Axioms for graph clustering objective functions

Twan van Laarhoven

Institute for Computing and Information Sciences
Radboud University Nijmegen, The Netherlands

28th June 2013

Outline

The motivation

- There is no strict definition of clustering.
- Can we formalize our intuition?
- Previous work is about distance based clustering (hierarchical clustering, K-means, etc.)
- What about graphs?

The setting

Definition (Graph)

A symmetric weighted *graph* is a pair (V, E) of

- a finite set V of *nodes*, and
- a function $E : V \times V \to \mathbb{R}_{\geq 0}$ of *edge weights*,

such that $E(i, j) = E(j, i)$ for all $i, j \in V$.

- Larger weight = stronger connection.
- We allow self loops.

The setting (cont.)

Definition (Clustering)

A *clustering* C of a graph $G = (V, E)$ is a partition of its nodes.

Definition (Clustering function)

A *graph clustering function* f is a function from graphs G to clusterings of G.

Definition (Objective function)

A *graph clustering objective function* Q is a function from graphs G and clusterings of G to $\mathbb{R}$.

- Larger objective value = better.

Outline

The form of axioms

Things that define clusterings

	Form	Notation
1	Clustering function	$f(G) = \text{argmax}_C\, Q(G, C)$
2	Objective function	$Q(G, C)$
3	Objective relation	$Q(G, C) \geq Q(G, D)$ or $C \geq_G D$

Basic axioms

Axiom 1: Scale invariance (first form)

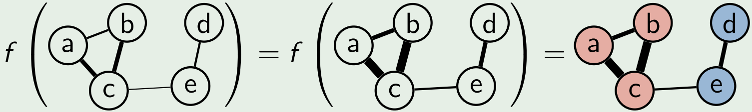
A graph clustering objective function Q is *scale invariant* if

- for all graphs $G = (V, E)$,
- all constants $\alpha > 0$,

$f(G) = f(\alpha G)$.

(where $\alpha G = (V, (i, j) \mapsto \alpha E(i, j))$.)

Example



Basic axioms

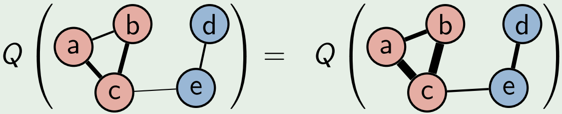
Axiom 1: Scale invariance (second form)

A graph clustering objective function Q is *scale invariant* if

- for all graphs $G = (V, E)$,
- all constants $\alpha > 0$,
- all clusterings C of G,

$Q(G, C) = Q(\alpha G, C)$.
(where $\alpha G = (V, (i, j) \mapsto \alpha E(i, j))$.)

Example



Basic axioms

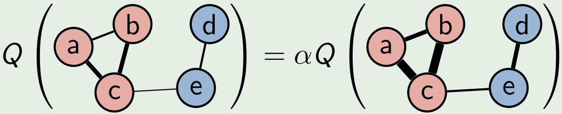
Axiom 1: Scale invariance (second form)

A graph clustering objective function Q is *scale invariant* if

- for all graphs $G = (V, E)$,
- all constants $\alpha > 0$,
- all clusterings C of G,

$Q(G, C) = \alpha Q(\alpha G, C)$???
(where $\alpha G = (V, (i,j) \mapsto \alpha E(i,j))$.)

Example



Basic axioms

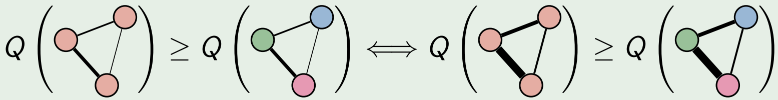
Axiom 1: Scale invariance (third form)

A graph clustering objective function Q is *scale invariant* if

- for all graphs $G = (V, E)$,
- all constants $\alpha > 0$,
- all clusterings C_1, C_2 of G,

$Q(G, C_1) \geq Q(G, C_2)$ if and only if $Q(\alpha G, C_1) \geq Q(\alpha G, C_2)$.
(where $\alpha G = (V, (i, j) \mapsto \alpha E(i, j))$.)

Example



Basic axioms

Axiom 2: permutation invariance

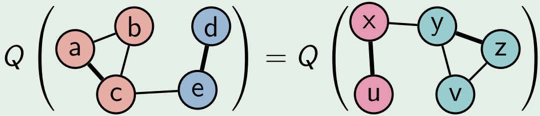
A graph clustering objective function Q is *permutation invariant* if

- for all graphs $G = (V, E)$ and
- all isomorphisms $f : V \to V'$,

it is the case that $Q(G, C) = Q(f(G), f(C))$.

(where f is extended to graphs and clusterings in the obvious way.)

Example



Basic axioms

Axiom 3: Richness

A graph clustering objective function Q is *rich* if

- for all sets V and
- all partitions C^* of V,

there is

- a graph $G = (V, E)$
- such that C^* is the optimal clustering of G.

Intuition:

- No trivial objective functions.
- No fixed number of clusters.

Basic axioms

Definition (Consistent improvement)

Let

- $G = (V, E)$ and $G' = (V, E')$ be graphs, and
- C be a clustering of G and G'.

Then G' *is a C-consistent improvement of G* if

- $E'(i,j) \geq E(i,j)$ for all $i \sim_C j$ and
- $E'(i,j) \leq E(i,j)$ for all $i \not\sim_C j$.

Intuition:

- Consistent improvements make a clustering fit better.

Basic axioms

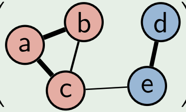
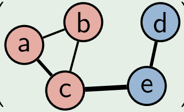
Axiom 4: Monotonicity

A graph clustering objective function Q is *monotonic* if

- for all graphs G,
- all clusterings C of G and
- all C-consistent improvements G' of G

it is the case that $Q(G', C) \geq Q(G, C)$.

Example

Q () $\geq Q$ ()

Local changes

Definition (agreement)

Let

- $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two graphs and
- $V_a \subseteq V_1 \cap V_2$.

The graphs *agree on* V_a if $E_1(i,j) = E_2(i,j)$ for all $i, j \in V_a$.

Definition (agreement on neighborhood)

The graphs also *agree on the neighborhood of* V_a if

$$\begin{aligned} E_1(i,j) &= E_2(i,j) && \text{for all } i \in V_a, j \in V_1 \cap V_2, \text{ and} \\ E_1(i,j) &= 0 && \text{for all } i \in V_a, j \in V_1 \setminus V_2, \text{ and} \\ E_2(i,j) &= 0 && \text{for all } i \in V_a, j \in V_2 \setminus V_1. \end{aligned}$$

What this means:

- For nodes/clusters in V_a, all incident edges are the same.

Local changes

Definition (agreement)

Let

- $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two graphs and
- $V_a \subseteq V_1 \cap V_2$.

The graphs *agree on* V_a if $E_1(i,j) = E_2(i,j)$ for all $i,j \in V_a$.

Definition (agreement on neighborhood)

The graphs also *agree on the neighborhood of* V_a if

$$\begin{aligned} E_1(i,j) &= E_2(i,j) && \text{for all } i \in V_a,\ j \in V_1 \cap V_2, \text{ and} \\ E_1(i,j) &= 0 && \text{for all } i \in V_a,\ j \in V_1 \setminus V_2, \text{ and} \\ E_2(i,j) &= 0 && \text{for all } i \in V_a,\ j \in V_2 \setminus V_1. \end{aligned}$$

What this means:

- For nodes/clusters in V_a, all incident edges are the same.

Local changes

Axiom 5: Locality

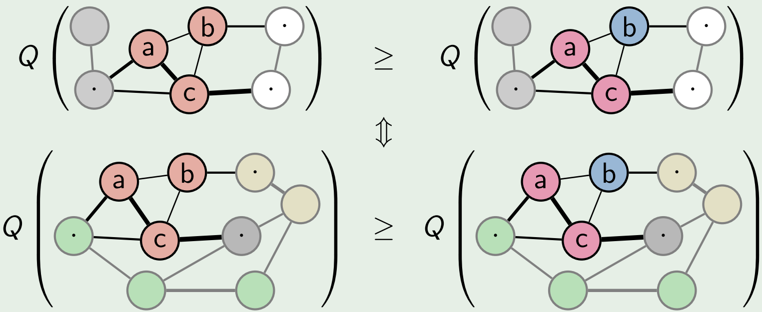
A graph clustering objective function Q is *local* if

- for all graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ that agree on a set V_a and its neighborhood,
- for all clusterings C_1 of $V_1 \setminus V_a$, C_2 of $V_2 \setminus V_a$ and C_a, D_a of V_a.

$$\begin{array}{ll} \text{if} & Q(G_1, C_a \cup C_1) \geq Q(G_1, D_a \cup C_1) \\ \text{then} & Q(G_2, C_a \cup C_2) \geq Q(G_2, D_a \cup C_2). \end{array}$$

Local changes

Example



Local changes

Special cases

- $G_1 = G_2$: change part of a clustering.
  In practice: optimize parts separately (divide and conquer).
- $V_a = \emptyset$: union of two disjoint graphs.

Interlude: Related work

Theorem (Kleinberg 2002)

There is no clustering function that is permutation invariant, scale invariant, monotonic and rich.

Theorem (Ackerman, Ben-David 2008)

There is a clustering quality function that is permutation invariant, scale invariant, monotonic and rich.

Discontinuity is magic

Theorem

There is a graph clustering function that is scale invariant, permutation invariant, monotonic, rich and local.

Connected components

$$f_{\mathrm{coco}}(G) = \text{the connected components of } G$$

$$Q_{\mathrm{coco}}(G, C) = \mathbf{1}[C \text{ are the connected components of } G]$$

Huh!?!?

- Doesn't this contradict Kleinberg's theorem?
- No: edge weight 0 = distance ∞.

Discontinuity is magic

Theorem

There is a graph clustering function that is scale invariant, permutation invariant, monotonic, rich and local.

Connected components

$$f_{\text{coco}}(G) = \text{the connected components of } G$$

$$Q_{\text{coco}}(G, C) = \mathbf{1}[C \text{ are the connected components of } G]$$

Huh!?!?

- Doesn't this contradict Kleinberg's theorem?
- No: edge weight 0 = distance ∞.

Discontinuity is magic

Theorem

There is a graph clustering function that is scale invariant, permutation invariant, monotonic, rich and local.

Connected components

$$f_{\text{coco}}(G) = \text{the connected components of } G$$

$$Q_{\text{coco}}(G, C) = \mathbf{1}[C \text{ are the connected components of } G]$$

Huh!?!?

- Doesn't this contradict Kleinberg's theorem?
- No: edge weight 0 = distance ∞.

Discontinuity is magic

Why I don't like it

- Adding/removing an edge with tiny weight ϵ changes the graph slightly, but the clustering completely.
- Possibly unstable.
- So don't allow it.

Axiom 6: continuity

An objective function Q is *continuous* if a small change in the graph leads to a small change in the objective value.

Outline

An objective function

Modularity

$$Q_{\text{modularity}}(G, C) = \sum_{c \in C} \left(\frac{w_c}{v_V} - \left(\frac{v_c}{v_V} \right)^2 \right).$$

Where

$$v_c = \sum_{i \in c} \sum_{j \in V} E(i, j) \quad \text{volume of cluster}$$

$$w_c = \sum_{i \in c} \sum_{j \in c} E(i, j) \quad \text{within cluster weight.}$$

Properties

The obvious:

- Modularity is permutation invariant.
- Modularity is scale invariant.
- Modularity is continuous.

The less obvious:

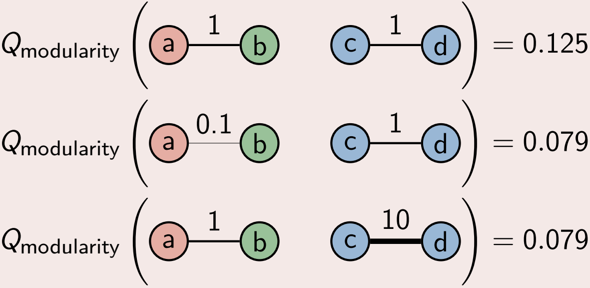
- Modularity is rich.

The bad:

- Modularity is *not* local.
- Modularity is *not* monotonic.

What goes wrong?

Modularity is not monotonic.



□

Outline

Fixed Scale modularity

Idea 1

Fix the scale

$$Q_{M\text{-fixed}}(G, C) = \sum_{c \in C} \left(\frac{w_c}{M} - \left(\frac{v_c}{M} \right)^2 \right)$$

Is it monotonic?

Take $v_c = w_c + b_c$ (within + between)

$$\frac{\partial Q_{M\text{-fixed}}(G, C)}{\partial w_c} = \frac{1}{M} - \frac{2w_c + 2b_c}{M^2}.$$

This is negative when $2v_c > M$

$\Rightarrow$ not monotonic

Fixed Scale modularity

Idea 1

Fix the scale

$$Q_{M\text{-fixed}}(G, C) = \sum_{c \in C} \left(\frac{w_c}{M} - \left(\frac{v_c}{M} \right)^2 \right)$$

Is it monotonic?

Take $v_c = w_c + b_c$ (within + between)

$$\frac{\partial Q_{M\text{-fixed}}(G, C)}{\partial w_c} = \frac{1}{M} - \frac{2w_c + 2b_c}{M^2}.$$

This is negative when $2v_c > M$

$\Rightarrow$ not monotonic

Fixed Scale modularity

Idea 1

Fix the scale

$$Q_{M\text{-fixed}}(G, C) = \sum_{c \in C} \left(\frac{w_c}{M} - \left(\frac{v_c}{M} \right)^2 \right)$$

Is it monotonic?

Take $v_c = w_c + b_c$ (within + between)

$$\frac{\partial Q_{M\text{-fixed}}(G, C)}{\partial w_c} = \frac{1}{M} - \frac{2w_c + 2b_c}{M^2}.$$

This is negative when $2v_c > M$

$\Rightarrow$ **not monotonic**

Adaptive Scale Modularity

Idea 2

Add some v_c to the denominator

$$Q_{M,\gamma}(G,C) = \sum_{c \in C} \left(\frac{w_c}{M + \gamma v_c} - \left(\frac{v_c}{M + \gamma v_c} \right)^2 \right).$$

Theorem

Adaptive scale modularity is monotonic for all $M \geq 0$ and $\gamma \geq 2$.

Theorem

Adaptive scale modularity is rich for all $M \geq 0$ and $\gamma \geq 1$.

Theorem

Adaptive scale modularity is scale invariant for $M = 0$.

Adaptive Scale Modularity

Idea 2

Add some v_c to the denominator

$$Q_{M,\gamma}(G,C) = \sum_{c \in C} \left(\frac{w_c}{M + \gamma v_c} - \left(\frac{v_c}{M + \gamma v_c} \right)^2 \right).$$

Theorem

Adaptive scale modularity is monotonic for all $M \geq 0$ and $\gamma \geq 2$.

Theorem

Adaptive scale modularity is rich for all $M \geq 0$ and $\gamma \geq 1$.

Theorem

Adaptive scale modularity is scale invariant for $M = 0$.

Adaptive Scale Modularity

Idea 2

Add some v_c to the denominator

$$Q_{M,\gamma}(G,C) = \sum_{c \in C} \left(\frac{w_c}{M + \gamma v_c} - \left(\frac{v_c}{M + \gamma v_c} \right)^2 \right).$$

Theorem

Adaptive scale modularity is monotonic for all $M \geq 0$ and $\gamma \geq 2$.

Theorem

Adaptive scale modularity is rich for all $M \geq 0$ and $\gamma \geq 1$.

Theorem

Adaptive scale modularity is scale invariant for $M = 0$.

Adaptive Scale Modularity

Idea 2

Add some v_c to the denominator

$$Q_{M,\gamma}(G,C) = \sum_{c\in C}\left(\frac{w_c}{M+\gamma v_c} - \left(\frac{v_c}{M+\gamma v_c}\right)^2\right).$$

Theorem

Adaptive scale modularity is monotonic for all $M \geq 0$ and $\gamma \geq 2$.

Theorem

Adaptive scale modularity is rich for all $M \geq 0$ and $\gamma \geq 1$.

Theorem

Adaptive scale modularity is scale invariant for $M = 0$.

Adaptive Scale Modularity: related objectives

- When $\gamma = 0$, we get fixed scale modularity. Equivalent to other modularity variants.
- When $\gamma = 0$ and $M = v_V$, we get modularity.
- When $M = 0$ we get

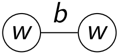
$$Q_{0,\gamma}(G, C) \propto \sum_{c \in C} \left(\frac{w_c}{v_c} - \frac{1}{\gamma} \right),$$

  i.e. normalized cut.
- When $M \to \infty$ we get

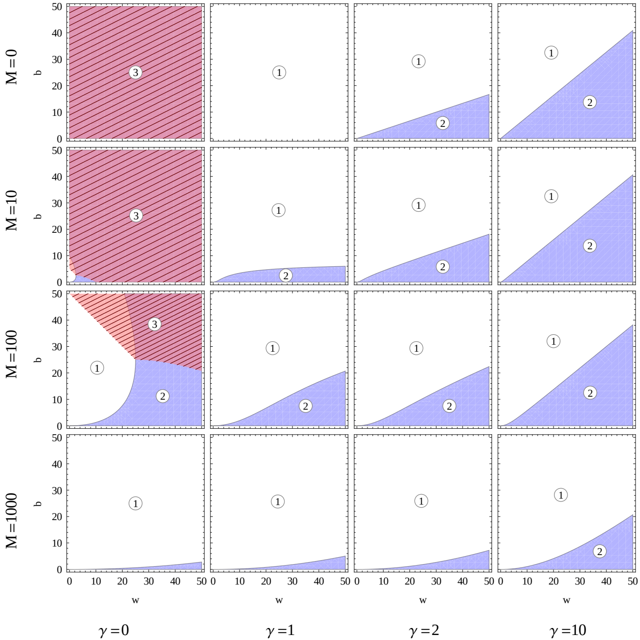
$$Q_{\infty,\gamma}(G, C) \propto \sum_{c \in C} w_c,$$

  i.e. unnormalized cut.

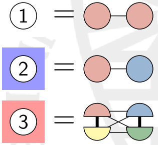
Adaptive Scale Modularity: behavior

Take a simple graph: 

- Two cliques each with w within weight
- Connected by edges with total weight b.
- Total volume $2w + 2b$.
- What is the behavior of adaptive scale modularity?



Legend:



Outline

Summary

- 6 axioms for graph clustering objectives.
- Graph setting allows for locality.
- Modularity is not monotonic.
- Non-monotonicity leads to splitting of cliques.
- Adaptive scale modularity satisfies all axioms (when $M = 0$).
- Generalizes both modularity and normalized cut.

Thank you for your attention.

Axioms for graph clustering objective functions

Twan van Laarhoven

Institute for Computing and Information Sciences
Radboud University Nijmegen, The Netherlands

28th June 2013