Axioms for graph clustering

Twan van Laarhoven

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

9th September 2013

Outline

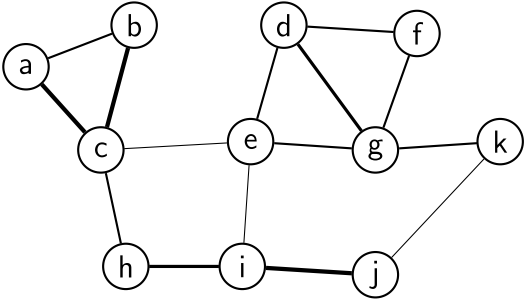
Introduction

Axioms

Modularity

Conclusion

Graphs

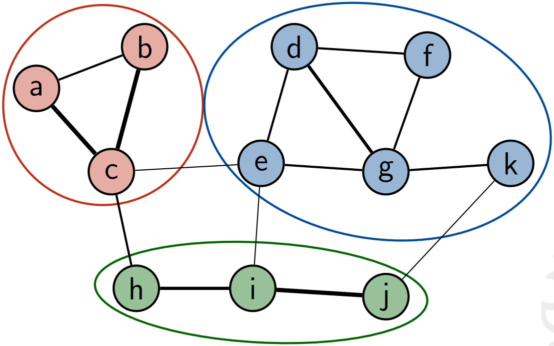


A symmetric weighted **graph** (or network) is a pair (V, E) of

- 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 $i, j \in V$.

Graph clustering



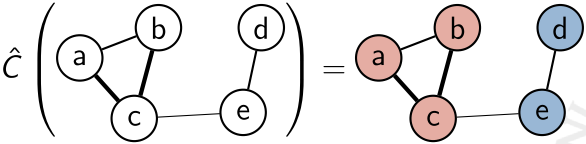
A **clustering** C of a graph $G = (V, E)$ is a partition of its nodes.

Applications

- Social networks
- Hyperlinks
- Protein interaction networks
- References between mathematical theorems
- Brain parcellation

Clustering methods

1. Clustering function
   $\hat{C}$: Graph $\to$ Clustering



2. Quality function
   Q : Graph $\times$ Clustering $\to \mathbb{R}$
3. Quality relation
   $\cdot \preceq^G \cdot \subseteq$ Clustering $\times$ Clustering

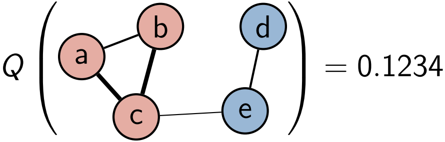
Clustering methods

1. Clustering function

   $\hat{C}$: Graph $\to$ Clustering

2. Quality function

   Q : Graph $\times$ Clustering $\to \mathbb{R}$

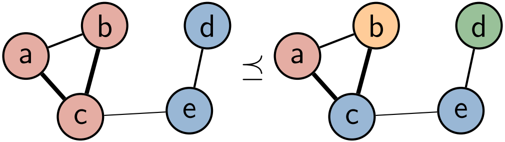


3. Quality relation

   $\cdot \preceq^{G} \cdot \subseteq$ Clustering $\times$ Clustering

Clustering methods

1. Clustering function
   $\hat{C}$: Graph $\to$ Clustering
2. Quality function
   Q : Graph $\times$ Clustering $\to \mathbb{R}$
3. Quality relation
   $\cdot \preceq^G \cdot \subseteq$ Clustering $\times$ Clustering



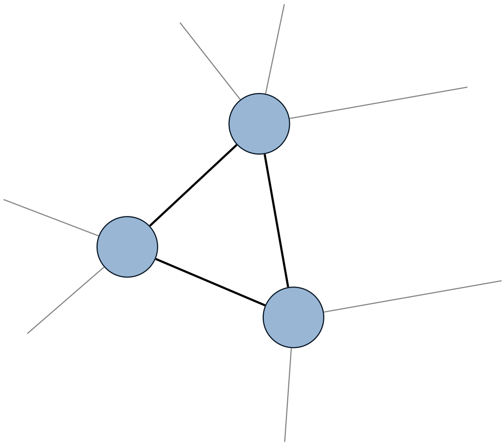
Clustering by optimization

- Graph clustering is NP hard.
- Top down:
  find best cut and repeat
- Bottom up:
  group nodes together
- Simulated annealing

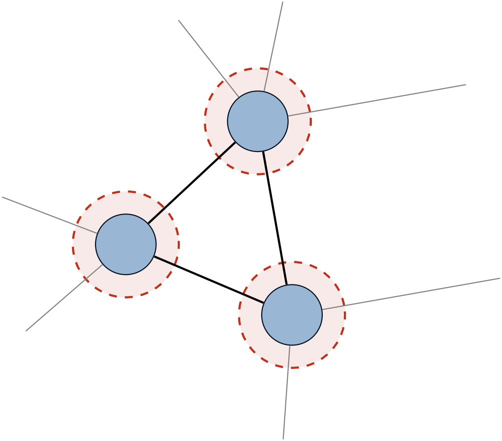
Louvain method

- V.D. Blondel, JL. Guillaume, R. Lambiotte, E. Lefebvre
  Fast unfolding of communities in large networks
  J. Stat. Mech. 2008
- Best graph clustering method in surveys.
- Method:
  1. Move nodes into neighboring clusters to improve quality.
  2. Repeat until local maximum.
  3. Now cluster the clusters.

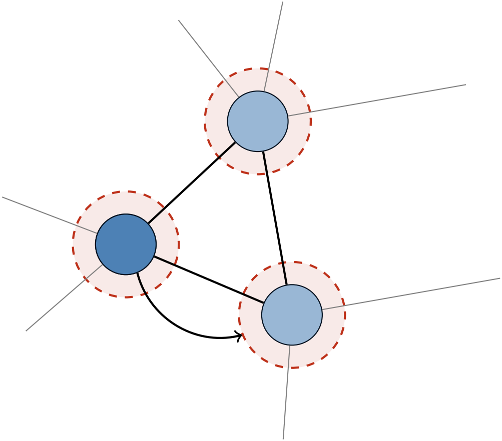
Louvain method (example)



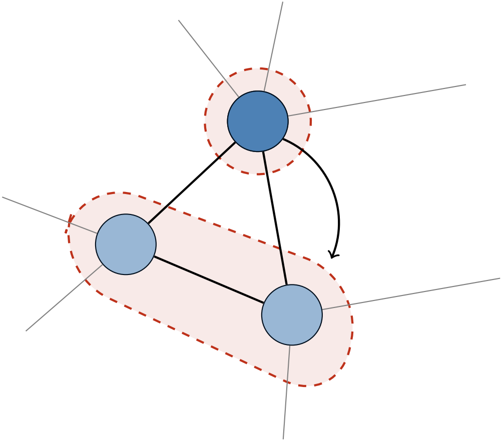
Louvain method (example)



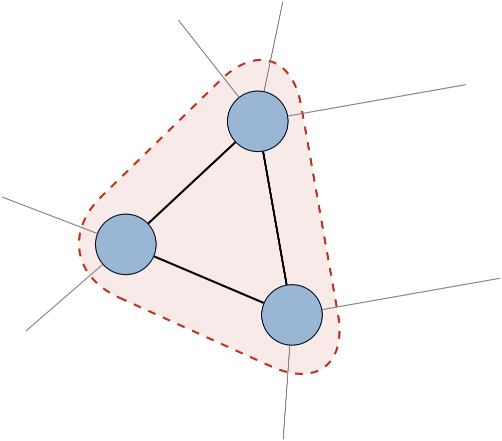
Louvain method (example)



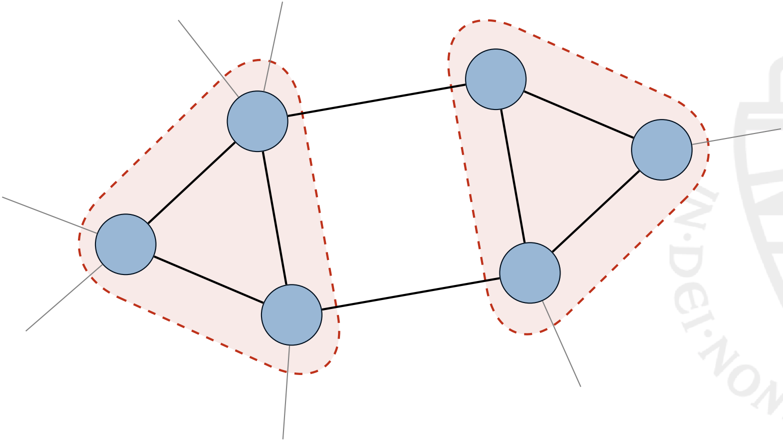
Louvain method (example)



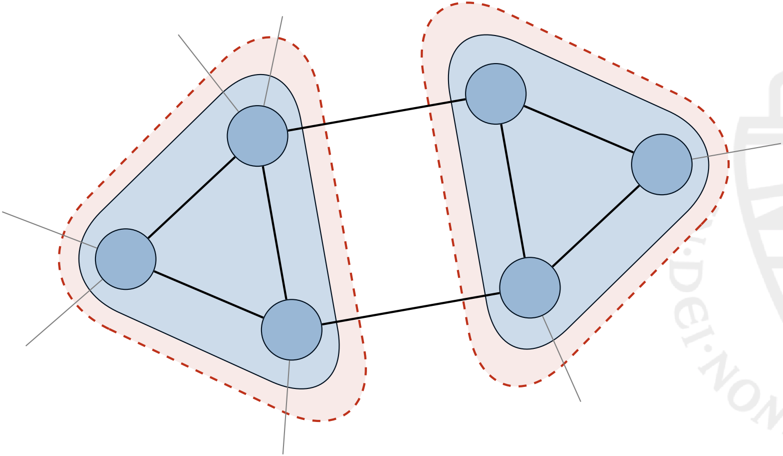
Louvain method (example)



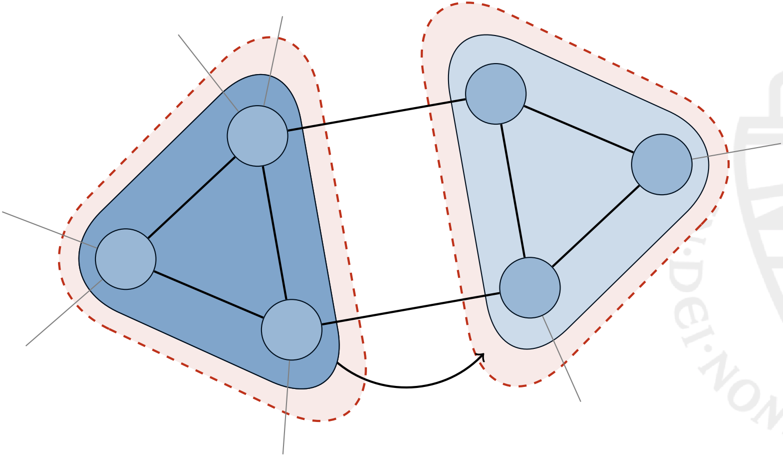
Louvain method (example)



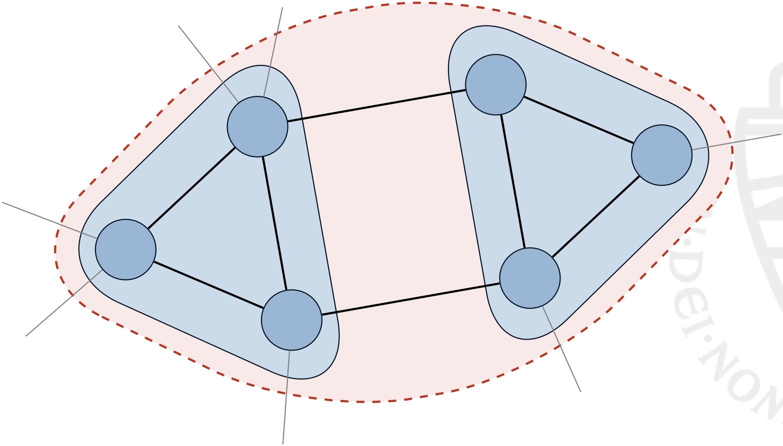
Louvain method (example)



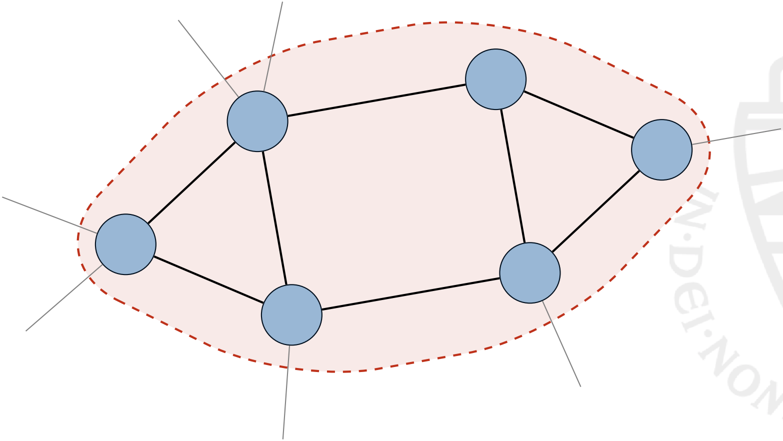
Louvain method (example)



Louvain method (example)



Louvain method (example)



Some quality functions

- Connected components
- Total weight of within cluster edges
$$Q(G, C) = \sum_{c \in C} w_c$$
- Modularity
$$Q(G, C) = \sum_{c \in C} \left(w_c / v_V - (v_c / v_V)^2 \right)$$
- Many more
$$Q(G, C) = \sum_{c \in C} -w_c \log(v_c / v_V)$$
$$\ldots$$

Families of quality functions

- Connected components with threshold
- Total weight of within cluster edges with penalty
  $$Q(G, C) = \sum_{c \in C} w_c - \alpha |C|$$
- Modularity
  $$Q_{\text{RB}}^{\gamma}(G, C) = \sum_{c \in C} \left(w_c / v_V - \gamma (v_c / v_V)^2 \right)$$
- Many more
  $$Q(G, C) = \sum_{c \in C} -w_c \log(v_c / \alpha)$$
  $\ldots$

Outline

Introduction

Axioms

Modularity

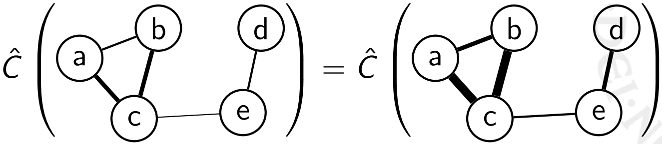
Conclusion

Why axioms?

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

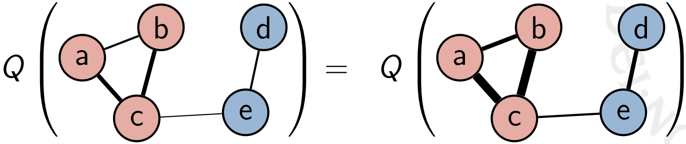
Axiom 1: Scale invariance

Intuition: The absolute value of the edge weights shouldn't matter.



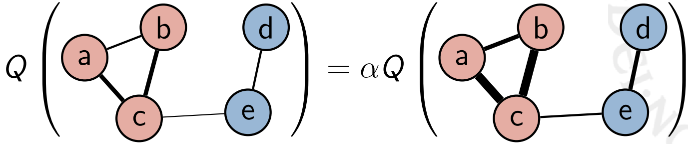
Axiom 1: Scale invariance

Intuition: The absolute value of the edge weights shouldn't matter.



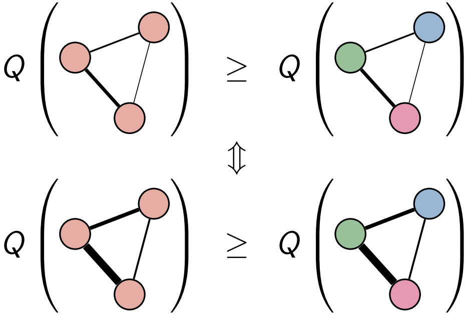
Axiom 1: Scale invariance

Intuition: The absolute value of the edge weights shouldn't matter.



Axiom 1: Scale invariance

Intuition: The absolute value of the edge weights shouldn't matter.



Axiom 1: Scale invariance

Intuition: The absolute value of the edge weights shouldn't matter.

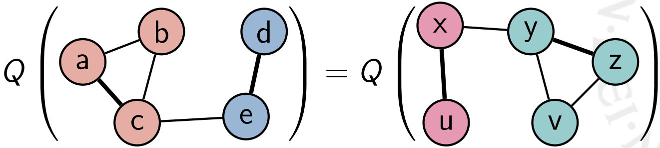
A quality function Q is **scale invariant** if

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

$Q(G, C_1) \geq Q(G, C_2)$ if and only if $Q(\alpha G, C_1) \geq Q(\alpha G, C_2)$.

Axiom 2: Permutation invariance

Intuition: Only the edge weights should matter.



Axiom 2: Permutation invariance

Intuition: Only the edge weights should matter.

A quality function Q is **permutation invariant** if

$$Q(G, C) = Q(f(G), f(C)).$$

for all

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

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

Axiom 3: Richness

Intuition:

- All clusterings must be possible.

So,

- no trivial quality functions.
- no fixed number of clusters.

A quality function Q is **rich** if

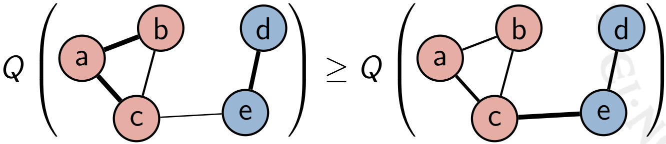
- for all sets V and
- all partitions C^* of V,

there is

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

Axiom 4: Monotonicity

Intuition: Adding edges inside a cluster or removing edges between clusters does not make the clustering worse.



Axiom 4: Monotonicity

Intuition: Adding edges inside a cluster or removing edges between clusters does not make the clustering worse.

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 $i \not\sim_C j$.

Axiom 4: Monotonicity

Intuition: Adding edges inside a cluster or removing edges between clusters does not make the clustering worse.

A quality function Q is **monotonic** if

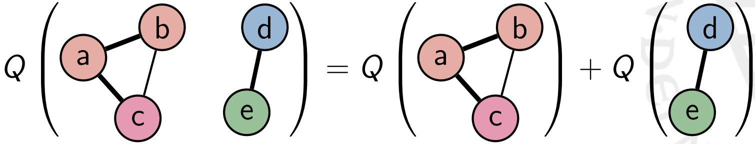
$$Q(G', C) \geq Q(G, C).$$

for all

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

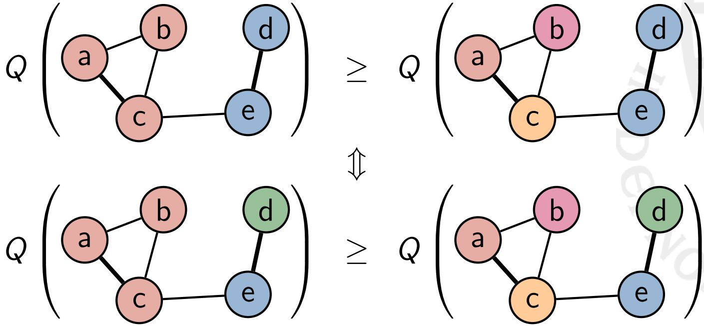
Axiom 5: Locality

Intuition: Local changes should have local effects.



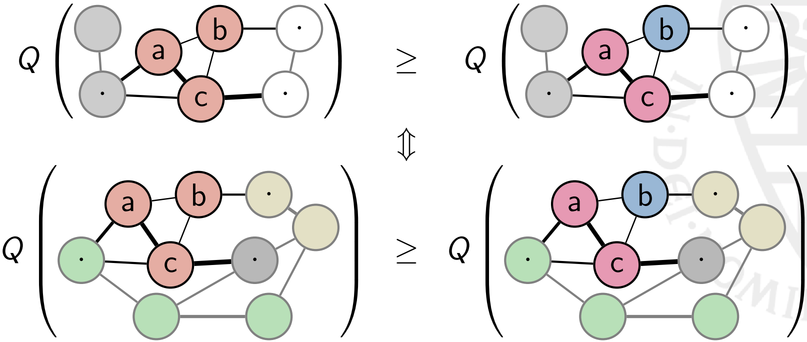
Axiom 5: Locality

Intuition: Local changes should have local effects.



Axiom 5: Locality

Intuition: Local changes should have local effects.



Axiom 5: Locality

Intuition: Local changes should have local effects.

Two graphs G_1 and G_2 **agree on the neighborhood of** $V_a \subseteq V_1 \cap V_2$ if

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

So, for nodes/clusters in V_a, all incident edges are the same.

Axiom 5: Locality

Intuition: Local changes should have local effects.

A quality 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.

if $\quad Q(G_1, C_a \cup C_1) \geq Q(G_1, D_a \cup C_1)$
then $Q(G_2, C_a \cup C_2) \geq Q(G_2, D_a \cup C_2)$.

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.

$$\hat{C}_{\text{coco}}(G) = \text{the connected components of } G$$

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

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

Discontinuity is magic

Theorem

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

$$\hat{C}_{\text{coco}}(G) = \text{the connected components of } G$$

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

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

Axiom 6: continuity

Intuition:

- Don't allow such unstable quality functions.

A quality function Q is **continuous** if for every graph G and every clustering C of G, a sufficiently small change in the edge weights leads to a small change in the objective value.

Outline

Modularity

Intuition:

- Balance within cluster edges against cluster volume.

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

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.

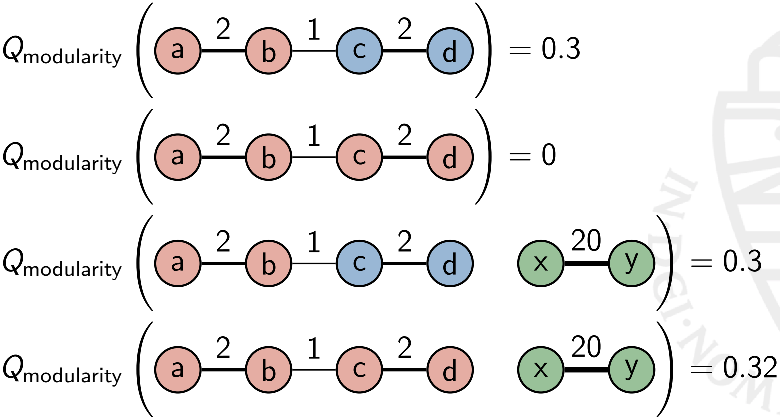
The less obvious:

- Modularity is rich.

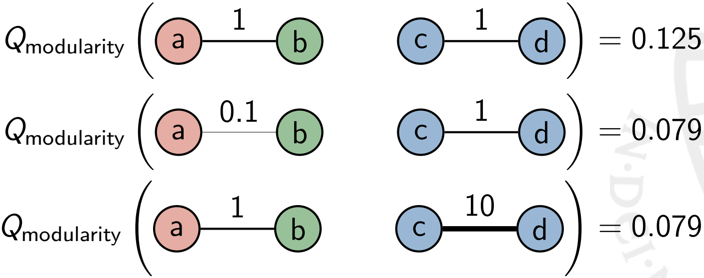
The bad:

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

Modularity is not local



Modularity is not monotonic



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 $2v_c > M$, so not monotonic.

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 $2v_c > M$, so not monotonic.

Idea 2: Add some v_c to the denominator

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

Adaptive scale modularity is

- permutation invariant, continuous and local.
- monotonic for all $M \geq 0$ and $\gamma \geq 2$.
- rich for all $M \geq 0$ and $\gamma \geq 1$.
- scale invariant for $M = 0$.

Idea 2: Add some v_c to the denominator

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

Adaptive scale modularity is

- permutation invariant, continuous and local.
- monotonic for all $M \geq 0$ and $\gamma \geq 2$.
- rich for all $M \geq 0$ and $\gamma \geq 1$.
- scale invariant for $M = 0$.

Related quality functions

- 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

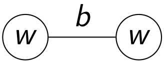
$$Q_{0,\gamma}(G, C) \propto \sum_{c \in C} \left(\frac{w_c}{v_c} - \frac{1}{\gamma} \right),$$

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

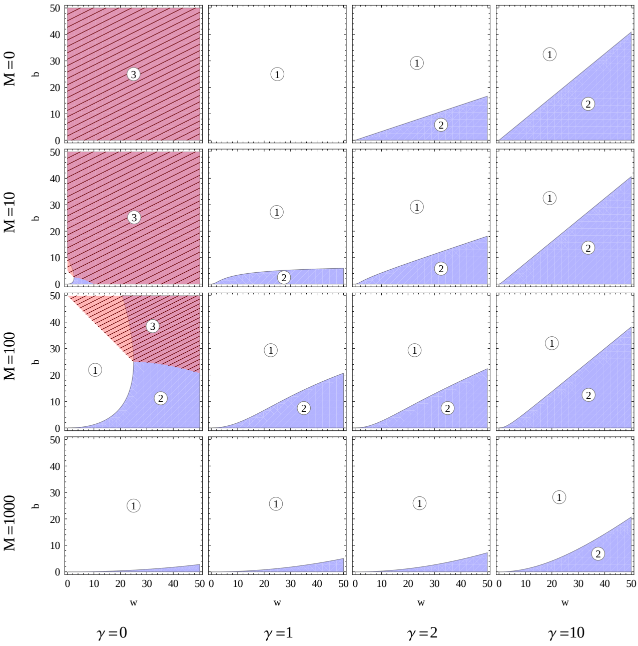
$$Q_{\infty,\gamma}(G, C) \propto \sum_{c \in C} w_c,$$

  i.e. unnormalized cut.

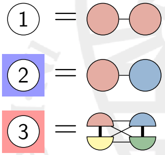
Adaptive Scale Modularity behavior

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

Introduction

Axioms

Modularity

Conclusion

Summary

- Graph clustering by optimization.
- 6 axioms for graph clustering quality functions.
- Graph setting allows for locality.
- Modularity is not monotonic.
- Non-monotonicity leads to splitting of cliques.
- Adaptive scale modularity satisfies all axioms.
- Generalizes both modularity and normalized cut.
- Two parameters to control size of clusters.

Thank you for your attention.

Axioms for graph clustering

Twan van Laarhoven

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

9th September 2013