on it! Something close to matching is pretty good in sparse networks, start with that.
- You just need the modularity of any good clustering, this is a lower bound for the optimum. Make up a clustering on a scalable sparse graph, calculate or bound its modularity. A simple quadratic grid might be a simple family.
- Strenuous and technical.

Dorothea Wagner - Hints (Algorithm Engineering for Graph Clustering)

Setting the density constraint to 1, the decision version is equivalent to the question: (min-cut clique partition) Is there a partition C of V into cliques such that the number of intra-cluster edges ist at least k? Try to reduce from the following problem:

*Exact Cover by 3-Sets (X3C)*: Given set *X* with |X| = 3q and collection *S* of 3-element subsets of *X*. Does *S* contain an exact cover for *X*, i.e., a subcollection  $S' \subseteq S$  such that every  $x \in X$  occurs in exactly one  $S \in S'$ ?

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



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



stroduc	ion Examplization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	Appondix
	Appendix	- - -
		< 토 → 토
	Offline Dynamic Clustering	<ul> <li>₹ Ξ ►</li> </ul>
	Online Dynamic Clustering	< <i>₫</i> >
	Dynamic Graph Clustering	+ - +
	Comparing Clusterings	
	The Role of Test Data in Algorithm Engineering	
	Experimental Evaluation	
	<ul> <li>Other Algorithmic Approaches</li> </ul>	
	Integer Linear Programs	
	Clustering with Minimum-Cut Tree	
	Local Moving and Multilevel	
	<ul> <li>Greedy Merge</li> </ul>	
3	Algorithmic Approaches	
	<ul> <li>Objective Functions</li> </ul>	
	Formalization of Aims and Objectives	
	• Example Applications	
	<ul> <li>Paradigm of Clustering</li> </ul>	
	Scenario: Network Analysis	

## Greedy Merge Modularity



Carefully note how entries in the merge matrix change. First find out what  $\Delta \text{mod}_{i,k}$  is, i.e., the change in modularity if  $C_i$  and  $C_k$  are merged. Then, given  $C_i$  and  $C_i$  are merged, show that  $\Delta \text{mod}_{(ii),k} = \Delta \text{mod}_{i,k} + \Delta \text{mod}_{i,k}$ Based on these observations and on how the greedy algorithm works, complete the argument.

Algorithmic Approaches

< 🗇 > < ∃ → < ∃ → Э nan

Appendix

Dynamic Graph Clustering

Experimental Evaluation

#### Greedy Merge General



- A vectorial point of view is one way of proving this. The key observation here is that when merging clusters, the resulting density is the vectorial (or component-wise) addition of the numerators and the denominators of the summands.
- Solutions can be found in [Schumm et al.: Density-constrained Graph clustering, 2011]

Algorithmic Approaches

Appendix

Dynamic Graph Clustering

Experimental Evaluation

#### Structure Introduction

	T
Karlsruher Institut für Teo	hnologie

	<ul><li>Scenario: Network Analysis</li><li>Paradigm of Clustering</li></ul>	
	• 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	
	<ul> <li>Other Algorithmic Approaches</li> </ul>	
	Experimental Evaluation	
	<ul> <li>The Role of Test Data in Algorithm Engineering</li> </ul>	
	<ul> <li>Comparing Clusterings</li> </ul>	
	Dynamic Graph Clustering	
	Online Dynamic Clustering	< ₫ >
	<ul> <li>Offline Dynamic Clustering</li> </ul>	< ≣ → < ≣ →
	Appendix	=
Introduct	tion Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	Appendix

#### **Data Structures**



Do not re-initialize an array repeatedly, re-use the array but keep track of which entries are valid.

Introduction Formalization of Aims and Objectives Algorithmic Approaches

Experimental Evaluation

Dynamic Graph Clustering Appendix

In

	Т
Karlsruher Institut für Tec	hnologie

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

Dorothea Wagner - Hints (Algorithm Engineering for Graph Clustering)

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

#### **Bounds Guaranteed by Min-Cuts**



**Problem 1:** We prove the assertion by contradiction. Suppose there is a cut (P, Q) in *C* such that  $c(P, Q) < \alpha \min\{|P|, |Q|\}$ . Without loss of generality we assume the representative r(C) is in *P*. Then it is

$$\begin{array}{lll} c_{\alpha}(P,\bar{P}) &=& c(P,Q) + c(P,V\setminus C) + \alpha \, |P| \\ &<& \alpha \, |Q| + c(P,V\setminus C) + \alpha \, |P| \\ &\leq& \alpha \, |Q| + c(P,V\setminus C) + \alpha \, |P| + c(Q,V\setminus C) \\ &=& \alpha \, |C| + c(C,V\setminus C) = c_{\alpha}(C,\bar{C}). \end{array}$$

This contradicts the fact that  $(C, \overline{C})$  is a minimum r(C)-t-cut in  $G_{\alpha}$ .

1

**Problem 2:** We prove the assertion by contradiction. Suppose  $c(C, V \setminus C) > \alpha |V \setminus C|$ . Then it is

$$c_{\alpha}(\{t\}, V) = \alpha |V \setminus C| + \alpha |C|$$

$$< c(C, V \setminus C) + \alpha |C|$$

$$= c_{\alpha}(C, \overline{C}).$$

This contradicts the fact that  $(C, \overline{C})$  is a minimum r(C)-t-cut in  $G_{\alpha}$ .

Introduction Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering Appendix Dorothea Wagner – Hints (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	∽ Q (~ Appendix
	Appendix	王
	<ul> <li>Offline Dynamic Clustering</li> </ul>	< ≣ >
	Online Dynamic Clustering	< ⊡ > < ⊇ >
	Dynamic Graph Clustering	• • • •
	Comparing Clusterings	
	The Role of Test Data in Algorithm Engineering	
	I Contraction of the second	
	Experimental Evaluation	
	Other Algorithmic Approaches	
	<ul> <li>Integer Linear Programs</li> </ul>	
	<ul> <li>Clustering with Minimum-Cut Tree</li> </ul>	
	<ul> <li>Local Moving and Multilevel</li> </ul>	
	Greedy Merge	
3	Algorithmic Approaches	
	Objective Functions	
	Formalization of Aims and Objectives	
	Example Applications	
	<ul> <li>Paradigm of Clustering</li> </ul>	
	Scenario: Network Analysis	



#### The last item is a bit more difficult, please find suggestions for solutions in [Schumm et al.: Density-constrained graph clustering, 2011, full technical report version]

Algorithmic Approaches

< 67 ▶ < ∃ > < ∃⇒ ∋ nac

Appendix

Formalization of Aims and Objectives Dorothea Wagner - Hints (Algorithm Engineering for Graph Clustering)

Erice, Italy, 25. September - 3. October, 2011 20/28

Dynamic Graph Clustering

Experimental Evaluation

In

		1	
Karlsruher In	stitut für '	Technolog	gle

	Appendix	200
	Appendix	< 문 → 토
	<ul> <li>Offline Dynamic Clustering</li> </ul>	<ul> <li>₹ Ξ +</li> </ul>
	Online Dynamic Clustering	
	Dynamic Graph Clustering	< • • •
	Comparing Clusterings	
	The Role of Test Data in Algorithm Engineering	
	Experimental Evaluation	
	<ul> <li>Other Algorithmic Approaches</li> </ul>	
	Integer Linear Programs	
	Clustering with Minimum-Cut Tree	
	Local Moving and Multilevel	
	<ul> <li>Greedy Merge</li> </ul>	
3	Algorithmic Approaches	
_	<ul> <li>Objective Functions</li> </ul>	
	Formalization of Aims and Objectives	
	<ul> <li>Example Applications</li> </ul>	
	Paradigm of Clustering	
	Scenario: Network Analysis	

Dorothea Wagner - Hints (Algorithm Engineering for Graph Clustering)

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



	<ul> <li>Decentation rectivent Analysis</li> <li>Paradigm of Clustering</li> <li>Example Applications</li> </ul>	
	<ul><li>Formalization of Aims and Objectives</li><li>Objective Functions</li></ul>	
	Algorithmic Approaches <ul> <li>Greedy Merge</li> <li>Local Moving and Multilevel</li> <li>Clustering with Minimum-Cut Tree</li> <li>Integer Linear Programs</li> <li>Other Algorithmic Approaches</li> </ul>	
4	Experimental Evaluation <ul> <li>The Role of Test Data in Algorithm Engineering</li> <li>Comparing Clusterings</li> </ul>	
	Dynamic Graph Clustering <ul> <li>Online Dynamic Clustering</li> <li>Offline Dynamic Clustering</li> </ul> Appendix	
troduct	ion Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	Appendix

Dorothea Wagner - Hints (Algorithm Engineering for Graph Clustering)

Scenario: Network Analysis

Erice, Italy, 25. September - 3. October, 2011 22/28

	Paradigm of Clustering	
	<ul> <li>Example Applications</li> </ul>	
	Formalization of Aims and Objectives	
	Objective Functions	
	Algorithmic Approaches	
	Greedy Merge	
	<ul> <li>Local Moving and Multilevel</li> </ul>	
	Clustering with Minimum-Cut Tree	
	Integer Linear Programs	
	<ul> <li>Other Algorithmic Approaches</li> </ul>	
4	Experimental Evaluation	
	The Role of Test Data in Algorithm Engineering	
	Comparing Clusterings	
	Dynamic Graph Clustering	< • • •
	Online Dynamic Clustering	< 🗗 >
	Offline Dynamic Clustering	< ≞ > < ≣ >
	Appendix	1
		940
roduc	tion Formalization of Aims and Objectives Algorithmic Approaches Experimental Evaluation Dynamic Graph Clustering	Appendix



Dorothea Wagner - Hints (Algorithm Engineering for Graph Clustering)

• Scenario: Network Analysis

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

In



Dorothea Wagner - Hints (Algorithm Engineering for Graph Clustering)

Scenario: Network Analysis

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

In

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



Dorothea Wagner - Hints (Algorithm Engineering for Graph Clustering)

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

In

troduct	ion Formalization of Aims and Objectives Algorithmic Approaches Exper	rimental Evaluation	Dynamic Graph Clustering	Appendix
	Appendix			ヨークへつ
	Offline Dynamic Clustering			<ul> <li>III →</li> <li>III</li> </ul>
				< ≥ >
	<ul> <li>Online Dynamic Clustering</li> </ul>			< 67 ×
5	Dynamic Graph Clustering			< • • •
_	Comparing Clusterings			
	The Role of Test Data in Algorithm Enginee	ring		
	Experimental Evaluation			
	<ul> <li>Other Algorithmic Approaches</li> </ul>			
	Integer Linear Programs			
	Clustering with Minimum-Cut Tree			
	Local Moving and Multilevel			
	Greedy Merge			
	Algorithmic Approaches			
	-			
	<ul> <li>Objective Functions</li> </ul>			
	Formalization of Aims and Objectives			
	Example Applications			
	Paradigm of Clustering			
	Scenario: Network Analysis			



Dorothea Wagner - Hints (Algorithm Engineering for Graph Clustering)

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

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



Dorothea Wagner - Hints (Algorithm Engineering for Graph Clustering)

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

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



Dorothea Wagner - Hints (Algorithm Engineering for Graph Clustering)

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