

A Message Passing Algorithm for the Minimum Cost Multicut Problem

Paul Swoboda
IST Austria

pswoboda@ist.ac.at

Bjoern Andres
MPI for Informatics

andres@mpi-inf.mpg.de

Abstract

We propose a dual decomposition and linear program relaxation of the NP-hard minimum cost multicut problem. Unlike other polyhedral relaxations of the multicut polytope, it is amenable to efficient optimization by message passing. Like other polyhedral relaxations, it can be tightened efficiently by cutting planes. We define an algorithm that alternates between message passing and efficient separation of cycle- and odd-wheel inequalities. This algorithm is more efficient than state-of-the-art algorithms based on linear programming, including algorithms written in the framework of leading commercial software, as we show in experiments with large instances of the problem from applications in computer vision, biomedical image analysis and data mining.

1. Introduction

Decomposing a graph into meaningful clusters is a fundamental primitive in computer vision, biomedical image analysis and data mining. In settings where no information is given about the number or size of clusters, and information is given only about the pairwise similarity or dissimilarity of nodes, a canonical mathematical abstraction is the minimum cost multicut (or correlation clustering) problem [17]. The feasible solutions of this problem, multicuts, relate one-to-one to the decompositions of the graph. A multicut is the set of edges that straddle distinct clusters. The cost of a multicut is the sum of costs attributed to its edges.

In the field of computer vision, the minimum cost multicut problem has been applied in [5, 6, 47, 8] to the task of unsupervised image segmentation defined by the BSDS data sets and benchmarks [36]. In the field of biomedical image analysis, the minimum cost multicut problem has been applied to an image segmentation task for connectomics [7]. In the field of data mining, applications include [9, 40, 15, 16].

Also, recently many computer vision problems that use the multicut problem as a building block have been proposed: image and mesh segmentation [31], instance-separating semantic segmentation [33], multiple object tracking [44], cell

tracking [26] and articulated human body pose estimation [3]. Moreover, one of the tightest relaxations for the multi-label Potts model is based on multicuts [30].

As the minimum cost multicut problem is NP-hard [12, 19], even for planar graphs [10] large and complex instances with millions of edges, especially those for connectomics, pose a challenge for existing algorithms.

Related Work. Due to the importance of multicuts for applications, many algorithms for the minimum cost multicut problem have been proposed. They are grouped below into three categories: primal feasible local search algorithms, linear programming algorithms and fusion algorithms.

Primal feasible local search algorithms [42, 38, 23, 21, 22] attempt to improve an initial feasible solution by means of local transformations from a set that can be indexed or searched efficiently. Local search algorithms are practical for large instances, as the cost of all operations is small compared to the cost of solving the entire problem at once. On the downside, the feasible solution that is output typically depends on the initialization. And even if a solution is found, optimality is not certified, as no lower bound is computed.

Also, the multicut problem can be transformed into a Markov random field and solved with primal heuristics there, as done for the “scribbles” dataset in [37, 11].

Linear programming algorithms [28, 29, 32, 39, 46] operate on an outer polyhedral relaxation of the feasible set. Their output is independent of their initialization and provides a lower bound. This lower bound can be used directly inside a branch-and-bound search for certified optimal solutions. Alternatively, the LP relaxation can be tightened by cutting planes. Several classes of planes are known that define a facet of the multicut polytope and can be separated efficiently [17]. On the downside, algorithms for general LPs that are agnostic to the structure of the multicut problem scale super-linearly with the size of the instance.

Fusion algorithms attempt to combine feasible solutions of subproblems obtained by combinatorial or random procedures into successively better multicuts. The fusion process can either rely on column generation [47], binary quadratic programming [14] or any algorithm for solving integer LPs [13]. In particular, [47] provides dual lower bounds

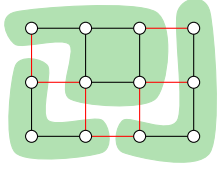


Figure 1. Decomposition of a graph into three components (green). The corresponding multicut consists of the edges straddling distinct components (red).

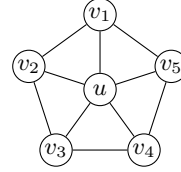


Figure 2. Odd Wheel

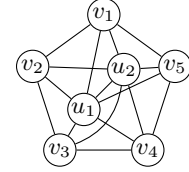


Figure 3. Odd Bicycle Wheel

but is restricted to planar graphs. [14, 13] explore the primal solution space in a clever way, but do not output dual information.

Outline. Below, a discussion of preliminaries (Sec. 2) is followed by the definition of our proposed decomposition (Sec. 3) and algorithm (Sec. 4) for the minimum cost multicut problem. Our approach combines the efficiency of local search with the lower bounds of LPs and the subproblems of fusion, as we show in experiments with large and diverse instances of the problem (Sec. 5). All code and data is available at http://github.com/pawelswoboda/LP_MP.

2. Preliminaries

2.1. Minimum Cost Multicut Problem

A *decomposition* (or *clustering*) of a graph $G = (V, E)$ is a partition $V_1 \cup \dots \cup V_k$ of the node set V such that $V_i \cap V_j = \emptyset \forall i \neq j$ and every cluster V_i , $i = 1, \dots, k$ is connected. The *multicut* induced by a decomposition is the subset of those edges that straddle distinct clusters (cf. Fig. 1). Such edges are said to be *cut*. Every multicut induced by any decomposition of G is called a multicut of G . We denote by \mathcal{M}_G the set of all multicuts of G .

Given, for every edge $e \in E$, a cost $c_e \in \mathbb{R}$ of this edge being cut, the instance of the *minimum cost multicut problem* w.r.t. these costs is the optimization problem (1) whose feasible solutions are all multicuts of G . For any edge $\{v, w\} = e \in E$, negative costs $\theta_e < 0$ favour the nodes v and w to be in distinct components. Positive costs $\theta_e > 0$ favour these nodes to lie in the same component.

$$\min_{M \in \mathcal{M}_G} \sum_{e \in M} \theta_e \quad (1)$$

This problem is NP-hard [12, 19], even for planar graphs [10]. Below, we recapitulate its formulation as a binary LP and then turn to LP relaxations: For any 01-labeling $x \in \{0, 1\}^E$ of the edges of G , the subset $x^{-1}(1)$ of those edges labeled 1 is a multicut of G if and only if x satisfies the system (3) of *cycle inequalities* [17]. Hence, (1) can be stated equivalently in the form of the binary LP (2)–(4).

$$\min_{x \in \mathbb{R}^E} \sum_{e \in E} \theta_e x_e \quad (2)$$

$$\text{subject to } \forall C \in \text{cycles}(G) : x_e \leq \sum_{e' \in C \setminus \{e\}} x_{e'} \quad (3)$$

$$x \in \{0, 1\}^E \quad (4)$$

An LP relaxation is obtained by replacing the integrality constraints (4) by $x \in P$ with $P \subseteq [0, 1]^E$. This results in an outer relaxation of the *multicut polytope*, which is the convex hull of the characteristic functions of all multicuts of G . The LP relaxation obtained for $P := [0, 1]^E$, i.e., with only the cycle inequalities, will not in general be tight.

A tighter LP relaxation is obtained by enforcing also the *odd wheel inequalities* [17]. A k -wheel is a cycle in G with k nodes all of which are connected to an additional node $u \in V$ that is not in the cycle and is called the *center* of the k -wheel (cf. Fig. 2). For any odd number $k \in \mathbb{N}$, any k -wheel of G , the cycle $C = (v_1 v_2, \dots, v_k v_1)$ and the center u of the k -wheel, every characteristic function $x \in \{0, 1\}^E$ of a multicut $x^{-1}(1)$ of G satisfies the odd wheel inequality

$$\sum_{i=1}^k x_{v_i v_{i+1}} - \sum_{i=1}^k x_{u v_i} \leq \lfloor \frac{k}{2} \rfloor \quad \text{with } v_{k+1} := v_1. \quad (5)$$

For completeness, we note that other inequalities known to further tighten the LP relaxation can be included in our algorithm, e.g., the *bicycle inequalities* [17] defined on graphs as in Fig. 3. We, however, do not consider inequalities other than cycles and odd wheels in the algorithm we propose.

2.2. Integer relaxed pairwise separable LPs

LP relaxations of the multicut problem can in principle be solved with algorithms for general LPs which are available in excellent software such as CPLEX [2] and Gurobi [25]. However, these algorithms scale super-linearly with the size of the problem and are hence impractical for large instances.

We define in Sec. 3 an LP relaxation of the multicut problem in form of an IRPS-LP (Def. 1). IRPS-LPs are a special case of dual decomposition [24]. In Def. 1, every $i \in \mathbb{V}$ defines a subproblem, and every edge $ij \in \mathbb{E}$ defines a dependency of subproblems. Def. 1 is more specific in that, firstly, the subproblems are binary and, secondly, the linear constraints (9) that describe the dependence of subproblems are defined by 01-matrices that map 01-vectors to 01-vectors. IRPS-LPs are amenable to efficient optimization by message passing in the framework of [43].

Definition 1 (IRPS-LP [43]). *Let $N \in \mathbb{N}$ and let $\mathbb{G} = (\mathbb{V}, \mathbb{E})$ be a graph with $\mathbb{V} = \{1, \dots, N\}$. For every $j \in \mathbb{V}$, let $d_j \in \mathbb{N}$, let $X_j \subseteq \{0, 1\}^{d_j}$, and let $\theta_j \in \mathbb{R}^{d_j}$. Let $\Lambda := \text{conv}(X_1) \times \dots \times \text{conv}(X_N)$. For every $\{j, k\} = e \in \mathbb{E}$, let*

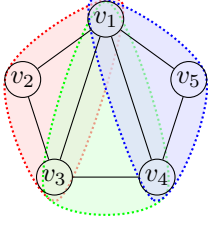


Figure 4. A triangulated cycle (black) covered by three triangles (red, green and blue)

$m_e \in \mathbb{N}$, $A_{(j,k)} \in \{0, 1\}^{m_e \times d_j}$ and $A_{(k,j)} \in \{0, 1\}^{m_e \times d_k}$ such that

$$\forall x \in X_j : A_{(j,k)}x \in \{0, 1\}^{m_e} \quad (6)$$

$$\forall x \in X_k : A_{(k,j)}x \in \{0, 1\}^{m_e} \quad (7)$$

Then, the LP written below is called integer relaxed pairwise separable w.r.t. the graph \mathbb{G} .

$$\min_{\mu \in \Lambda} \sum_{j \in \mathbb{V}} \sum_{k=1}^{d_j} \theta_{jk} \mu_{jk} \quad (8)$$

$$\text{subject to } \forall \{j, k\} \in E : A_{(j,k)}\mu_j = A_{(k,j)}\mu_k \quad (9)$$

3. Dual Decomposition

A straight-forward decomposition of the minimum cost multicut problem (2)–(4) in the form of an IRPS-LP (Def. 1) consists of one subproblem for every edge, one subproblem for every cycle inequality and one subproblem for every odd-wheel inequality. From a computational perspective, it is however advantageous to triangulate cycles and odd wheels, and to consider the resulting smaller subproblems. Below, three classes of subproblems are defined rigorously.

Edge Subproblems. For every edge $e \in E$, we consider a subproblem $e \in \mathbb{V}$ with the feasible set $X_e := \{0, 1\}$, encoding whether edge e is cut (1) or uncut (0).

Triangle Subproblems For every cycle $C = \{v_1 v_2, v_2 v_3, \dots, v_k v_1\} \subseteq E$, we consider the triangles $v_1 v_2 v_3$ to $v_{k-1} v_k v_1$, as depicted in Fig. 4. If some edge uv of a triangle C_i is not in E , we add it to E with cost zero, i.e., we triangulate the cycle in G . For each triangle uvw , we introduce a subproblem $uvw \in \mathbb{V}$ whose feasible set consists of the five feasible multicuts of the triangle, i.e., $X_{uvw} := \{(0, 0, 0), (0, 1, 1), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}$.

Lollipop Subproblems For every odd number $k \in \mathbb{N}$ and every k -wheel of G consisting of a center node u and cycle nodes v_1, \dots, v_k , we introduce two classes of subproblems. For the 5-wheel depicted in Fig. 2, these subproblems are depicted in Fig. 5.

For every $j \in \{2, \dots, k\}$, we add the triangle subproblem $uv_1 v_j \in \mathbb{V}$, as described in the previous section.

For every $j \in \{2, \dots, k-1\}$, we add the subproblem $uv_j v_{j+1}, v_1 \in \mathbb{V}$ for the lollipop graph that consists of the triangle $uv_j v_{j+1}$ and the additional edge uv_1 . The feasible set $X_{uvw,s}$ of a lollipop graph uvw, s has ten elements, five feasible multicuts of the triangle times two for accounting for the additional edge.

3.1. Dependencies

The dependency between triangle subproblems and edge subproblems are expressed below in the form of a linear system. It fits into the form (9) of an IRPS-LP.

$$\mu_{uv} = \mu_{uvw}(1, 1, 0) + \mu_{uvw}(1, 0, 1) + \mu_{uvw}(1, 1, 1)$$

$$\mu_{uw} = \mu_{uvw}(1, 1, 0) + \mu_{uvw}(0, 1, 1) + \mu_{uvw}(1, 1, 1)$$

$$\mu_{vw} = \mu_{uvw}(1, 0, 1) + \mu_{uvw}(0, 1, 1) + \mu_{uvw}(1, 1, 1)$$

The dependency between a lollipop subproblem with edge set $L = \{e_1, e_2, e_3, e_4\}$ and a triangle subproblem with edge set $T = \{e'_1, e'_2, e'_3\}$ is stated below as a linear system with sums over edges not shared between L and T . This linear system has the form (9) of an IRPS-LP.

$$\forall x_{L \cap T} : \sum_{x_{L \setminus T}} \mu_L(x_{L \cap T}, x_{L \setminus T}) = \sum_{x_{T \setminus L}} \mu_T(x_{T \cap L}, x_{T \setminus L})$$

3.2. Remarks

Remark 1. The triangulation of cycles can be understood as the constructing of a junction tree [45] in such a way that the minimum cost multicut problem over the cycle can be solved by dynamic programming. The triangulation of cycles can also be understood as a tightening of an outer polyhedral relaxation of the multicut polytope: A cycle inequality (3) defines a facet of the multicut polytope if and only if the cycle is chordless [17]. By triangulating a cycle, we obtain a set of minimal chordless cycles (triangles) whose cycle inequalities together imply that of the entire cycle.

Remark 2. Technically, we would not have needed to include triangle subproblems for odd wheels. Instead, we could have introduced dependencies between lollipops directly in the form of an IRPS-LP. However, by introducing triangle factors in addition and by expressing dependencies between lollipops and triangles, we couple lollipop factors from different odd wheels more tightly whenever they share the same triangles.

4. Algorithm

We now define an algorithm for the minimum cost multicut problem (2)–(4). This algorithm takes an instance of the problem as input and alternates for a fixed number of iterations between two main procedures.

The first procedure, defined in Sec. 4.1, solves an instance of a dual of the IRPS-LP relaxation defined in the

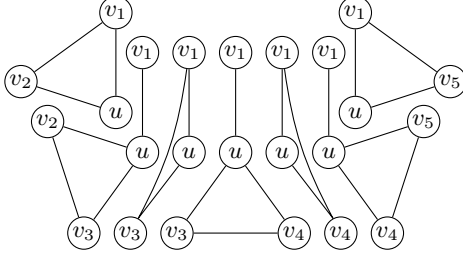


Figure 5. Triangulation of the odd wheel from Figure 2. It consists of the triangles $uv_1v_2, uv_1v_3, uv_1v_4, uv_1v_5$ and the lollipop graphs $(uv_2v_3, v_1), (uv_3v_4, v_1), (uv_4v_5, v_1)$.

previous section. The output consists in a lower bound and a re-parameterization of the instance of the minimum cost multicut problem given as input. The second procedure tightens the IRPS-LP relaxation by adding subproblems for cycle inequalities (3) and odd wheel inequalities (5) violated by the current solution. Separation procedures for finding such violated inequalities, more efficiently than in cutting plane algorithms for the primal [28, 29, 32], are defined in Sec. 4.2.

To find feasible solutions of the instance of the minimum cost multicut problem given as input, we apply a state-of-the-art local search algorithm on the computed re-parameterizations, a procedure commonly referred to as *rounding* (Sec. 4.3).

4.1. Message Passing

Like other algorithms based on dual decomposition, the algorithm we propose does not solve the IRPS-LP directly, in the primal domain, but optimizes a dual of (8)–(9). Specifically, it operates on a space of re-parametrizations of the problem defined below: For any two dependent subproblems $jk \in \mathbb{E}$, we can change the costs θ_j and θ_k by an arbitrary vector Δ according to the *update rules*

$$\theta'_j := \theta_j + A_{(j,k)}^\top \Delta \quad (10)$$

$$\theta'_k := \theta_k - A_{(k,j)}^\top \Delta \quad (11)$$

We refer to any update of θ according to the rules (10)–(11) as *message passing*. Message passing does not change the cost of any primal feasible solution, as

$$\begin{aligned} & \langle \theta'_j, \mu_j \rangle + \langle \theta'_k, \mu_k \rangle \\ &= \langle \theta_j + A_{(j,k)}^\top \Delta, \mu_j \rangle + \langle \theta_k - A_{(k,j)}^\top \Delta, \mu_k \rangle \end{aligned} \quad (12)$$

$$= \langle \theta_j, \mu_j \rangle + \langle \theta_k, \mu_k \rangle + \langle \Delta, A_{(j,k)} \mu_j - A_{(k,j)} \mu_k \rangle \quad (13)$$

$$\stackrel{(9)}{=} \langle \theta_j, \mu_j \rangle + \langle \theta_k, \mu_k \rangle \quad (14)$$

Message passing does, however, change the *dual lower bound* $L(\theta)$ to (8) given by

$$L(\theta) := \sum_{j \in \mathbb{V}} \min_{x \in X_j} \langle \theta_j, x_j \rangle \quad (15)$$

Algorithm 1: Message passing for the multicut problem

Data: $\{i_1, \dots, i_k\} = \mathbb{V}, (\theta_i)_{i \in \mathbb{V}}, (A_{(j,i)}, A_{(i,j)})_{ij \in \mathbb{E}}$

for $i = i_1, \dots, i_k$ **do**

if i is an edge subproblem uv : **then**

Receive messages:

for $w \in \mathbb{V} : uvw \in T$ **do**

$$\delta := \min_{x_{uw}, x_{vw}} \theta_{uvw}(1, x_{uw}, x_{vw}) - \min_{x_{uw}, x_{vw}} \theta_{uvw}(0, x_{uw}, x_{vw})$$

$$\theta_{uv} += \delta$$

$$\forall x_{uw}, x_{vw} : \theta_{uvw}(1, x_{uw}, x_{vw}) -= \delta$$

end

Send messages:

$$\delta := |\{w \in \mathbb{V} : uvw \in T\}|^{-1} \theta_{uv}$$

$$\theta_{uv} := 0$$

for $w \in \mathbb{V} : uvw \in T$ **do**

$$\forall x_{uw}, x_{vw} : \theta_{uvw}(1, x_{uw}, x_{vw}) += \delta$$

end

end

if i is a triangle subproblem uvw with edges C :

then

Receive messages:

for lollipops L with $L \cap C \neq \emptyset$ **do**

$$\delta(x_{L \cap C}) := \min_{x_{L \setminus C}} \theta_L(x_{L \cap C}, x_{L \setminus C})$$

$$\theta_C(x_{L \cap C}, x_{C \setminus L}) += \delta(x_{L \cap C})$$

$$\theta_L(x_{L \cap C}, x_{L \setminus C}) += \delta(x_{L \cap C})$$

end

Send messages:

$$\alpha := |\{L \text{ a lollipop} : L \cap C \neq \emptyset\}|$$

for lollipops L with $L \cap C \neq \emptyset$ **do**

$$\delta_L(x_{L \cap C}) :=$$

$$\min_{x_{C \setminus L}} \theta_{uvw}(x_{L \cap C}, x_{C \setminus L})$$

$$\theta_L(x_{L \cap C}, x_{L \setminus C}) += \frac{1}{1+\alpha} \delta_L(x_{L \cap C})$$

end

for lollipops L with $L \cap C \neq \emptyset$ **do**

$$\theta_C(x_{L \cap C}, x_{C \setminus L}) += \frac{1}{1+\alpha} \delta_L(x_{L \cap C})$$

end

end

end

The maximum of $L(\theta)$ over all costs obtainable by message passing is equal to the minimum of (8), by linear programming duality. We seek to alter the costs θ by means of message passing so as to maximize the lower bound $L(\theta)$. For the general IRPS-LP, a framework of algorithms to achieve this goal is defined in [43]. For the minimum cost multicut problem, we define and implement Alg. 1 within this framework. The specifics of this algorithm for the minimum cost multicut problem are discussed below. General properties of message passing for IRPS-LP s are discussed in [43].

Factor Order. Alg. 1 iterates through all edge and triangle subproblems. The order is specified as follows: We assume that a node order is given. With respect to this node order, edges $uv \in E$ are ordered lexicographically. For every triangle and its edge set $C = \{e_1, e_2, e_3\} \subseteq E$ with $e_1 < e_2 < e_3$, we define the ordering constraint $e_1 < C < e_3$. For every lollipop graph and its edge set $L = \{e_1, e_2, e_3, e_4\}$ with $e_1 < e_2 < e_3 < e_4$, we define the ordering constraint $e_1 < L < e_4$. The strict partial order defined by these constraints is extended to a total order by topological sorting.

Message Passing Description. When an edge subproblem $uv \in \mathbb{E}$ is visited, Alg. 1 receives messages from all dependent triangle subproblems. Having received a message from triangle $uvw \in \mathbb{E}$, the costs θ_{uvw} satisfy the condition

$$\min_{x_{uw}, x_{vw}} \theta_{uvw}(0, x_{uw}, x_{vw}) = \min_{x_{uw}, x_{vw}} \theta_{uvw}(1, x_{uw}, x_{vw}) .$$

In other words, the cost of the triangle factor θ_{uvw} has no preference for either $x_{uv} = 0$ or $x_{uv} = 1$. Sending messages from θ_{uv} is analogous: Having sent messages from uv , we have $\theta_{uv} = 0$, i.e., there is again no preference for either $x_{uv} = 0$ or $x_{uv} = 1$.

When we visit a triangle subproblem uvw , we do the analogous with all dependent lollipop subproblems: Once messages have been received, lollipop subproblems have no preference for incident edges. Once messages have been sent, this holds true for the triangle subproblems.

Once Alg. 1 has visited all subproblems and terminates, we reverse the order of subproblems and invoke Alg. 1 again. This double call of Alg. 1 is repeated for a fixed number of iterations that is a parameter of our algorithm.

4.2. Separation

Applying Alg. 1 with all cycles and all odd wheels of a graph G is impractical, as the number of triangles for cycle inequalities (3) is cubic, and the number of lollipop graphs for odd wheels (5) is quartic in $|E|$. In order to arrive at a practical algorithm, we take a cutting plane approach in which we separate and add subproblems for violated cycle and odd wheel inequalities periodically. Initially, \mathbb{V} contains only one element for every edge $e \in E$, and \mathbb{E} is empty.

In the primal, given some fractional $x \in [0, 1]^E$, it is common to look for maximally violated inequalities (3) and (5). This is possible in polynomial time via shortest path computations [17, 20]. In our dual formulation, we have no primal solution x to search for violated inequalities. Here, a suitable criterion is to consider those additional triangle or lollipop subproblems that necessarily increase the dual lower bound $L(\theta)$ by some constant $\epsilon > 0$. Among these subproblems, we choose those for which the increase is maximal and add them to the graph (\mathbb{V}, \mathbb{E}) . A similar dual cutting plane approach has shown to be useful for graphical models in [41].

Algorithm 2: Separation of cycle inequalities (3)

```

Data:  $G = (V, E)$ ,  $\epsilon \geq 0$ ,  $\theta_e \in \mathbb{R}$ 
 $l := 1$ 
for  $uv \in E$  do
  if  $\theta_{uv} \geq \epsilon$  then
     $\text{union}(u, v)$ 
  end
end
for  $uv \in E$  do
  if  $\theta_{uv} \leq -\epsilon$  and  $\text{find}(u) = \text{find}(v)$  then
     $C_l := \text{shortest-path}(u, v, \epsilon)$ 
     $l := l + 1$ 
  end
end

```

As we discuss below, separation is more efficient in the dual than in the primal.

4.2.1 Cycle Inequalities

We characterize those cycles whose subproblem increases the dual lower bound $L(\theta)$ by at least ϵ .

Proposition 1. *Let $C = \{e_1, \dots, e_k\}$ be a cycle with $\theta_{e_1} \leq -\epsilon$ and $\theta_{e_l} \leq \epsilon$ for $l > 1$. Then, the dual lower bound $L(\theta)$ can be increased by ϵ by including a triangulation of C .*

In order to find such cycles, we apply Alg. 2. This algorithm first records in a disjoint set data structure¹ whether distinct nodes $u, v \in V$ are connected via edges with weight $\geq \epsilon$. Specifically, the disjoint-set operation $\text{union}(u, v)$ in Algorithm 2 joins the connected components of u and v . Then, we visit all edges $uv \in E$ with $\theta_{uv} \leq -\epsilon$. Querying the disjoint-set datastructure via $\text{find}(u) = \text{find}(v)$ reveals whether u and v are connected via a path with edges of weight $\geq \epsilon$. If so we search for a shortest one with a breadth first search.

In the primal, finding a maximally violated cycle inequality (3) is more expensive, requiring, for every edge $uv \in E$, the search for a uv -path with minimum cost x [17] by, e.g., Dijkstra's algorithm.

4.2.2 Odd Wheel Inequalities

We characterize those odd wheels whose lollipop subproblem increases the lower bound $L(\theta)$ by at least ϵ .

Proposition 2. *Let O an odd wheel with center node u and cycle nodes v_1, \dots, v_k . Adding the lollipop subproblems for O increases $L(\theta)$ by at least ϵ if the costs $\theta_{uv_i v_{i+1}}$ of each triangle $uv_i v_{i+1}$ are such that the minimal cost of any edge*

¹https://en.wikipedia.org/wiki/Disjoint-set_data_structure

Algorithm 3: Separation of odd wheel inequalities (5)

Data: Triangles uvw , costs θ_{uvw} , $\epsilon \geq 0$
 $l := 0$
for $u \in V$ **do**
 $G' = (V', E')$, $V' = \emptyset$, $E' = \emptyset$, $Connect = \emptyset$
 for triangles uvw **do**
 if (16) holds true **then**
 $V' := V' \cup \{v, v', w, w'\}$
 $E' := E' \cup \{vw', v'w\}$
 union(v, v')
 union(w, w')
 end
 end
 for $v \in V' \cap V$ **do**
 if find(u) = find(v) **then**
 $P' := \text{shortest-path}_{G'}(u, v, \epsilon)$
 $C = \{uv \in E \mid uw' \in P' \vee u'v \in P'\}$
 if C is a simple cycle in G **then**
 $O_l := \{u, P\}$
 $l := l + 1$
 end
 end
 end
end

labeling of the triangle cutting precisely one edge incident to u is smaller by ϵ than the minimal cost of any edge labeling of the triangle cutting 0 or 2 edges incident to u . That is:

$$\begin{aligned} & \min_{\{x: x_{uv_i} + x_{uv_{i+1}} = 1\}} \theta_{uv_i v_{i+1}}(x) + \epsilon \\ \leq & \min_{\{x: x_{uv_i} + x_{uv_{i+1}} \neq 1\}} \theta_{uv_i v_{i+1}}(x) . \end{aligned} \quad (16)$$

In order to find such odd wheels, we apply Alg. 3. This algorithm builds on our observation that we need to look only at triangles whose subproblem has already been added. Hence, Alg. 3 visits each node $u \in V$ and builds a bipartite graph $G' = (V', E')$ as follows. (An example is depicted in Fig. 6 for a 5-wheel and (16) holding true for all triangles of the wheel.) For each triangle uvw such that (16) holds true, four nodes $v, v', k, k' \in V'$ are added to V' , two copies of each original node. These are joined by edges $uv', u'v \in E'$. If a path from u to u' exists in G' , we have found a violated odd wheel inequality (5). As G' is bipartite, a uu' -path in G' corresponds to an odd cycle in G . As before, the search for paths is accelerated by connectivity tests via a disjoint set data structure and is carried out by breadth first search.

In the primal, finding a maximally violated odd wheel inequality (5) entails the same construction of the bipartite graph G' for each node $u \in V$ [20]. However, a shortest path search w.r.t. edge costs $\frac{1}{2} - x_{vw} + \frac{1}{2}(x_{uv} + x_{uw})$ needs to be carried out by Dijkstra's algorithm instead of breadth

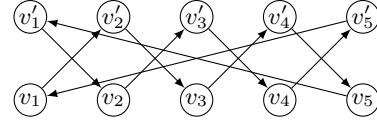


Figure 6. The bipartite graph G' constructed by Alg. 3 for separating the 5-wheel depicted in Fig. 2.

first search. Further complication in the primal comes from the fact that a separation algorithm needs to visit all $v \in V'$ in order to compute the shortest vv' -path in G' .

4.3. Rounding

Our message passing Alg. 1 improves a dual lower bound on (2), but does not provide a feasible solution of (2)–(4). In order to obtain a feasible multicut, we apply a local search algorithm defined in [31], namely greedy additive edge contraction (GAEC), followed by Kernighan-Lin with joins (KLj). GAEC computes a multicut by greedily contracting those edges for which the join decreases the cost maximally. It stops as soon as no contraction of any edge strictly decreases the cost. KLj attempts to improve a given multicut recursively by applying transformations from three classes: (1) moving nodes between two components, (2) moving nodes from a given component to a newly forming one or (3) joining two components. GAEC and KLj are local search algorithms that output a feasible multicut that need not be optimal.

We apply GAEC and KLj not only to the instance of the minimum cost multicut problem given as input but also to the re-parameterization of this instance output by Alg. 1. The rationale for doing so comes from LP duality:

Proposition 3. Assume θ maximizes the dual lower bound $L(\theta)$ and the relaxation is tight, i.e.

$$L(\theta) = \min_{\{x \in \{0,1\}^E \mid x^{-1}(1) \in \mathcal{M}_G\}} \langle \theta, x \rangle . \quad (17)$$

Moreover, let $\hat{x} \in \{0,1\}^E$ such that $\hat{x}^{-1}(1)$ is an optimal multicut of G . Then,

$$\theta_e \begin{cases} \leq 0 & \text{if } \hat{x}_e = 1 \\ \geq 0 & \text{if } \hat{x}_e = 0 \end{cases} \quad (18)$$

Having run Alg. 1 for a while, we expect θ to fulfill the sign condition of Prop. 3 approximately. Therefore, the sign of θ_e will be a good hint of the edge e being cut. Thus, informally, we expect local search algorithms operating on the re-parameterized instance of the problem to yield better feasible multicuts than local search algorithms operating on the given instance.

For MAP-inference in Markov random fields it is known [34, 35] that primal rounding can be improved greatly when applied on costs re-parameterized by message passing.

5. Experiments

Solvers We compare against several state of the art algorithms.

- The algorithm **MC-ILP** [29] is an efficient implementation of a cutting plane algorithm solving (2) using cycle inequalities (3) in a cutting plane fashion. CPLEX [2] is used to solve the underlying ILP problems. The integrality conditions in (4) are directly given to the solver. According to [29] this is beneficial due to the excellent branch and cut capabilities of CPLEX [2].
- Cut, Glue & Cut [14], abbreviated as **CGC**, is a move making algorithm using planar max-cut subproblems to improve multicuts.
- Fusion moves for correlation clustering [13], abbreviated as **CC-Fusion**, fuses multicuts generated by various proposal generator with the help of auxiliary multicut problems, solved in turn by **MC-ILP**. We use randomized hierarchical clustering and randomized watersheds as proposal generators, identified by the suffixes **-RHC** and **-RWS**. We use parameters for the proposal generators as recommended by the authors [13].
- **MP-C** denotes Algorithm 1 when we only separate for cycle inequalities (3) by Algorithm 2, while **MP-COW** denotes that we additionally separate for odd wheel inequalities (5) by Algorithm 3. We search for triangles and lollipops to add every 10th iteration.
- **KL** is the GAEC and KLj implementation [31] described in Section 4.3 for computing multicuts. We let **KL** run every 100th iteration of **MP-C** and **MP-COW** on the current reparametrized edge costs.

MC-ILP, **CGC** and **CC-Fusion** are implemented as part of the OpenGM suite [27]. Only **MC-ILP** and our solvers **MP-C** and **MP-COW** generate dual lower bounds. **CGC** outputs the trivial dual lower bound $\sum_{e \in E} \min(0, \theta_e)$, with edge weights θ_e as given by the problem. It has been shown that **CGC**, **CC-Fusion** and **KL** outperform other primal heuristics [13], hence we do not compare to any other heuristic algorithm. Also **MC-ILP** outperforms the LP-based solver [39], due to the latter using the slower COIN-OR CLP [18] solver internally, hence we exclude it from the comparison as well.

All solvers were run on a laptop computer with a i5-5200 CPU with 2.2 GHz and 8GB RAM.

Datasets We compare on 8 datasets of diverse origin.

- `image-seg` consists of images of the Berkeley segmentation dataset [36], presegmented with superpixels, for which pairwise affinity values have been computed as in [6].

- The `knott-3d-{150|300|450|550}` datasets come from a neural circuit reconstruction problem of tissue [7] with $[150]^3$, $[300]^3$, $[450]^3$ and $[900]^3$ voxels. The data is presegmented into supervoxels.
- `modularity clustering` aims to cluster a social network into subgroups based on affinity between individual persons.
- `CREMI-{small|large}` datasets [4] were constructed as part of the CREMI [1] challenge, which aims to reconstruct neural circuits of the adult fly brain. The images are taken by electron microscopy. The `-small` instances are cropped versions of the `-large` ones. To our knowledge, the `CREMI-large` dataset contain the largest multicut problems approached with LP-based methods.

The `image-seg`, `knott-3d` and `modularity clustering` datasets were taken from the OpenGM benchmark [27], while the CREMI datasets were kindly provided by their authors and are not yet published.

The dataset consists of 100, 8, 8, 8, 8, 6, 3 and 3 instances, in total 144. Dataset details can be found in Table 1.

Evaluation We have set a timelimit of one hour for all algorithms, but exit early when the primal/dual gap vanishes or no progress can be observed anymore. In Table 1 results averaged over all instances in specific datasets are reported. In Figure 7 primal solution energy and dual lower bound (where applicable) averaged over all instances in specific datasets are drawn against runtime.

As can be seen from Table 1, the primal rounding heuristics, while faster, never give better primal energies than the LP-based approaches **MC-ILP**, **MP-C** and **MP-COW**. The branch-and-cut solver **MC-ILP** outperforms our algorithms **MP-C** and **MP-COW** on the small `image-seg`, `knott-3d-150` and `knott-3d-300` instances and has higher lower bounds for the `CREMI-small` problems. Whenever **MC-ILP** beats our algorithms, it is by a very small margin though. On the other hand, **MC-ILP** is significantly slower on all problems and does not even finish a single iteration on the large `CREMI-large` instances. On the larger `knott-3d-450`, `knott-3d-550` and `CREMI-large` datasets, our algorithms outperform **MC-ILP** markedly. As can be seen from Fig. 7 our dual lower and upper primal bounds usually converge faster than **MC-ILP**'s.

We conjecture that **MP-C** and **MP-COW** inside a branch-and-bound solver can significantly extend the reach of exact methods for the multicut problem and close the gap to **MC-ILP** on the smaller datasets. Also, unlike **MC-ILP**, our reparametrized costs can be used to improve heuristic primal algorithms. An example of this can be seen in Fig. 8, where reparametrized costs improve **KL**'s solutions.

Dataset / Algorithm			MP-C	MP-COW	CGC	MC-ILP	CC-Fusion-RWS	CC-Fusion-RHC	
image-seg	#I	100	UB	4436.25	4435.94	4600.81	4434.91	4447.06	4436.33
	#V	< 3764	LB	4434.17	4434.44	4129.70	4434.91	‡	‡
	#E	< 10970	time(s)	9.05	21.85	0.14	11.89	1.19	1.30
modularity clustering	#I	6	UB	-0.49	-0.49	-0.30	-0.44	0.00	-0.44
	#V	< 115	LB	-0.54	-0.52	-0.79	-0.52	‡	‡
	#E	< 6555	time(s)	87.02	1071.17	0.15	2911.10	0.00	17.78
knott-3d-150	#I	8	UB	-4571.21	-4571.65	-4220.66	-4571.69	-4534.76	-4552.51
	#V	< 972	LB	-4572.21	-4571.72	-4855.18	-4571.69	‡	‡
	#E	< 5656	time(s)	0.75	0.81	0.04	2.37	0.26	0.53
knott-3d-300	#I	8	UB	-27299.78	-27301.97	-24864.59	-27302.78	-27242.03	-27247.29
	#V	< 5896	LB	-27304.96	-27303.16	-28901.58	-27302.78	‡	‡
	#E	< 36221	time(s)	19.33	34.27	2.73	227.33	2.96	8.15
knott-3d-450	#I	8	UB	-78466.04	-78472.24	-70865.27	-78391.32	-78386.14	-78381.06
	#V	< 17074	LB	-78485.04	-78481.01	-83272.85	-78522.51	‡	‡
	#E	< 107060	time(s)	385.14	598.60	31.56	1840.47	16.52	119.23
knott-3d-550	#I	8	UB	-136517.72	-136523.39	-123841.47	-135766.90	-136464.05	-136395.89
	#V	< 31249	LB	-136570.96	-136564.10	-144703.64	-136755.36	‡	‡
	#E	< 195271	time(s)	1804.28	2218.64	102.42	3683.22	72.94	594.60
CREMI-small	#I	3	UB	-213189.20	-213193.09	-194616.60	-209594.49	-168905.17	-213117.84
	#V	< 35523	LB	-213212.43	-213209.67	-215473.98	-213208.94	‡	‡
	#E	< 235966	time(s)	915.70	1238.59	319.01	2775.81	3543.61	2555.48
CREMI-large	#I	3	UB	-3887113.00	-3887078.88	‡	‡	-3772597.37	-3619190.20
	#V	< 623435	LB	-3888461.86	-3888482.06	‡	‡	‡	‡
	#E	< 4172314	time(s)	4088.51	4093.91	‡	‡	5978.08	23139.40

Table 1. Primal solution energy (UB)/dual lower bound (LB)/runtime in seconds averaged over all instances of datasets. #I is number of instances in dataset, #V and #E are number of vertices and edges in multicut instances. † means method did not finish after one hour. ‡ means method does not output dual lower bound. **Bold** numbers denote lowest primal solution energy, highest lower bound, fastest runtime.

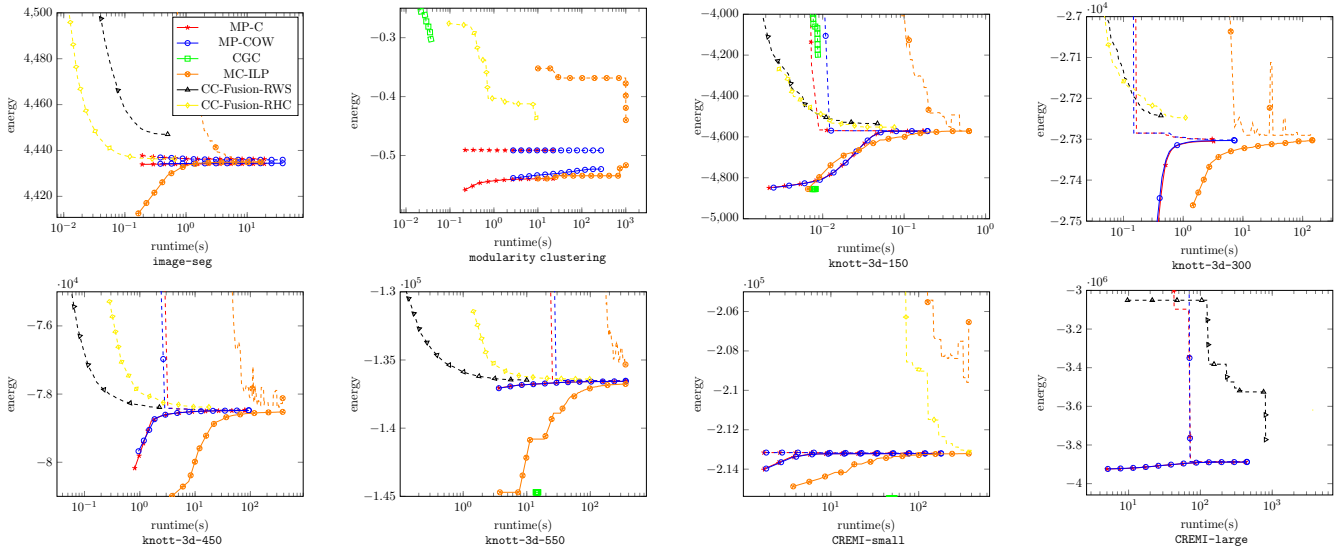


Figure 7. Averaged runtime plots for image-seg, modularity clustering, knott-3d-150, knott-3d-300, knott-3d-450, knott-3d-550, CREMI-small and CREMI-large datasets. Continuous lines denote dual lower bounds and dashed ones primal energies. Values are averaged over all instances of the dataset. The x-axis is logarithmic.

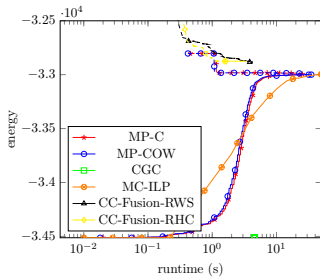


Figure 8. Instance gm_knott_3d_072 from dataset knott-3d-300 where reparametrized costs improve KL's solutions.

6. Acknowledgments

The authors would like to thank Vladimir Kolmogorov for helpful discussions and gratefully acknowledge Fred Hamprecht's and Thorsten Beier's help in comparing OpenGM's multicut solvers. This work is partially funded by the European Research Council under the European Unions Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no 616160. We thank anonymous reviewers for pointing out the references [37, 11].

References

- [1] CREMI MICCAI Challenge on circuit reconstruction from Electron Microscopy Images. <https://cremi.org>.
- [2] IBM ILOG CPLEX Optimizer. <http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/>.
- [3] *DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation.*, June 2016.
- [4] Multicut brings automated neurite segmentation closer to human performance. *Nature Methods*, 14:101–102, 2017.
- [5] A. Alush and J. Goldberger. Break and conquer: Efficient correlation clustering for image segmentation. In E. R. Hancock and M. Pelillo, editors, *SIMBAD*, volume 7953 of *Lecture Notes in Computer Science*, pages 134–147. Springer, 2013.
- [6] B. Andres, J. H. Kappes, T. Beier, U. Köthe, and F. A. Hamprecht. Probabilistic image segmentation with closedness constraints. In D. N. Metaxas, L. Quan, A. Sanfeliu, and L. J. V. Gool, editors, *ICCV*, pages 2611–2618. IEEE Computer Society, 2011.
- [7] B. Andres, T. Kröger, K. L. Briggman, W. Denk, N. Korogod, G. Knott, U. Köthe, and F. A. Hamprecht. Globally optimal closed-surface segmentation for connectomics. In A. W. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, editors, *ECCV (3)*, volume 7574 of *Lecture Notes in Computer Science*, pages 778–791. Springer, 2012.
- [8] B. Andres, J. Yarkony, B. S. Manjunath, S. Kirchhoff, E. Turetken, C. C. Fowlkes, and H. Pfister. Segmenting planar superpixel adjacency graphs w.r.t. non-planar superpixel affinity graphs. *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, 2013.
- [9] A. Arasu, C. Ré, and D. Suci. Large-scale deduplication with constraints using dedupalog. In Y. E. Ioannidis, D. L. Lee, and R. T. Ng, editors, *ICDE*, pages 952–963. IEEE Computer Society, 2009.
- [10] Y. Bachrach, P. Kohli, V. Kolmogorov, and M. Zadimoghadam. Optimal coalition structure generation in cooperative graph games. In *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence, July 14-18, 2013, Bellevue, Washington, USA.*, 2013.
- [11] S. Bagon and M. Galun. A multiscale framework for challenging discrete optimization. In *NIPS*, 2012.
- [12] N. Bansal, A. Blum, and S. Chawla. Correlation clustering. *Machine Learning*, 56(1):89–113, 2004.
- [13] T. Beier, F. A. Hamprecht, and J. H. Kappes. Fusion moves for correlation clustering. In *CVPR*, pages 3507–3516. IEEE Computer Society, 2015.
- [14] T. Beier, T. Kröger, J. H. Kappes, U. Köthe, and F. A. Hamprecht. Cut, glue & cut: A fast, approximate solver for multicut partitioning. In *CVPR. Proceedings*, 2014.
- [15] Y. Chen, S. Sanghavi, and H. Xu. Clustering sparse graphs. In P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *NIPS*, pages 2213–2221, 2012.
- [16] F. Chierichetti, N. Dalvi, and R. Kumar. Correlation clustering in mapreduce. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14*, pages 641–650, New York, NY, USA, 2014. ACM.
- [17] S. Chopra and M. R. Rao. The partition problem. *Mathematical Programming*, 59(1):87–115, 1993.
- [18] COIN-OR CLP, 2016. <http://www.coin-or.org/projects/Clp.xml>.
- [19] E. D. Demaine, D. Emanuel, A. Fiat, and N. Immerlica. Correlation clustering in general weighted graphs. *Theor. Comput. Sci.*, 361(2):172–187, Sept. 2006.
- [20] M. M. Deza and M. Laurent. *Geometry of Cuts and Metrics*. Springer Publishing Company, Incorporated, 1st edition, 2009.
- [21] M. Elsner and E. Charniak. You talking to me? a corpus and algorithm for conversation disentanglement. In K. McKeown, J. D. Moore, S. Teufel, J. Allan, and S. Furui, editors, *ACL*, pages 834–842. The Association for Computer Linguistics, 2008.
- [22] M. Elsner and W. Schudy. Bounding and comparing methods for correlation clustering beyond ILP. In *Proceedings of the Workshop on Integer Linear Programming for Natural Language Processing, ILP '09*, pages 19–27, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.
- [23] A. Gionis, H. Mannila, and P. Tsaparas. Clustering aggregation. *ACM Trans. Knowl. Discov. Data*, 1(1):4, 2007.
- [24] M. Guignard and S. Kim. Lagrangean decomposition for integer programming: theory and applications. *Revue française d'automatique, d'informatique et de recherche opérationnelle. Recherche opérationnelle*, 21(4):307–323, 1987.
- [25] Gurobi Optimization, Inc., 2015. <http://www.gurobi.com>.
- [26] F. Jug, E. Levinkov, C. Blasse, E. W. Myers, B. Andres, undefined, undefined, and undefined. Moral lineage tracing. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 00:5926–5935, 2016.
- [27] J. H. Kappes, B. Andres, F. A. Hamprecht, C. Schnörr, S. Nowozin, D. Batra, S. Kim, B. X. Kausler, T. Kröger, J. Lellmann, N. Komodakis, B. Savchynskyy, and C. Rother. A comparative study of modern inference techniques for structured discrete energy minimization problems. *International Journal of Computer Vision*, 115(2):155–184, 2015.
- [28] J. H. Kappes, M. Speth, B. Andres, G. Reinelt, and C. Schnörr. Globally optimal image partitioning by multicuts. In *EMMCVPR*. Springer, Springer, 2011.
- [29] J. H. Kappes, M. Speth, G. Reinelt, and C. Schnörr. Higher-order segmentation via multicuts. *CoRR*, abs/1305.6387, 2013.
- [30] J. H. Kappes, M. Speth, G. Reinelt, and C. Schnörr. Higher-order segmentation via multicuts. *Computer Vision and Image Understanding*, 143:104–119, 2016.
- [31] M. Keuper, E. Levinkov, N. Bonneel, G. Lavoué, T. Brox, and B. Andres. Efficient decomposition of image and mesh graphs by lifted multicuts. In *ICCV*, 2015.
- [32] S. Kim, S. Nowozin, P. Kohli, and C. D. Yoo. Higher-order correlation clustering for image segmentation. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. C. N. Pereira, and K. Q. Weinberger, editors, *NIPS*, pages 1530–1538, 2011.
- [33] A. Kirillov, E. Levinkov, B. Andres, B. Savchynskyy, and C. Rother. InstanceCut: from edges to instances with MultiCut, 2016.

- [34] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(10):1568–1583, 2006.
- [35] V. Kolmogorov. A new look at reweighted message passing. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(5):919–930, 2015.
- [36] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 416–423 vol.2, 2001.
- [37] O. Meir, M. Galun, S. Yagev, R. Basri, I. Yavneh, undefined, undefined, undefined, and undefined. A multiscale variable-grouping framework for mrf energy minimization. volume 00, pages 1805–1813, Los Alamitos, CA, USA, 2015. IEEE Computer Society.
- [38] V. Ng and C. Cardie. Improving machine learning approaches to coreference resolution. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*, (July):104, 2001.
- [39] S. Nowozin and S. Jegelka. Solution stability in linear programming relaxations: graph partitioning and unsupervised learning. In A. P. Danyluk, L. Bottou, and M. L. Littman, editors, *ICML*, volume 382 of *ACM International Conference Proceeding Series*, pages 769–776. ACM, 2009.
- [40] E. Sadikov, J. Madhavan, L. Wang, and A. Halevy. Clustering query refinements by user intent. In *World Wide Web Conference (WWW)*. ACM Press, April 2010.
- [41] D. Sontag, D. K. Choe, and Y. Li. Efficiently searching for frustrated cycles in MAP inference. In *UAI*, pages 795–804. AUAI Press, 2012.
- [42] W. M. Soon, H. T. Ng, and D. C. Y. Lim. A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 27(4):521–544, 2001.
- [43] P. Swoboda, J. Kuske, and B. Savchynskyy. A dual ascent framework for Lagrangean decomposition of combinatorial problems. In *CVPR*, 2017.
- [44] S. Tang, B. Andres, M. Andriluka, and B. Schiele. Subgraph decomposition for multi-object tracking. In *Computer Vision and Pattern Recognition (CVPR)*, June 2015.
- [45] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.
- [46] J. Yarkony, T. Beier, P. Baldi, and F. A. Hamprecht. Parallel multicut segmentation via dual decomposition. In *New Frontiers in Mining Complex Patterns - Third International Workshop, NFMCP 2014, Held in Conjunction with ECML-PKDD 2014, Nancy, France, September 19, 2014, Revised Selected Papers*, pages 56–68, 2014.
- [47] J. Yarkony, A. Ihler, and C. C. Fowlkes. *Fast Planar Correlation Clustering for Image Segmentation*, pages 568–581. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.