Probabilistic Correlation Clustering and Image Partitioning Using Perturbed Multicuts

Jörg Hendrik Kappes[1], Paul Swoboda[2], Bogdan Savchynskyy[1], Tamir Hazan[3], and Christoph Schnörr[1,2]

[1] Heidelberg Collaboratory for Image Processing, Heidelberg University
[2] Image and Pattern Analysis Group, Heidelberg University
[3] Department of Computer Science, University of Haifa

Abstract. We exploit recent progress on globally optimal MAP inference by integer programming and perturbation-based approximations of the log-partition function. This enables to locally represent uncertainty of image partitions by approximate marginal distributions in a mathematically substantiated way, and to rectify local data term cues so as to close contours and to obtain valid partitions. Our approach works for any graphically represented problem instance of correlation clustering, which is demonstrated by an additional social network example.

Keywords: Correlation Clustering, Multicut, Perturb and MAP

1 Introduction

Clustering, image partitioning and related NP-hard decision problems abound in the fields image analysis, computer vision, machine learning and data mining, and much research has been done on alleviating the combinatorial difficulty of such inference problems using various forms of relaxations. A recent assessment of the state-of-the-art using discrete graphical models has been provided by [13]. A subset of specific problem instances considered there (Potts-like functional minimisation) are closely related to continuous formulations investigated, e.g., by [7, 17].

From the viewpoint of statistics and Bayesian inference, such *Maximum-A-Posteriori (MAP)* point estimates have been always criticised as falling short of the scope of probabilistic inference, that is to provide – along with the MAP estimate – "error bars" that enable to assess sensitivities and uncertainties for further data analysis. Approaches to this more general objective are less uniquely defined than the MAP problem. For example, a variety of approaches have been suggested from the viewpoint of clustering (see more comments and references below) which, on the other hand, differ from the variational marginalisation problem in connection with discrete graphical models [25]. From the computational viewpoint, these more general problems are not less involved than the corresponding MAP(-like) combinatorial inference problems.

In this paper, we consider the general multicut problem [8], also known as correlation clustering in other fields [5], which includes the image partitioning problem as special case. Our work is based on

(i) recent progress [19, 10] on the probabilistic analysis of perturbed MAP problems applied to our setting in order to establish a sound link to basic variational approximations of inference problems [25],
(ii) recent progress on *exact* solvers of the multicut problem [15, 16], which is required in connection with (i).

Figure 1 provides a first illustration of our approach. Our general problem formulation enables to address not only the image partitioning problem. We demonstrate this in the experimental section by applying correlation clustering to a problem instance from machine learning that involves network data on a general graph.

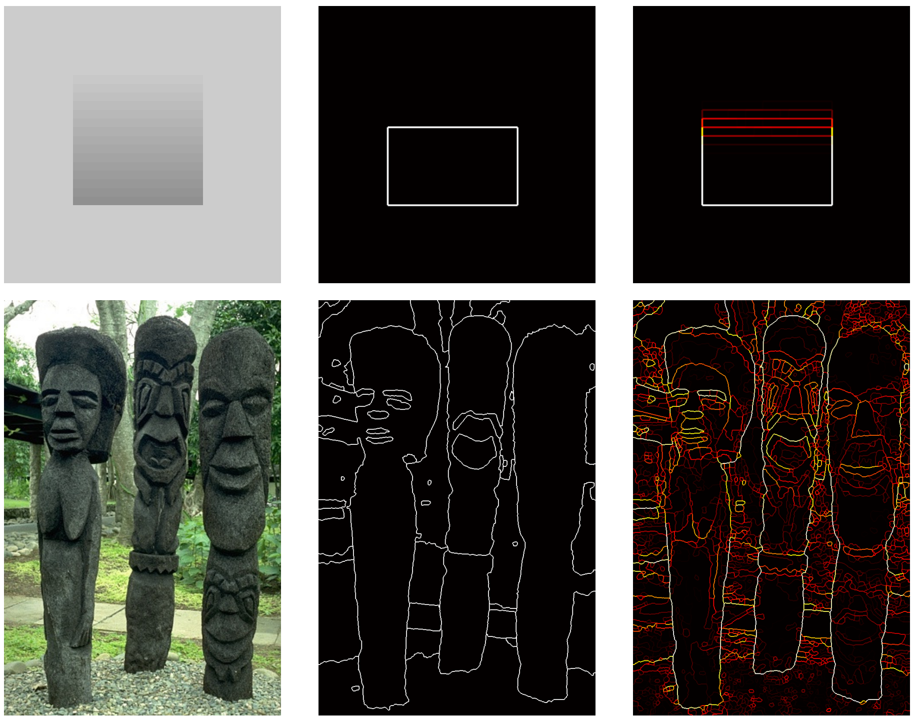


Fig. 1. Two examples demonstrating our approach. **Left column:** images subject to unsupervised partitioning. **Center column:** globally optimal partitions. **Right column:** probabilistic inference provided along with the partition. The color order: white → yellow → red → black, together with decreasing brightness, indicate uncertainty, cf. Fig. 2. We point out that *all local* information provided by our approach is intrinsically *non-locally* inferred and relates to partitions, that is to *closed* contours.

Related Work. The susceptibility of clustering to noise is well known. This concerns, in particular, clustering approaches to image partitioning that typically employ spectral relaxation [22, 12, 23]. Measures proposed in the literature [24] to quantitatively assess confidence in terms of stability, employ data perturbations and various forms of cluster averaging. While this is intuitively plausible, a theoretically more convincing substantiation seems to be lacking, however.

In [11], a deterministic annealing approach to the unsupervised graph partitioning problem (called pairwise clustering) was proposed by adding an entropy term weighted by an artificial temperature parameter. Unlike the simpler continuation method of Blake and Zisserman [6], this way of smoothing the combinatorial partitioning problem resembles the variational transition from marginalisation to MAP estimation, by applying the log-exponential function to the latter objective [25]. As in [6], however, the primary objective of [11] is to compute a single "good" local optimum by solving a sequence of increasingly non-convex problems parametrised by an artificial temperature parameter, rather than sampling various "ground states" (close to zero-temperature solutions) in order to assess stability, and to explicitly compute alternatives to the single MAP solution. The latter has been achieved in [18] using a non-parametric Bayesian framework. Due to the complexity of model evaluation, however, authors have to resort to MCMC sampling.

Concerning continuous problem formulations, a remarkable approach to assess "error bars" of variational segmentations has been suggested by [20]. Here, the starting point is the "smoothed" version of the Mumford-Shah functional in terms of the relaxation of Ambrosio and Tortorelli [2] that is known to Γ-converge to the Mumford-Shah functional in the limit of corresponding parameter values. Authors of [20] apply a particular perturbation ("polynomial chaos") that enables to locally infer confidence of the segmentation result. Although being similar in scope to our approach, this approach is quite different. An obvious drawback results from the fact that minima of the Ambrosio-Tortorelli functional do not enforce partitions, i.e. may involve contours that are not closed.

Finally, we mention recent work [21] that addresses the same problem using – again – a quite different approach: "stochastic" in [21] just refers to the relaxation of binary indicator vectors to the probability simplex, and this relaxation is solved by a *local* minimisation method. Our approach, on the other hand, is based on random perturbations of *exact* solutions of the correlation clustering problem. This yields a truly probabilistic interpretation in terms of the induced approximation of the log-partition function, whose derivatives generate the expected values of the variables of interest.

Organization. Sec. 2 defines the combinatorial correlation clustering problem and introduces multicuts. The variational formulation for probabilistic inference is presented in Sec. 3, followed by the perturbation approach in Sec. 4. A range of experiments demonstrate the approach in Sec. 5. Since alternative approaches rely on quite different methods, as explained above, a re-implementation is beyond the scope of this paper. We therefore restrict our comparison to the evaluation of *local* potentials that we consider as an efficient alternative. This comparison reveals that contrary to this local method, our perturbation approach effectively enforces *global* topological constraints so as to sample from most likely partitions.

Basic Notation. We set $[n] := \{1, 2, \ldots, n\}$, $n \in \mathbb{N}$ and use the indicator function $\mathbb{I}(p) = 1$ if the predicate p is true, and $\mathbb{I}(p) = 0$ otherwise. $|S|$ denotes the cardinality of a finite set S. $\langle x, y\rangle = \sum_{i\in[n]} x_i y_i$ denotes the Euclidean inner product of vectors $x, y \in \mathbb{R}^n$. $\mathbb{E}[X]$ denotes the expected value of a random variable X. $\Pr[\Omega]$ denotes the probability of an event Ω.

2 Correlation Clustering and Multicuts

The correlation clustering problem is defined in terms of partitions of an undirected weighted graph

$$G = (V, E, w), \quad V = [n], \quad E \subseteq V \times V, \tag{1a}$$

$$w \colon E \to \mathbb{R}, \qquad e \mapsto w_e := w(e) \tag{1b}$$

with signed edge-weight function w. A positive weight $w_e > 0$, $e \in E$ indicates that two adjacent nodes should be merged, whereas a negative weight indicates that these nodes should be separated into distinct clusters S_i, S_j.

We formally define valid partitions and interchangeably call them segmentations or clusterings.

Definition 1 (partition, segmentation, clustering). *A set of subsets* $\{S_1, \ldots, S_k\}$*, called* shores, components or clusters*, is a* (valid) partition *of a graph* $G = (V, E, w)$ *iff* ***(a)*** $S_i \subseteq V$, $i \in [k]$, ***(b)*** $S_i \neq \emptyset$, $i \in [k]$, ***(c)*** *the induced subgraphs* $G_i := \big(S_i, (S_i \times S_i) \cap E\big)$ *are connected,* ***(d)*** $\bigcup_{i\in[k]} S_i = V$, ***(e)*** $S_i \cap S_j = \emptyset$, $i, j \in [k]$, $i \neq j$. *The set of all valid partitions of* G *is denoted by* $\mathcal{S}(G)$.

The number $|\mathcal{S}(G)|$ of all possible partitions is upper-bounded by the Bell number [1] that grows very quickly with $|V|$.

The *correlation clustering* or *minimal cost multicut problem* is to find a partition that minimizes the cost of intra cluster edges as defined by the weight function w. This problem can be formulated as a minimization problem of a Potts model

$$\arg\min_{x\in V^{|V|}} \sum_{ij\in E} w_{ij} \mathbb{I}(x_i \neq x_j). \tag{2}$$

Because any node can form its own cluster, $|V|$ labels are needed to represent all possible assignments in terms of variables x_i, $i \in V$.

A major drawback of this formulation is the huge inflated space representing the assignments. Furthermore, due to the lack of an external field (unary terms), any permutation of an optimal assignment results in another optimal labeling. As a consequence, the standard relaxation in terms of the so-called local polytope [25] becomes too weak.

In order to overcome these problems, we adopt an alternative representation of partitions based on the set of *inter* cluster edges [8]. We call the edge set

$$\delta(S_1, \ldots, S_k) := \big\{uv \in E \colon u \in S_i,\, v \in S_j,\, i \neq j,\, i, j \in [k]\big\} \tag{3}$$

a *multicut*. To obtain a polyhedral representation of multicuts, we define *indicator vectors* $\chi(E') \in \{0,1\}^{|E|}$ for each subset $E' \subseteq E$ by

$$\chi_e(E') := \begin{cases} 1, & \text{if } e \in E', \\ 0, & \text{if } e \in E \setminus E'. \end{cases}$$

The *multicut polytope* $\mathcal{MC}(G)$ then is given by the convex hull

$$\mathcal{MC}(G) := \operatorname{conv}\big\{\chi\big(\delta(S)\big) \colon S \in \mathcal{S}(G)\big\}. \tag{4}$$

The vertices of this polytope are the indicator functions of valid partitions and denoted by

$$\mathcal{Y}(G) := \big\{\chi\big(\delta(S)\big) \colon S \in \mathcal{S}(G)\big\}. \tag{5}$$

The *correlation clustering problem* then amounts to find a partition $S \in \mathcal{S}(G)$ that minimizes the sum of the weights of edges cut by the partition

$$\arg\min_{S \in \mathcal{S}(G)} \sum_{e \in E} w_e \cdot \chi_e(\delta(S)) = \arg\min_{y \in \mathcal{MC}(G)} \sum_{e \in E} w_e \cdot y_e. \tag{6}$$

Although problem (6) is a linear program, solving it is NP-hard, because a representation of the multicut polytope $\mathcal{MC}(G)$ by half-spaces is of exponential size and moreover, unless $P = NP$, it is not separable in polynomial time. However, one can develop efficient separation procedures for an outer relaxation of the multicut polytope which involves all facet-defining cycle inequalities. Together with integrality constraints, this guarantees globally optimal solutions of problem (6) and performs best on benchmark datasets [14, 13].

3 Probabilistic Correlation Clustering

A major limitation of solutions to the correlation clustering problem is that the most likely segmentations are returned without any measurement of the corresponding uncertainty. To overcome this, one would like to compute the marginal probability that an edge is an inter-cluster edge or, in other words, that an edge is cut.

The most direct approach to accomplish this is to associate a Gibbs distribution with the Potts model in (2)

$$p(x|w,\beta) = \exp\Big(-\beta \sum_{ij \in E} w_{ij} \mathbb{I}(x_i \neq x_j) - \log\big(Z_x(w,\beta)\big)\Big), \tag{7a}$$

$$Z_x(w,\beta) = \sum_{x \in \mathcal{X}} \exp\Big(-\beta \sum_{ij \in E} w_{ij} \mathbb{I}(x_i \neq x_j)\Big), \tag{7b}$$

where $\mathcal{X}$ denotes the feasible set of (2)

$$\mathcal{X} := \mathcal{X}_1 \times \ldots \times \mathcal{X}_{|V|} := V^{|V|}, \qquad \mathcal{X}_i = V,\ i \in V. \tag{8}$$

Parameter β is a free parameter (in physics: "inverse temperature") and $Z(w,\beta)$ the partition function. Performing the reformulation

$$-\beta \sum_{ij\in E} w_{ij}\mathbb{I}(x_i \neq x_j) = \sum_{ij\in E} \sum_{\substack{x'_i\in\mathcal{X}_i \\ x'_j\in\mathcal{X}_j}} \underbrace{-\beta\, w_{ij}\, \mathbb{I}(x'_i \neq x'_j)}_{:=\theta_{ij;x'_i,x'_j}} \cdot \underbrace{\mathbb{I}(x_i = x'_i \vee x_j = x'_j)}_{:=\phi_{ij;;x'_i,x'_j}(x)}, \quad (9)$$

we recognise the distribution as a member of the exponential family with model parameter θ and sufficient statistics $\phi(x)$:

$$p(x|\theta) = \exp\big(\langle\theta,\phi(x)\rangle - \log\big(Z_x(\theta)\big)\big), \quad (10a)$$

$$Z_x(\theta) = \sum_{x\in\mathcal{X}} \exp\big(\langle\theta,\phi(x)\rangle\big). \quad (10b)$$

Note that the dimension $d = |V|\cdot|V|\cdot|E|$ of the vectors θ,ϕ is large. Therefore, while (10) in principle provides the "correct" basis for assessing uncertainty in terms of marginal distributions $p(x_i,x_j|\theta)$, $ij \in E$, this is infeasible computationally due to the huge space X and the aforementioned permutation invariance.

To overcome this problem, we resort to the problem formulation (6) in terms of multicuts, define the model parameter vector θ and the sufficient statistics $\phi(y)$ by

$$\theta = -\beta\, w, \qquad \phi(y) = y, \quad (11)$$

to obtain the distribution

$$p(y|\theta) = \exp\big(\langle\theta,y\rangle - \log\big(Z(\theta)\big)\big), \quad (12a)$$

$$Z(\theta) = \sum_{y\in\mathcal{Y}(G)} \exp\left(\langle\theta,y\rangle\right). \quad (12b)$$

Note that the dimension $d = |E|$ of the vectors w,y is considerably smaller than in problem (10).

Applying basic results that hold for distributions of the exponential family [25], the following holds regarding (12). For the random vector $Y = (Y_e)_{e\in E}$ taking values in $\mathcal{Y}(G)$, the marginal distributions, also called *mean parameters* in a more general context, are defined by

$$\mu_e := \mathbb{E}[\phi_e(Y)] = \sum_{y\in\mathcal{Y}(G)} \phi_e(y)p(y|\theta), \qquad \forall e \in E. \quad (13)$$

Likewise, the entire vector $\mu \in \mathbb{R}^{|E|}$ results as convex combination of the vectors $\phi(y)$, $y \in \mathcal{Y}(G)$. The closure of the convex hull of all such vectors corresponds to the (closure) of vectors μ that can be generated by valid distributions. This results in the representation of the multicut polytope (4)

$$\mathcal{MC}(G) = \operatorname{conv}\{\phi(y)\colon y \in \mathcal{Y}(G)\} \quad (14a)$$

$$= \Big\{\mu \in \mathbb{R}^{|E|}\colon \mu = \sum_{y\in\mathcal{Y}(G)} p(y)\phi(y) \text{ for some } p(y) \geq 0,\ \sum_{y\in\mathcal{Y}(G)} p(y) = 1\Big\}. \quad (14b)$$

Furthermore, the log-partition function generates the mean parameters through

$$\mu = \nabla_\theta \log Z(\theta), \tag{15}$$

which a short computation using (12) shows. Due to this relation, approximate probabilistic inference rests upon approximations of the log-partition function. In connection with discrete models, the Bethe-Kikuchi approximation and the local polytope relaxation provide basic examples for the marginal polytope [25].

In connection with the multicut polytope (14), however, we are not aware of an established outer relaxation and approximation of the log-partition function that is both tight enough and of manageable polynomial size. It is this fact that makes our approach presented in the subsequent section an attractive alternative, because it rests upon progress on solving several times problem (6) instead, together with perturbing the objective function.

4 Perturbation & MAP for Correlation Clustering

Recently, Hazan and Jaakkola [10] showed the connection between extreme value statistics and the partition function, based on the pioneering work of Gumbel [9]. In particular they provided a framework for approximating and bounding the partition function using MAP-inference with randomly perturbed models.

Analytic expressions for the statistics of a random MAP perturbation can be derived for general discrete sets, whenever independent and identically distributed random perturbations are applied to every assignment.

Theorem 1 ([9]). *Given a discrete Gibbs distribution $p(x) = 1/Z(\theta)\, \exp(\theta(x))$ with $x \in \mathcal{X}$ and $\theta : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$, let Γ be a vector of i.i.d. random variables Γ_x indexed by $x \in \mathcal{X}$, each following the Gumbel distribution whose cumulative distribution function is $F(t) = \exp\big(-\exp(-(t+c))\big)$ (here c is the Euler-Mascheroni constant). Then*

$$\Pr\big[\hat{x} = \arg\max_{x\in\mathcal{X}} \{\theta(x) + \Gamma_x\}\big] = 1/Z(\theta) \cdot \exp\big(\theta(\hat{x})\big), \tag{16a}$$

$$\mathbb{E}\big[\max_{x\in\mathcal{X}} \{\theta(x) + \Gamma_x\}\big] = \log Z. \tag{16b}$$

For our problem at hand the set $\mathcal{X} = \mathcal{Y}(G)$ is complex and thus Thm. 1 not directly applicable. Hazan and Jaakkola [10] develop computationally feasible approximations and bounds of the partition function based on *low*-dimensional random MAP perturbations.

Theorem 2 ([10]). *Given a discrete Gibbs distribution $p(x) = 1/Z(\theta)\, \exp(\theta(x))$ with $x \in \mathcal{X} = [L]^n$, $n = |V|$ and $\theta : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$. Let Γ' be a collection of i.i.d. random variables $\{\Gamma'_{i;x_i}\}$ indexed by $i \in V = [n]$ and $x_i \in X_i = [L]$, $i \in V$, each following the Gumbel distribution whose cumulative distribution function is $F(t) = \exp\big(-\exp(-(t+c))\big)$ (here c is the Euler-Mascheroni constant). Then*

$$\log Z(\theta) = \mathbb{E}_{\Gamma'_{1;x_1}}\Big[\max_{x_1\in\mathcal{X}_1} \cdots \mathbb{E}_{\Gamma'_{N;x_n}}\big[\max_{x_n\in\mathcal{X}_n} \theta(x) + \sum_{i\in V} \Gamma'_{i;x_i}\big] \dots\Big]. \tag{17}$$

Note that the random vector Γ' includes only nL random variables. Appying Jensen's inequality, we arrive at a computationally feasible upper bound of the log partition function,

$$\log Z(\theta) \leq \mathbb{E}_{\Gamma'}\Big[\max_{x \in X} \; \theta(x) + \sum_{i \in V} \Gamma'_{i;x_i}\Big]. \tag{18}$$

In the case of graph partitioning, we specifically have

$$\theta(y) = \begin{cases} \langle \theta, y \rangle & \text{if } y \in \mathcal{Y}(G) \\ -\infty & \text{else} \end{cases}, \qquad y \in \{0,1\}^{|E|} \tag{19}$$

with $\theta = -\beta\, w$ due to (11) which after insertion into Eq. (18) yields

$$\log Z(\theta) \leq \mathbb{E}_{\Gamma'}\Big[\max_{y \in \mathcal{Y}(G)} \langle \theta, y \rangle + \sum_{e \in E} \Gamma'_{e;y_e}\Big] =: \tilde{A}(\theta). \tag{20}$$

Our final step towards estimating the marginals (13) consists in replacing the log-partition function in (15) by the approximation (20) and computing estimates for the mean parameters

$$\mu \approx \tilde{\mu} := \nabla_\theta \tilde{A}(\theta) := \mathbb{E}_{\Gamma'}\Big[\operatorname*{arg\,max}_{y \in \mathcal{Y}(G)} \Big\{ \langle \theta, y \rangle + \sum_{e \in E} \Gamma'_{e;y_e} \Big\}\Big] \tag{21a}$$

$$\approx \frac{1}{M} \sum_{k=1}^{M} \operatorname*{arg\,max}_{y \in \mathcal{Y}(G)} \Big\{ \langle \theta, y \rangle + \sum_{e \in E} \gamma'^{(n)}_{e;y_e} \Big\}, \quad \gamma'^{(n)}_{e;y_e} \sim \Gamma'_{e;y_e}. \tag{21b}$$

Note that the expression in the brackets $[\dots]$ is a subgradient of the corresponding objective function. Thus, in words, we *define* our mean parameter estimate as empirical average of specific subgradients of the randomly perturbed MAP objective function.

5 Experiments

5.1 Setup

For the empirical evaluation of our approach we consider standard benchmark datasets for correlation clustering [14]. As solver for the correlation clustering problems we use the cutting-plane solver suggested by Kappes et al. [16], which can solve these problems to global optimality. We use the publicly available implementation of OpenGM2 [3].

For each instance we compare the globally optimal solution (mode)

$$\mu^* = \arg\max_{y \in \mathcal{Y}(G)} \sum_e w_e \cdot y_e \tag{22}$$

and the *local* boundary probabilities $\bar{\mu}$ given as softmax-function of the edge-weight

$$\bar{\mu}_e = \Pr_{\beta}^{\text{local}}(y_e = 1) := \frac{\exp(-\beta \cdot w_e)}{\exp(-\beta \cdot w_e) + 1} \tag{23}$$

with our estimates (21) for the boundary marginals based on the global model

$$\tilde{\mu}_e \overset{\text{Eq.(21)}}{\approx} \Pr_{\beta}(y_e = 1) := \sum_{y' \in \mathcal{Y}(G), y'_e = y_e} \frac{1}{Z(w, \beta)} \exp\Big(-\beta \cdot \sum_e w_e \cdot y'_e \Big) \tag{24}$$

for the same β as in eq. 23 and $M = 100$ samples for eq. 21. While μ^* and $\tilde{\mu}$ are by definition contained in the multicut polytope $\mathcal{MC}(G)$ and hence valid mean parameters, for $\bar{\mu}$ this is not necessarily the case, as the experiments will clearly show. For visualization we use the color map shown in Fig. 2.

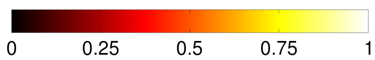


Fig. 2. Color coding used for visualization of boundary probabilities.

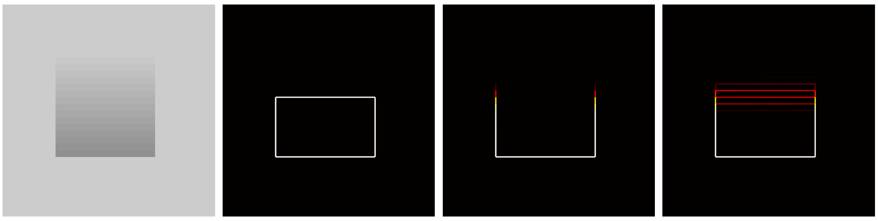
5.2 Evaluation and Discussion

Synthetic Example. We considered the image shown in Fig. 3(a). Local boundary detection was simply estimated by gray-value difference, i.e.

$$w_{ij} = |I(i) - I(j)| - 0.1. \qquad\qquad \forall ij \in E.$$

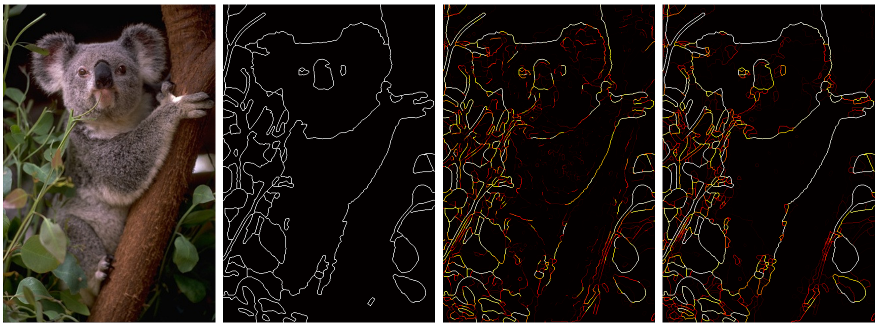
As shown in Fig. 3(c) this gives a strong boundary prediction in the lower part, but obviously no response in the upper part of the image. Applying correlation clustering to find the most likely clustering returns the partition shown in Fig. 3(b). However, this gives no information on the uncertainty of the solution. Fig. 3(d) shows our estimated mean parameters. These not only encode uncertainty but also enforce the boundary probability to be topologically consistent in terms of a convex combination of *valid* partitions.

Image Segmentation. For real world examples we use the public available benchmark model of Andres et al. [4, 14]. This model is based on super pixels and local boundary probabilities are learned by a random forest. Fig. 4 shows as example one of the 100 instances. Contrary to the mode (Fig. 4(b)), the boundary marginals (Fig. 4(d)) describe the uncertainty of the boundary and alternative contours. In contrast to the local boundary probability learned by a random forest, shown in Fig. 4(c), our marginal contours are closed and have no dangling contour-parts. This leads to a better boundary of the brown tree and removes or closes local artefacts in the koalas head. Note that Fig. 4(c) cannot be described as a convex combinations of valid clusterings.



(a) Image (b) μ^* (c) $\bar{\mu}$ (d) $\tilde{\mu}$

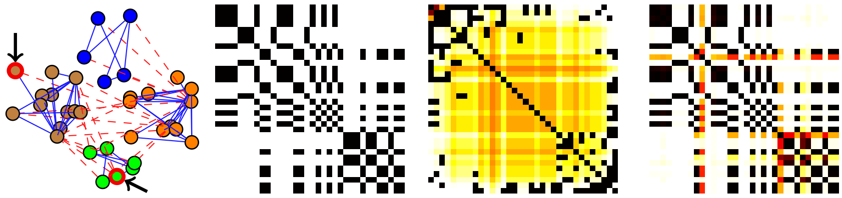
Fig. 3. The optimal clustering **(b)** encodes no uncertainty, the local probability **(c)** is topological not consistent. Our estimate **(d)** encodes uncertainty and is topological consistent.



(a) Image (b) μ^* (c) $\bar{\mu}$ (d) $\tilde{\mu}$

Fig. 4. The proposed global boundary probability **(d)** can only guarantee topological consistency and reflect uncertainty. This leads to a better boundary probabilities of the brown tree and removes or closes local artefacts in the koalas head compared to **(c)**. The optimal partitioning **(b)** and the local boundary probabilities **(c)** can handle only either aspect and, in the latter case, signal *invalid* partitions.

Social Networks. As an example for data mining and to demonstrate the generality of our approach, we consider the karate network [13]. Nodes in the graph correspond to members of the karate club and edges indicate friendship of members. The task is to cluster the graph such that the modularity is maximized, which can be reformulated into a correlation clustering problem over a fully connected graph with the same nodes. Because edge weights are not probabilistically motivated for this model, the local boundary probabilities are poor (Fig. 5(c)). Global inference helps to detect the two members (nodes) for which the assignment to the cluster is uncertain (Fig. 5(d)). Fig. 5(a) shows the clustering that maximizes the modularity. Our result enables the conclusion that the two uncertain nodes (marked with red boundary and arrows) can be moved to another cluster without much worsening the modularity.



(a) network (b) μ^* (c) $\bar{\mu}$ (d) $\tilde{\mu}$

Fig. 5. The clustering of members of a karate club is a example for correlation clustering in social networks. Figure **(a)** and **(b)** show the clustering that maximizes the modularity. Nodes marked with a red boundary in **(a)** are nodes with an uncertain assignment. The uncertainty is measured by the marginal probabilities **(d)**. Pseudo probabilities calculated by local weights only, shown in **(c)**, do not reveal this detailed information. Our result **(d)** enables to conclude that for the network graph **(a)** the modularity would not change much if the two nodes with uncertain assignment would be moved to the orange and brown cluster, respectively.

6 Conclusion

We presented a probabilistic approach to correlation clustering and showed how perturbed MAP estimates can be used to efficiently calculate globally consistent approximations to marginal distributions. Regarding image partitioning, by enforcing this marginal consistency, we are able to close open contour parts caused by imperfect local detection and thus reduce local artefacts by topological priors. In future work we would like to speed up our method by making use of warm start techniques, to reduce the computation time from a few minutes to seconds.

Acknowledgments. We thank Johannes Berger for inspiring discussions. This work has been supported by the German Research Foundation (DFG) within the program Spatio/Temporal Graphical Models and Applications in Image Analysis", grant GRK 1653.

References

1. M. Aigner. *Combinatorial Theory.* Springer, 1997.
2. L. Ambrosio and V. Tortorelli. Approximation of Functionals Depending on Jumps by Elliptic Functionals via γ-Convergence. *Comm. Pure Appl. Math.*, 43(8):999–1036, 1990.
3. B. Andres, T. Beier, and J. H. Kappes. OpenGM: A C++ library for Discrete Graphical Models. *CoRR*, abs/1206.0111, 2012.
4. B. Andres, J. H. Kappes, T. Beier, U. Köthe, and F. A. Hamprecht. Probabilistic image segmentation with closedness constraints. In *ICCV*, pages 2611–2618. IEEE, 2011.
5. N. Bansal, A. Blum, and S. Chawla. Correlation clustering. *Machine Learning*, 56(1–3):89–113, 2004.
6. A. Blake and A. Zisserman. *Visual Reconstruction.* MIT Press, 1987.

7. A. Chambolle, D. Cremers, and T. Pock. A Convex Approach to Minimal Partitions. *SIAM J. Imag. Sci.*, 5(4):1113–1158, 2012.
8. S. Chopra and M. Rao. The partition problem. *Mathematical Programming*, 59(1-3):87–115, 1993.
9. E. Gumbel. *Statistical theory of extreme values and some practical applications: a series of lectures*. Applied mathematics series. U. S. Govt. Print. Office, 1954.
10. T. Hazan and T. Jaakkola. On the partition function and random maximum a-posteriori perturbations. In *ICML*. icml.cc / Omnipress, 2012.
11. T. Hofman and J. Buhmann. Pairwise Data Clustering by Deterministic Annealing. *IEEE Trans. Patt. Anal. Mach. Intell.*, 19(1):1–14, 1997.
12. R. Kannan, S. Vempala, and A. Vetta. On clusterings: good, bad and spectral. *J. ACM*, 51(3):497–515, 2004.
13. J. Kappes, B. Andres, F. Hamprecht, C. Schnörr, S. Nowozin, D. Batra, S. Kim, B. 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. 2014. http://arxiv.org/abs/1404.0533.
14. J. H. Kappes, B. Andres, F. A. Hamprecht, C. Schnörr, S. Nowozin, D. Batra, S. Kim, B. X. Kausler, J. Lellmann, N. Komodakis, and C. Rother. A comparative study of modern inference techniques for discrete energy minimization problems. In *CVPR*, 2013.
15. J. H. Kappes, M. Speth, B. Andres, G. Reinelt, and C. Schnörr. Globally optimal image partitioning by multicuts. In *EMMCVPR*, pages 31–44. Springer, 2011.
16. J. H. Kappes, M. Speth, G. Reinelt, and C. Schnörr. Higher-order segmentation via multicuts. *CoRR*, abs/1305.6387, 2013.
17. J. Lellmann and C. Schnörr. Continuous Multiclass Labeling Approaches and Algorithms. *SIAM J. Imaging Science*, 4(4):1049–1096, 2011.
18. P. Orbanz and J. Buhmann. Nonparametric Bayesian Image Segmentation. *Int. J. Comp. Vision*, 77(1-3):25–45, 2008.
19. G. Papandreou and A. Yuille. Perturb-and-MAP Random Fields: Using Discrete Optimization to Learn and Sample from Energy Models. In *Proc. ICCV*, 2011.
20. T. Pätz, R. Kirby, and T. Preusser. Ambrosio-Tortorelli Segmentation of Stochastic Images: Model Extensions, Theoretical Investigations and Numerical Methods. *Int. J. Comp. Vision*, 103(2):190–212, 2013.
21. N. Rebagliati, S. Rota Bulo, and M. Pelillo. Correlation clustering with stochastic labellings. In *Similarity-Based Pattern Recognition*, volume 7953 of *LNCS*, pages 120–133. Springer Berlin Heidelberg, 2013.
22. J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, aug 2000.
23. U. von Luxburg. A Tutorial on Spectral Clustering. *Statistics and Computing*, 17(4):395–416, 2007.
24. U. von Luxburg. Clustering Stability: An Overview. *Found. Trends Mach. Learning*, 2(3):235–274, 2009.
25. M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1:1–305, 2008.