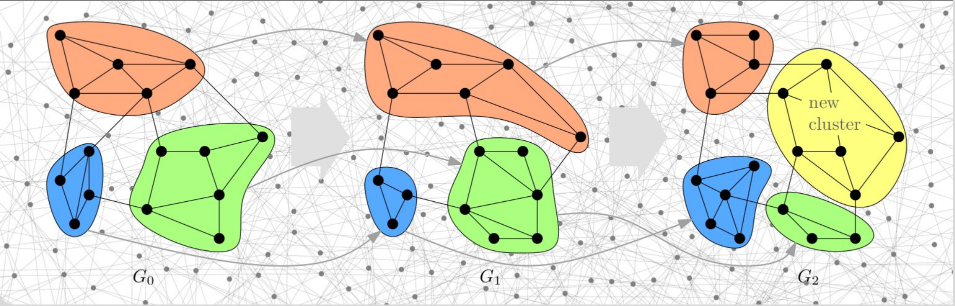
Exercises
AE for Graph Clustering

School on Graph Theory, Algorithms and Applications

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

KARLSRUHE INSTITUTE OF TECHNOLOGY – INSTITUTE OF THEORETICAL INFORMATICS



Structure

Structure

Structure

Structure

Requirements for Clusterings

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

Structure

Structure

Structure

Properties of Objective Functions

Karlsruher Institut für Technologie

1. Consider the desiderata and the objective functions mentioned in the lecture. Try to discover violations.
2. How large is the best cluster editing set you can find for Zachary's Karate Club? At most 54 should be doable . . .
3. For some family of sparse graphs, prove that the modularity of the modularity-optimal clustering approaches 1.
4. Prove that plugging performance instead of coverage into the concept of modularity yields an equivalent measure.
5. 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).

Structure

Structure

Greedy Merge Modularity

1. 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.
2. 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

1. Prove the following reachability result:
   Given density α, clustering $\mathcal{C}$ and a subset $\mathcal{D} \subseteq \mathcal{C}$ of dense clusters:
   $\forall D \in \mathcal{D}$: density$(D) \geq \alpha$.
   If $\mathcal{D}$'s union $U = \bigcup_{D \in \mathcal{D}} D$ is dense, i.e., density$(U) \geq \alpha$, then there exist $A, B \in \mathcal{D}$ such that density$(A \cup B) \geq \alpha$.
2. Prove that maximum isolated (and pairwise) inter-cluster density have unbounded merge behavior (actually you can try to prove the whole list!).

Structure

Data Structures

Karlsruher Institut für Technologie

1. 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.

Structure

Bounds Guaranteed by Min-Cuts

Karlsruher Institut für Technologie

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 $(X, \bar{X})$ 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 $\mathcal{C}$ denote the clustering returned by the Cut-Clustering Algorithm with respect to α and let $C \in \mathcal{C}$ 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 $\mathcal{C}$ denote the clustering returned by the Cut-Clustering Algorithm with respect to α and let $C \in \mathcal{C}$ 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.

Structure

Variables and Constraints

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

1. cluster-leaders: a binary variable stating whether a vertex is the unique leader of its cluster
2. $|\mathcal{C}| \leq$ given value
3. containment: binary variables stating whether vertex v is in cluster C_i
4. $\forall C \in \mathcal{C} : |C| \leq$ given value
5. cluster-size variables: binary variable stating whether $|C_i| = \ell$
6. $\forall C \in \mathcal{C} :$ density$(C) \geq$ given value

Structure

Structure

Structure

Structure

Structure

Structure

Structure

Structure