Solution Stability in Linear Programming Relaxations: Graph Partitioning and Unsupervised Learning

Sebastian Nowozin and Stefanie Jegelka

Department Empirical Inference for Machine Learning and Perception
Max Planck Institute for Biological Cybernetics
Tübingen, Germany

June 15, 2009

MAX-PLANCK-GESELLSCHAFT

BIOLOGISCHE KYBERNETIK

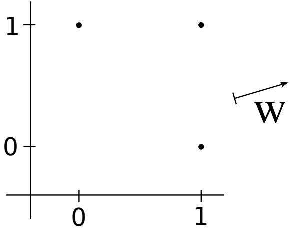
Combinatorial Problems

Many ML problems take the form

$$(P1) \qquad \mathbf{z}^* := \operatorname{argmin}_{\mathbf{z}\in\mathcal{B}} \mathbf{w}^\top \mathbf{z},$$

where

- $\mathcal{B} \subseteq \{0,1\}^n$: finite set of binary indicator vectors of length n.



- Despite simplicity: very general model

Stability Analysis

Solution stability with respect to **w** for a single problem instance.

Why?

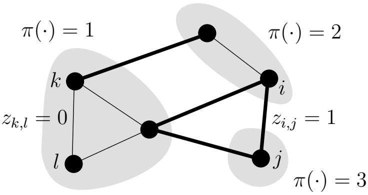
- **w** from noisy measurements: stable solutions $\longrightarrow$ robust to noise
- **w** dependent on model parameters: stable solutions $\longrightarrow$ trust in the model
- insight into data: multiple stable solutions can indicate different regimes
- parametrized solutions: solution paths, regularization paths, etc.

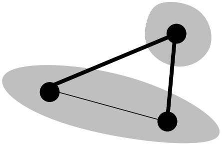
How?

- Linear programming analysis on LP relaxations
- Running example: graph partitioning

Example: Graph Partitioning / Clustering

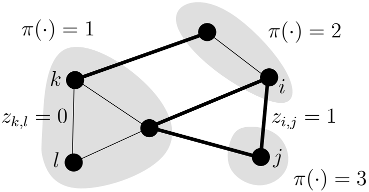
- Graph $G = (V, E)$, undirected, connected, simple
- $\mathbf{z}^* := \operatorname{argmin}_{\mathbf{z} \in \mathcal{B}} \mathbf{w}^\top \mathbf{z}$
- Variables $\mathbf{z} \in \{0, 1\}^E$,
  - $z_{i,j} = 1$ "different partition",
  - $z_{i,j} = 0$ "same partition",
- Solutions $\mathcal{B} \subset \{0, 1\}^E$: all valid graph partitionings (multicut polytope)
- Weights $w \in \mathbb{R}^E$:
  - $w_{i,j} > 0$ "prefer to be in the same partition",
  - $w_{i,j} < 0$ "prefer to be in different partition".

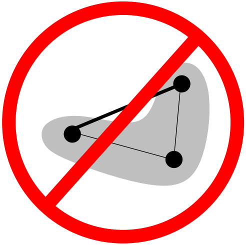




Example: Graph Partitioning / Clustering

- Graph $G = (V, E)$, undirected, connected, simple
- $\mathbf{z}^* := \operatorname{argmin}_{\mathbf{z} \in \mathcal{B}} \mathbf{w}^\top \mathbf{z}$
- Variables $\mathbf{z} \in \{0,1\}^E$,
  - $z_{i,j} = 1$ "different partition",
  - $z_{i,j} = 0$ "same partition",
- Solutions $\mathcal{B} \subset \{0,1\}^E$: all valid graph partitionings (multicut polytope)
- Weights $w \in \mathbb{R}^E$:
  - $w_{i,j} > 0$ "prefer to be in the same partition",
  - $w_{i,j} < 0$ "prefer to be in different partition".





Example (cont')

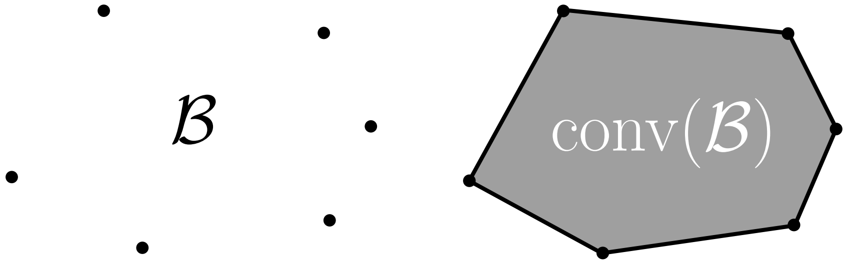
Setting covers popular methods

Method	Weights
Correlation Clustering	$w_{i,j}$ = similarity ratings
Clustering Aggregation	$w(i,j) = \frac{1}{m}\sum_{k=1}^{m}\left(1-2r_{i,j}^{k}\right), \forall (i,j) \in V \times V$ (Expected similarity with proposal clusterings)
Modularity Clustering	$w(i,j) = \frac{1}{2\lvert E\rvert}\left(\eta_{i,j} - \frac{\deg(i)\deg(j)}{2\lvert E\rvert}\right), \forall (i,j) \in V \times V$ Difference between achieved and expected fraction of intra-cluster edges
Relative Performance Significance Clustering	(Generalization of modularity clustering to more general measures of performance)
Bias: diff. of cluster sizes	$\lambda \sum_{k,l=1}^{K}(\lvert C_k\rvert - \lvert C_l\rvert)^2$
Bias: squared cluster sizes	$\lambda \sum_{k=1}^{K}\lvert C_k\rvert^2$

Linear Programming Relaxations

- $\mathcal{B}$ is a finite but large set
- Optimization over the convex hull $\mathrm{conv}(\mathcal{B})$ is exact, i.e.

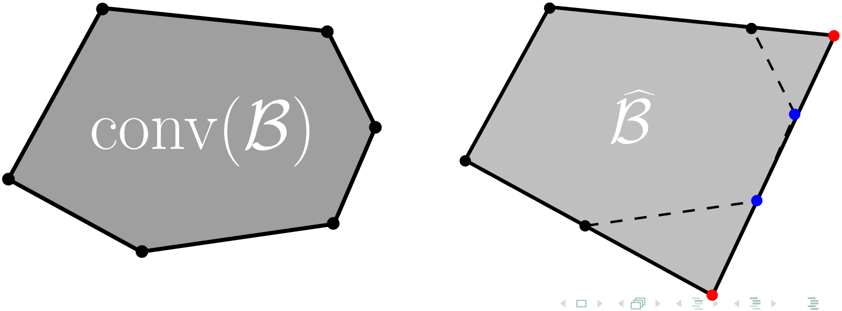
$$\operatorname{argmin}_{\mathbf{z}\in\mathcal{B}} \mathbf{w}^\top \mathbf{z} = \operatorname{argmin}_{\mathbf{z}\in\mathrm{conv}(\mathcal{B})} \mathbf{w}^\top \mathbf{z}.$$



Linear Programming Relaxations (cont)

- $\mathcal{B}$ and thus $\mathrm{conv}(\mathcal{B})$ is hard to describe
- Idea: approximate $\mathrm{conv}(\mathcal{B})$ by a larger set $\widehat{\mathcal{B}}$

$$\mathrm{argmin}_{\mathbf{z}\in\mathrm{conv}(\mathcal{B})}\,\mathbf{w}^\top\mathbf{z} \geq \mathrm{argmin}_{\mathbf{z}\in\widehat{\mathcal{B}}}\,\mathbf{w}^\top\mathbf{z}.$$



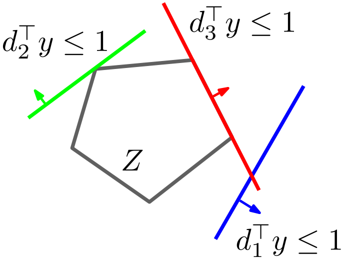
Facets and Valid Inequalities

Convex polytopes have two equivalent representations

- As a convex combination of extreme points
- As a set of facet-defining linear inequalities

A linear inequality with respect to a polytope can be

- *valid*, does not cut off the polytope,
- *representing a face*, valid and touching,
- *facet-defining*, representing a face of dimension one less than the polytope.



Solving Linear Relaxations

Delayed constraint generation

- Optimize over a subset of linear inequalities
- Identify violated inequalities over the full set

Returns solution and lower bound on the optimal objective

$S \leftarrow [0,1]^n$ {Initial feasible set}
loop
  $\mathbf{z} \leftarrow \operatorname{argmin}_{\mathbf{z} \in S} \mathbf{w}^\top \mathbf{z}$ {Solve LP relaxation}
  $S_{\text{violated}} \leftarrow \textsc{SeparateInequalities}(\mathcal{B}, \mathbf{z})$
  if no violated inequality found **then**
    break
  end if
  $S \leftarrow S \cap S_{\text{violated}}$ {Cut $\mathbf{z}$ from feasible set}
end loop
$\text{optimal} \leftarrow (\mathbf{z} \in \{0,1\}^n)$ {Integrality check}
$(f, \mathbf{z}^*) \leftarrow (\mathbf{w}^\top \mathbf{z}, \mathbf{z})$

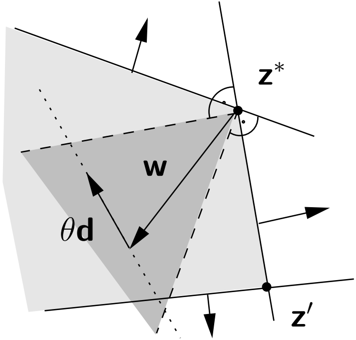
Stability Range

Setting

- *optimal solution* $\mathbf{z}^* := \operatorname{argmin}_{\mathbf{z}\in\operatorname{conv}(\mathcal{B})} \mathbf{w}^\top \mathbf{z}$,
- *perturbation vector* $\mathbf{d} \in \mathbb{R}^n$,
- *modified weights* $\mathbf{w}'(\theta) = \mathbf{w} + \theta\mathbf{d}$

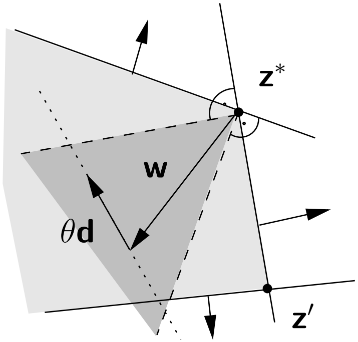
Stability Range

- *stability range*: θ-interval $[\rho_{\mathbf{d},-}, \rho_{\mathbf{d},+}] \in (\{-\infty,\infty\} \cup \mathbb{R})^2$ for which $\mathbf{z}^*$ remains optimal
- *perturbed problem* $\min_{\mathbf{z}\in\operatorname{conv}(\mathcal{B})} \mathbf{w}'(\theta)^\top \mathbf{z}$.



Stability Analysis (1)

- LP geometry: solution becomes suboptimal when $\mathbf{w} + \theta\mathbf{d}$ leaves the cone of negative constraint normals at $\mathbf{z}^*$
- Standard LP stability analysis: basis matrix approach
- Here: does not work, not all binding constraints at $\mathbf{z}^*$ are known, additionally degeneracy



Stability Analysis (2)

Idea from Jansen [7]

- explicitly search cone of constraint normals

$$\begin{aligned} \min_{\alpha \in \mathbb{R}, \mathbf{z} \in \mathbb{R}^n} \quad & \mathbf{w}^\top \mathbf{z} + \alpha \mathbf{w}^\top \mathbf{z}^* \\ \text{sb.t.} \quad & (\frac{1}{\alpha}\mathbf{z}) \in \mathrm{conv}(\mathcal{B}), \\ & (\mathbf{d}^\top \mathbf{z}^*)\alpha - \mathbf{d}^\top \mathbf{z} = t \ : \ \lambda, \\ & 0 \le \mathbf{z}_i \le \alpha, \qquad i = 1, \dots, n. \end{aligned}$$

- $(\frac{1}{\alpha}\mathbf{z}) \in \mathrm{conv}(\mathcal{B})$ still linear $(A(\frac{1}{\alpha}\mathbf{z}) \le b \Leftrightarrow A\mathbf{z} - \alpha b \le 0)$
- Separation routine recycling: given $(\mathbf{z}, \alpha)$ we can still separate from $\mathrm{conv}(\mathcal{B})$
- Complexity: identical to canonical problem

Lagrange multiplier λ provides $\rho_{\mathbf{d},-}$ for the left boundary $(t = -1)$ or $\rho_{\mathbf{d},+}$ for the right boundary $(t = 1)$.

Stability Analysis (3)

Theorem (Stability Inclusion)

Let $\mathbf{z}^*$ be the optimal solution of P1 for a given $\mathcal{B} \subseteq \{0,1\}^n$ and weights $\mathbf{w} \in \mathbb{R}^n$. For a perturbation $\mathbf{d} \in \mathbb{R}^n$, let $[\xi_{\mathbf{d},-}, \xi_{\mathbf{d},+}]$ be the true stability range for θ on $conv(\mathcal{B})$. If $\widehat{\mathcal{B}} \supseteq conv(\mathcal{B})$ is a polyhedral relaxation of $\mathcal{B}$ using only facet-defining inequalities and if $\mathbf{z}^*$ is a vertex of $\widehat{\mathcal{B}}$, then the stability range $[\rho_{\mathbf{d},-}, \rho_{\mathbf{d},+}]$ on $\widehat{\mathcal{B}}$, i.e., for the relaxation $min_{\mathbf{z} \in \widehat{\mathcal{B}}} \mathbf{w}^\top \mathbf{z}$, is included in the true range: $[\rho_{\mathbf{d},-}, \rho_{\mathbf{d},+}] \subseteq [\xi_{\mathbf{d},-}, \xi_{\mathbf{d},+}]$.

Simply put

- estimated stability is conservative
- never overestimates the true stability

Multicut Polytope (1)

- Convex hull of the set of all partitionings of a graph: *multicut polytope*
- Extensive results in late eighties and early nineties [5, 6, 2, 3, 4].
- Classes of facet-defining inequalities known

Polynomial-time separable facet-defining inequalities for the multicut polytope

- Cycle inequalities
- Odd-wheel inequalities

Multicut Polytope: Cycle Inequalities

- generalize triangle inequalities,
- valid graph partitioning **z** satisfies a *transitivity* relation: there is no all-zero path between any two adjacent vertices i, j that are in different subsets of the partition.

For chord-free cycles $((i,j),p)$, $p \in \mathrm{Path}(i,j)$, where $\mathrm{Path}(i,j)$ is the set of paths between i and j, we have the facet-defining inequalities

$$z_{i,j} \leq \sum_{(s,t)\in p} z_{s,t}, \qquad (i,j) \in E, \quad p \in \mathrm{Path}(i,j).$$

- Complete graphs: all cycles longer than three edges contain chords $\longrightarrow$ reduces to triangle inequalities,
- Separation procedure: series of shortest path problems.

Multicut Polytope: Odd-wheel Inequalities

- A *q-wheel* is a connected subgraph $S = (V_s, E_s)$ with a central vertex $j \in V_s$ and a cycle of the q vertices in $C = V_s \setminus \{j\}$,
- For each $i \in C$ there exists an edge $(i, j) \in E_s$.

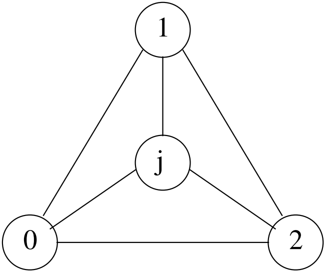


Figure: 3-wheel with center j.

Then, for every q-wheel, a valid partitioning $\mathbf{z}$ satisfies

$$\sum_{(s,t)\in E(C)} z_{s,t} - \sum_{i\in C} z_{i,j} \leq \lfloor \frac{1}{2} q \rfloor,$$

- Polynomial-time separable [3, 2]

Experiment: Relaxation Tightness

Examine tightness of multicut polytope relaxation

- Maximize modularity objective on popular benchmark data sets [1, 8]
- Kernighan-Lin: popular graph-partitioning heuristic (VLSN)
- LP-C: LP relaxation with cycle-inequalities only
- LP-CO: LP relaxation with cycle- and oddwheel-inequalities

	Kernighan-Lin		LP-C		LP-CO	
dolphins	0.5268	0.4s	(0.5315)	4.2s	**0.5285**	9.1s
karate	**0.4198**	0.1s	**0.4198**	0.2s	**0.4198**	0.2s
polbooks	0.5226	7.0s	(0.5276)	147.4s	**0.5272**	148.5s
lesmis	0.5491	1.5s	(0.5609)	6.9s	**0.5600**	11.7s
att180	0.6559	14.5s	(0.6633)	302.3s	**0.6595**	1119.6s

- LP-CO achieves global optimum, LP-C only on smallest problem
- Heuristic fast but suboptimal

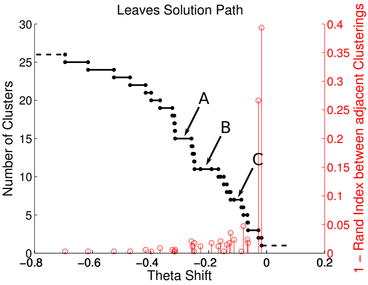
Experiment: Tracing Solution Path

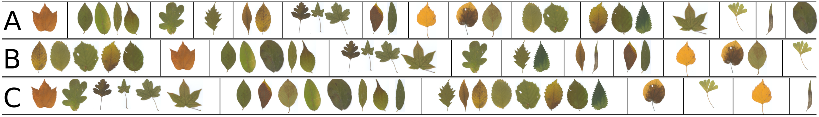
- Stability quantifies when the perturbed solution becomes suboptimal
- $\longrightarrow$ can be used to compute solution path

We can efficiently trace all solutions along a piecewise linear path in weightspace ("parametric programming").

Experiment: Tracing Solution Path

- Data: 26 classes of leaves, **w**: pairwise confusion rates from human experiments (courtesy of Frank Jäkel)
- Task: clustering leaves by human "confusion rates similarity"
- **d** = **1**, uniform bias toward fewer/more clusters
- Trace solution path, identify stable solutions





Experiment: "Critical" Edges

In some cases, stability can be visualized in the input data.

- Social network data: edges indicate social contact
- Modularity clustering: grouping
- Question: which friendships are essential in that their removal would change the grouping?
- Answer: for each friendship between i and j, check stability range for $\mathbf{d} = \mathbf{w}(E \setminus \{(i,j)\}) - \mathbf{w}(E)$

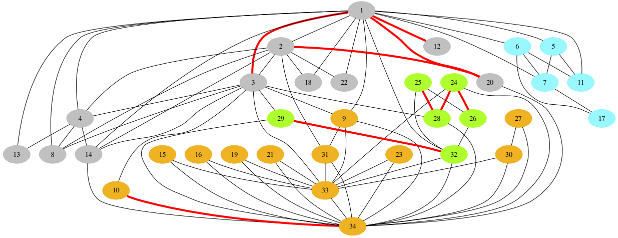


Figure: Critical edges in karate social network.

Conclusions

- Proposed a general method to quantify solution stability for combinatorial optimization problems
  - Requires only separation oracle
  - Works for problems with exponentially many inequalities
- Computed stability is conservative
- Is exact if relaxation is *locally exact*
- Allows computation of stability ranges and solution paths

References

U. Brandes, D. Delling, M. Gaertler, R. Görke, M. Hoefer, Z. Nikoloski, and D. Wagner.
On modularity clustering.
IEEE Trans. Knowl. Data Eng, 20(2):172–188, 2008.

S. Chopra and M. R. Rao.
The partition problem.
Math. Program, 59:87–115, 1993.

M. M. Deza, M. Grötschel, and M. Laurent.
Clique-web facets for multicut polytopes.
Mathematics of Operations Research, 17(4):981–1000, 1992.

M. M. Deza and M. Laurent.
Geometry of cuts and metrics.
Springer, 1997.

M. Grötschel and Y. Wakabayashi.
A cutting plane algorithm for a clustering problem.
Math. Prog., 45:59–96, 1989.

M. Grötschel and Y. Wakabayashi.
Facets of the clique partitioning polytope.
Math. Prog., 47, 1990.

B. Jansen, J. Jong, C. Roos, and T. Terlaky.
Sensitivity analysis in linear programming: Just be careful!
European Journal of Operational Research, 101:15–28, 1997.

M. E. J. Newman and M. Girvan.
Finding and evaluating community structure in networks.
Physical Review E, 69(026113), 2004.

LP-C/LP-CO tightness example

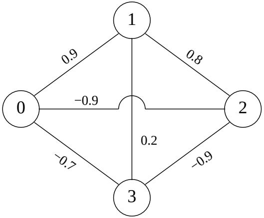


Figure: Example input graph with four vertices and edge weights as shown.

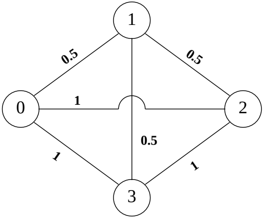


Figure: Fractional solution with $f(\mathbf{z}^*) = -1.55$, obtained by the simple LP relaxation (without odd wheel inequalities).

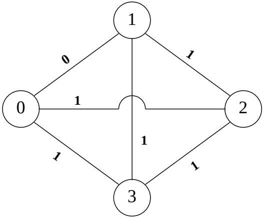


Figure: Integer solution with $f(\mathbf{z}^*) = -1.5$, obtained by adding the odd wheel inequality $z_{0,2} + z_{0,3} + z_{2,3} - z_{0,1} - z_{1,2} - z_{1,3} \leq 1$.

Limitations of the Basis Matrix Approach: Example

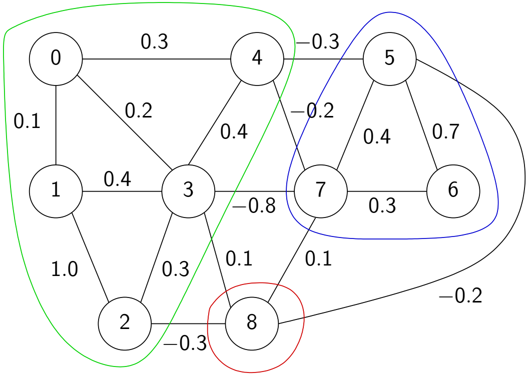


Figure: Toy example input graph with signed edge weights shown. The optimal graph partitioning has an objective of −1.6 and produces the three sets as shown.

Limitations of the Basis Matrix Approach: Example

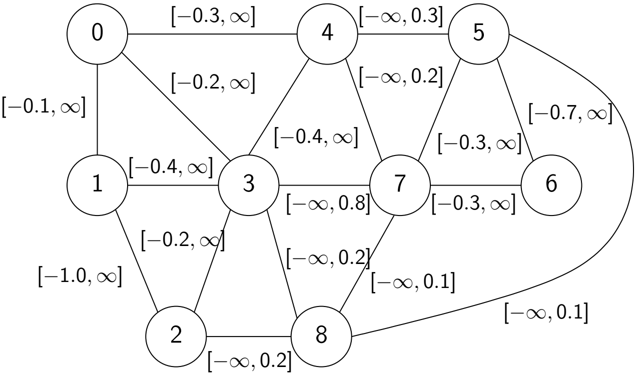


Figure: Per-edge weight sensitivities at the optimal solution, estimated by the basis matrix method.

Limitations of the Basis Matrix Approach: Example

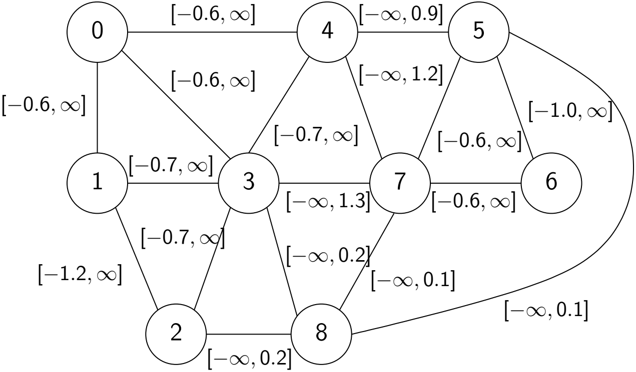


Figure: Per-edge weight sensitivities at the optimal solution, exact by the auxiliary linear programming method.