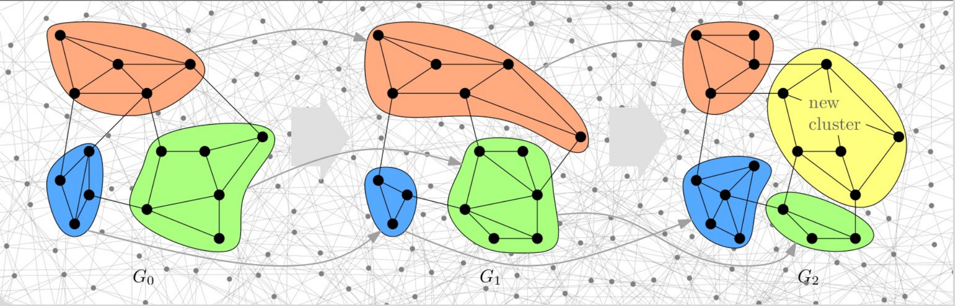
Karlsruher Institut für Technologie

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



Structure

Structure

Structure

Structure

Requirements for Clusterings

Karlsruher Institut für Technologie

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

Structure

Structure

Karlsruher Institut für Technologie

Structure

1. Introduction
   - 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
   - Online Dynamic Clustering
   - Offline Dynamic Clustering
6. Appendix

Properties of Objective Functions

1. Consider the desiderata and the objective functions mentioned in the lecture. Try to discover violations.
2. 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.
3. 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.
4. Strenuous and technical.
5. Setting the density constraint to 1, the decision version is equivalent to the question: *(min-cut clique partition)* Is there a partition $\mathcal{C}$ 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 $\mathcal{S}$ of 3-element subsets of X. Does $\mathcal{S}$ contain an exact cover for X, i.e., a subcollection $\mathcal{S}' \subseteq \mathcal{S}$ such that every $x \in X$ occurs in exactly one $S \in \mathcal{S}'$?

Structure

Structure

Greedy Merge Modularity

1. 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_j are merged, show that
   $\Delta\text{mod}_{(ij),k} = \Delta\text{mod}_{i,k} + \Delta\text{mod}_{j,k}$
   Based on these observations and on how the greedy algorithm works, complete the argument.
2. -

Greedy Merge General

Karlsruher Institut für Technologie

1. 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.
2. Solutions can be found in [Schumm et al.: Density-constrained Graph clustering, 2011]

Structure

Data Structures

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

Structure

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

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

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

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

$$\begin{aligned} c_\alpha(\{t\}, V) &= \alpha\,|V \setminus C| + \alpha\,|C| \\ &< c(C, V \setminus C) + \alpha\,|C| \\ &= c_\alpha(C, \bar{C}). \end{aligned}$$

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

Structure

Variables and Constraints

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]

Structure

Structure

Structure

Structure

Structure

Structure

Structure

Structure