UNIFORMLY REWEIGHTED BELIEF PROPAGATION FOR DISTRIBUTED BAYESIAN HYPOTHESIS TESTING

Federico Penna[1]*, Henk Wymeersch*[2]*, Vladimir Savic*[3]

[1]Politecnico di Torino, Italy. Email: federico.penna@polito.it
[2]Chalmers University of Technology, Gothenburg, Sweden. Email: henkw@chalmers.se
[3]Universidad Politecnica de Madrid, Spain. Email: vladimir@gaps.ssr.upm.es

ABSTRACT

Belief propagation (BP) is a technique for distributed inference in wireless networks and is often used even when the underlying graphical model contains cycles. In this paper, we propose a uniformly reweighted BP scheme that reduces the impact of cycles by weighting messages by a constant "edge appearance probability" $\rho \leq 1$. We apply this algorithm to distributed binary hypothesis testing problems (e.g., distributed detection) in wireless networks with Markov random field models. We demonstrate that in the considered setting the proposed method outperforms standard BP, while maintaining similar complexity. We then show that the optimal ρ can be approximated as a simple function of the average node degree, and can hence be computed in a distributed fashion through a consensus algorithm.

1. INTRODUCTION

Many problems in wireless networks, such as classification or detection, can be formulated as distributed hypothesis tests, where multiple nodes have to choose among a set of possible alternatives based on certain observable data. Applications include sensor networks [1], surveillance systems [2], spectrum sensing in cognitive radio networks [3].

Such hypothesis testing problems can be addressed by a Bayesian inference approach and thus mapped on probabilistic graphical models that help devise distributed solutions. We focus here on the case of "heterogeneous" hypotheses, i.e., when to each node corresponds a different state variable to be estimated.[1] Markov random fields (MRF) are a typical graphical model used to represent the structure of this class of problems, establishing a one-to-one connection between nodes and variables, and accounting for (pairwise) correlations between neighboring nodes. Once the communication graph is mapped onto a statistical graph, distributed inference can be performed. The usual tool adopted for distributed inference on MRF models is belief propagation (BP) [4], in its sum-product or max-product variants. It is well known that if the MRF contain cycles, BP does not converge to the exact solution, but often yields reasonable approximations. Algorithms for exact inference on loopy graphs (e.g., generalized BP [5]) are much more complex than standard BP and not suitable for a distributed implementation.

This research is supported, in part, by the FPU fellowship from Spanish Ministry of Science and Innovation, ICT project FP7-ICT-2009-4-248894-WHERE-2, and program CONSOLIDER-INGENIO 2010 under grant CSD2008-00010 COMONSENS.

[1]Other models assume, on the contrary, the same underlying hypothesis for all network nodes. These are referred to as "consensus problems".

In this paper, we address the problem of distributed hypothesis testing in networks with loops by applying a simplified version of the tree-reweighted (TRW)-BP algorithm, introduced by Wainwright *et al.* in [6, 7]. We show that the proposed algorithm provides an improved approximation of marginal a posteriori probabilities compared to loopy belief propagation (LBP), while maintaining (in the proposed simplified version) essentially the same complexity. The same approach can be adopted to address problems involving continuous domain variables (e.g., cooperative localization: see [10]).

The paper is organized as follows: the mathematical model is defined in Sec. 2; in Sec. 3 we introduce BP and its variations, and we discuss the problem of optimizing the edge appearance probability; Sec. 4 contains simulation results to validate the proposed BP algorithm; Sec. 5 concludes.

2. MATHEMATICAL MODEL

Consider a wireless network composed of K nodes. Each node is characterized by a state h_i, taking values in some discrete set of possible events, and collects some observation, expressed in general by vector $\boldsymbol{y}_i$ which depends on the underlying state through a likelihood function

$$\varphi_i(h_i) = p(\boldsymbol{y}_i|h_i). \tag{1}$$

We denote by $\boldsymbol{h} \triangleq [h_1, \ldots, h_K]$ and by $\boldsymbol{Y} \triangleq [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_K]$ the set of all nodes' states and observations, respectively. For simplicity, we focus here on the case of binary hypothesis, i.e., $h_i \in \{0, 1\}$, which is of particular interest for the problem of signal detection. We also assume that each node knows its own observation likelihood functions for both states, $\varphi_i(0)$ and $\varphi_i(1)$ (in case of detection problems, e.g., [3], these functions depend on the relative powers of the noise and of the signal under test). Inter-dependencies among nodes in the network are modeled by a pairwise MRF, i.e., the state of each node depends on the states of its neighbors (i.e., devices within communication range) through pairwise correlation terms. As such, the joint a priori distribution of vector $\boldsymbol{h}$ is given by

$$p(\boldsymbol{h}) = \prod_{i=1}^{K} \prod_{j \in \mathcal{N}_i, j<i} \psi_{ij}(h_i, h_j), \tag{2}$$

where $\mathcal{N}_i$ is the set of neighbors of i, and condition $j < i$ simply avoids double-counting the same term. Functions ψ_{ij} are specific to the problem of interest and may depend on the distance between node i and j. In Sec. 4 we consider exponential MRFs, which are one of the most typical and widely adopted MRF models. Assuming that communication and statistical graph can be mapped to each other, a natural graphical representation of the above MRF models is an undirected graph $G = (V, E)$ where vertices represent network

nodes and each pair of neighboring nodes (i,j) is linked by an edge with weight given by ψ_{ij}.

Solving the multiple hypothesis testing problem means estimating, in a distributed way, the marginal *a posteriori probabilities* (APP) of variables h_i,

$$p(h_i|\boldsymbol{Y}) = \sum_{\boldsymbol{h}\setminus h_i} p(\boldsymbol{h}|\boldsymbol{Y}). \tag{3}$$

Based on Bayes rule, the joint APP is

$$p(\boldsymbol{h}|\boldsymbol{Y}) \propto p(\boldsymbol{Y}|\boldsymbol{h})p(\boldsymbol{h}) = \prod_{i=1}^{K}\left[\varphi_i(h_i)\prod_{j\in\mathcal{N}_i, j<i}\psi_{ij}(h_i,h_j)\right], \tag{4}$$

where conditional mutual independence of different nodes' observations is assumed. In the next section we introduce BP and modified BP algorithms to approximate the marginals (3).

3. BP AND REWEIGHTED BP

3.1. Loopy BP

According to traditional BP [4], each node iteratively exchanges with its neighbors messages of the form

$$\mu_{i\to j}(h_j) \propto \sum_{h_i}\left(\varphi_i(h_i)\psi_{ij}(h_i,h_j)\prod_{n\in\mathcal{N}_i\setminus j}\mu_{n\to i}(h_i)\right), \tag{5}$$

with initialization $\mu_{i\to j}=1\ \forall(i,j)$. Beliefs are updated as

$$b_i(h_i) \propto \varphi_i(h_i)\prod_{n\in\mathcal{N}_i}\mu_{n\to i}(h_i), \tag{6}$$

normalized such that $b_i(0)+b_i(1)=1$. If G is a tree, after a sufficient number of iterations belief b_i converges to the corresponding marginal APP (3). However, when G has cycles, BP may provide poor performance or even fail to converge [8].

3.2. Tree-reweighted BP

Tree-reweighted BP is a generalization of BP introduced in [6, 7]. While ordinary BP corresponds to finding a stationary point in the variational problem associated to Bethe's free energy approximation, TRW-BP is build on an improved upper bound of the log-partition function consisting of a convex combination of spanning trees[2]. From this idea, a local message passing algorithm analogous to BP is derived. Still it does not provably converge to the exact marginal APP, but in certain cases it provides a much better approximation than ordinary BP. The TRW-BP algorithm is defined by the following update rules:

$$\mu_{i\to j}(h_j) \propto \sum_{h_i}\left(\varphi_i(h_i)\psi_{ij}^{1/\rho_{ij}}(h_i,h_j)\frac{\prod_{n\in\mathcal{N}_i\setminus j}\mu_{n\to i}^{\rho_{in}}(h_i)}{\mu_{j\to i}^{1-\rho_{ij}}(h_i)}\right) \tag{7}$$

$$b_i(h_i) \propto \varphi_i(h_i)\prod_{n\in\mathcal{N}_i}\mu_{n\to i}^{\rho_{in}}(h_i), \tag{8}$$

where coefficients ρ_{ij} are called *edge appearance probabilities*. The vector of all edge appearance probabilities is denoted by $\boldsymbol{\rho}$ and has length $|E|$. According to [7], valid choices of $\boldsymbol{\rho}$ must belong to the *spanning tree polytope*: given a distribution $p(T)$ over the possible spanning trees $\mathbb{T}(G)$ of G, ρ_{ij} is given by

$$\rho_{ij} = \sum_{T\in\mathbb{T}(G)} p(T)n_T(i,j), \tag{9}$$

where $n_T(i,j)$ is 1 if edge $(i,j)\in T$, 0 otherwise.

Notice that configuration $\boldsymbol{\rho}=\mathbf{1}$ amounts to ordinary BP, and based on the above condition is valid only if G is itself a tree. In general, convexity properties of the TRW-BP formulation guarantee that an optimal choice of $\boldsymbol{\rho}$ (that minimizes the tree-based upper bound of the log-partition function) always exists, and can be found by solving a convex optimization problem over $\mathbb{T}(G)$, e.g., using the gradient descent algorithm proposed in [7].

Unfortunately, a direct application of TRW-BP to our distributed problem is not feasible, as it involves computation of all possible spanning trees, and iterative optimization to find the best $\boldsymbol{\rho}$. Implementing these tasks in a distributed fashion would be prohibitive due to the huge amount of information to be passed throughout the network.

3.3. Uniformly-reweighted BP

We propose a simplified version of reweighted BP, which we call uniformly-reweighted (URW)-BP. It has the same structure as TRW-BP, but we assign a constant appearance probability to all edges:

$$\rho_{ij}=\rho \quad \forall(i,j)\in E, \quad \text{with } 0<\rho\leq 1. \tag{10}$$

In doing so, we relax the tree-consistency requirement and reduce the degrees of freedom from $|E|$ to 1. Yet, this simplified reweighting scheme turns out to outperform BP in graphs with cycles. Notice that in graphs satisfying certain symmetry conditions (e.g., [7], example 3), uniform edge appearance probabilities are an optimal choice.

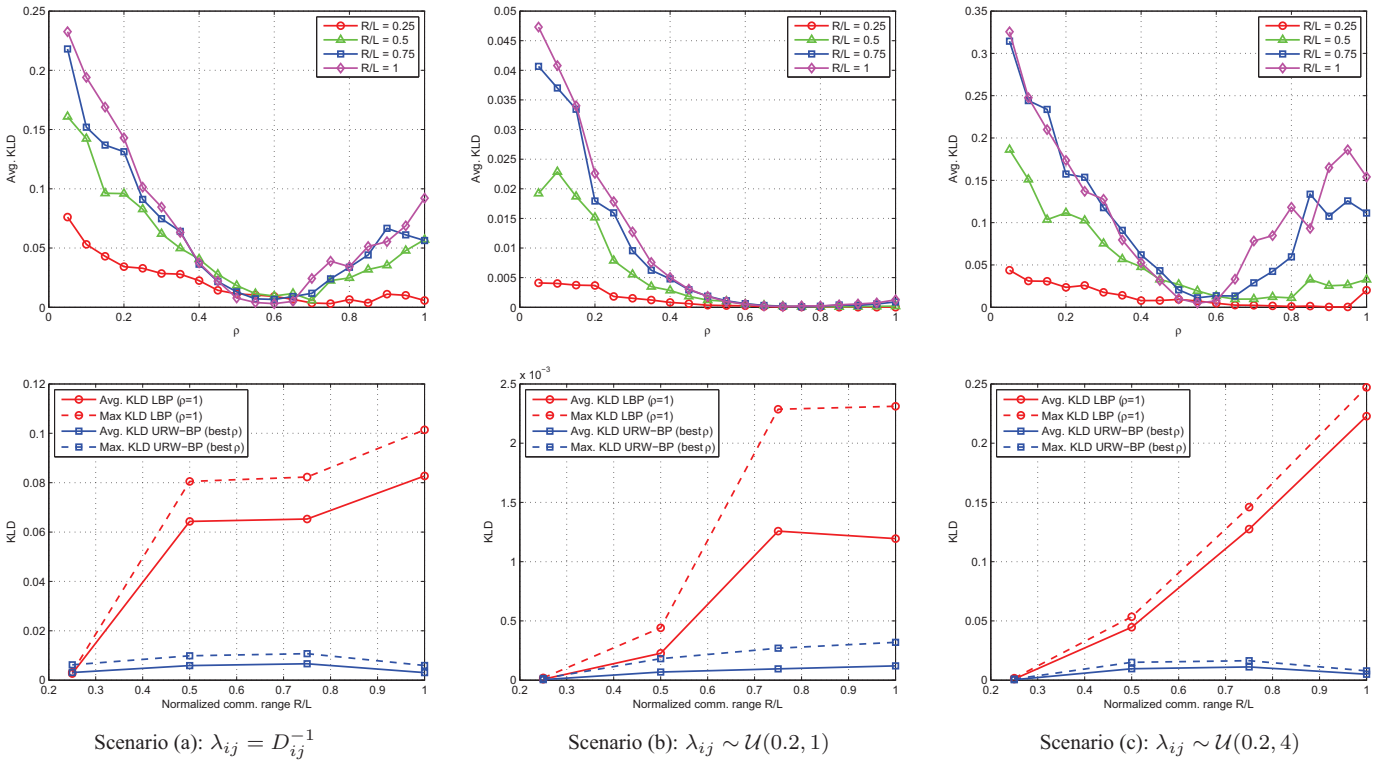
3.4. Optimizing the Edge Appearance Probability

The main question for applying URW-BP in practice is how to set ρ. Since $\rho=1$ corresponds to standard BP, intuitively we expect that if the network has a low degree of connectivity (hence few loops) the optimal value of ρ will be around 1; on the other hand, we expect lower values of ρ to perform better as connectivity increases.

To give insight into the dynamics of the algorithm, let us inspect the message update rule (7). Denote by d_k the degree of vertex k, i.e., the number of nodes connected to vertex k. Then, a generic message $\mu_{i\to j}$ includes d_i-1 messages from nodes at 1-hop distance, weighted by ρ. Each of them (say $\mu_{n\to i}$) in turn includes d_n-1 messages coming from nodes at 2-hop distance from i, resulting in weight ρ^2, and so on. Therefore, if $\rho<1$, the algorithm tends to reduce the weight of messages coming from nodes that are not in direct proximity, which is beneficial because (due to loops) these nodes might have been already reached through different paths. When the average degree $\overline{d}$ increases, incoming messages from nodes at a given distance are more and more likely to be double-counted. For this reason, we expect the optimal value of ρ to decrease with the average node degree.

Some more precise results can be given in case of symmetric graphs (i.e., where functions φ_i, ψ_{ij} exhibit symmetry). In this case, as mentioned in Sec. 3.3 a uniform edge appearance probability is optimal in the sense of TRW-BP, and the best ρ (denoted as ρ^*) can be approximated as $\rho^*\approx\frac{|V|-1}{|E|}$ (see [7, Sec. V-D]). This

[2]For a definition of Bethe's free energy, log partition-function, and spanning tree, we refer the reader to [7].



Scenario (a): $\lambda_{ij} = D_{ij}^{-1}$ Scenario (b): $\lambda_{ij} \sim \mathcal{U}(0.2, 1)$ Scenario (c): $\lambda_{ij} \sim \mathcal{U}(0.2, 4)$

Fig. 1. First row: Average KLD vs. ρ for a simple network of $K = 4$ nodes, different communication ranges (R/L) from 0.25 to 1. Second row: Avg./Max. KLD vs. connectivity range (R/L) using LBP and URW-BP with $\rho = \rho^*$.

can be related to the average node degree $\overline{d}$ as follows. For a general graph with $|V|$ vertices and $|E|$ edges, every edge must connect two vertices, so that a simple counting argument yields $\overline{d} = 2|E|/|V|$. Substitution gives

$$\rho^* \approx 2/\overline{d}. \tag{11}$$

As shown in Sec. 4.3, the above choice of ρ turns out to be very accurate in the considered MRF models, in spite of the fact that symmetry conditions are not strictly satisfied. For this reason, (11) can be used to set ρ when applying URW-BP in practice to address distributed problems in wireless networks. Note that the value of $\overline{d}$, and therefore of ρ^*, can be computed by running a simple average consensus algorithm [9] over the network.

4. CASE STUDY

4.1. Scenarios and Evaluation Metrics

As a reference scenario, we consider K nodes randomly deployed in a circular region with diameter L, with random observation likelihood functions $\varphi_i(0) \sim \mathcal{U}(0, 1)$, $\varphi_i(1) = 1 - \varphi_i(0)$, and pairwise interactions modeled as an exponential MRF:

$$\psi_{ij}(h_i, h_j) \propto e^{\lambda_{ij}\delta(h_i,h_j)} \tag{12}$$

where $\delta(h_i, h_j) = 1$ if $h_i = h_j$ and 0 otherwise. This model, that is a Gibbs distribution which can be derived from the maximum entropy principle, captures the structure of many practical problems where neighboring nodes are likely to have the same underlying state (e.g., [3]). The correlation strength is given by factors λ_{ij}. For generality, we consider two possible models:

(i) *Distance-based model*: $\lambda_{ij} = D_{ij}^{-\gamma}$, where D_{ij} is the distance (normalized by $L/2$) between nodes i and j, and γ is a decay exponent;

(ii) *Random correlation model*: $\lambda_{ij} \sim \mathcal{U}(\lambda_{\min}, \lambda_{\max})$.

In both cases, we assume that $\lambda_{ij} = 0$ when nodes i and j are at a distance greater than R (communication range).

Denoting by b_i the belief of node i computed through LBP or URW-BP after a number of iterations sufficient to reach convergence, we use as a performance evaluation metric the Kullback-Leibler divergence (KLD) between true APP and belief, defined as

$$\mathrm{KLD}_i = \sum_{h_i=\{0,1\}} b_i(h_i) \log \frac{b_i(h_i)}{p(h_i|\boldsymbol{Y})}. \tag{13}$$

4.2. Results for $K = 4$

Algorithms are first evaluated in a small network of 4 nodes, with four possible levels of connectivity, i.e., $R/L = \{0.25, 0.5, 0.75, 1\}$. For each value of connectivity a simulation set of 100 Monte Carlo runs is carried out. At every run a new topology is generated, with random positions of the nodes and different coefficients λ_{ij}, drawn (a) according to a distance-based model, with $\gamma = 1$, or (b) according to a random model with $\lambda_{\min} = 0.2, \lambda_{\max} = 1$, or (c) with $\lambda_{\min} = 0.2, \lambda_{\max} = 4$. For all BP methods, message passing is stopped after 6 iterations, which are enough to reach convergence[3] of all beliefs in a network of 4 nodes. For each of the above cases,

[3] Cases of non-convergence were never encountered in simulations.

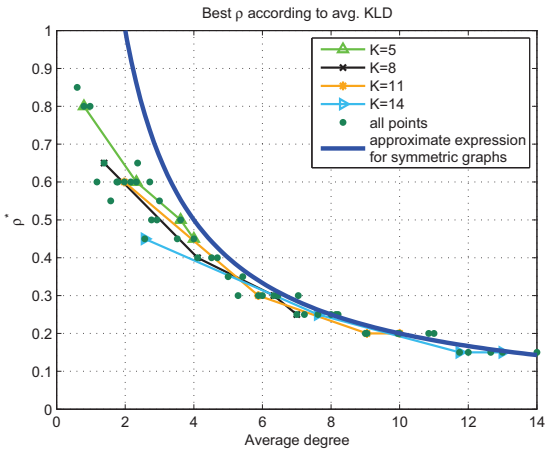


Fig. 2. URW-BP: optimal edge appearance probability (ρ^*) vs. average degree ($\overline{d}$), based on avg. KLD. Data from multiple simulations vs. theoretical approximation of ρ^* for symmetric graphs. Scenario (a).

average and maximum KLD over K nodes are computed for LBP and URW-BP for $\rho \in (0, 1]$ (with a step of 0.05).

Fig. 1 shows that for low values of connectivity range (e.g., $R/L = 0.25$), the best ρ remains close to 1, which indicates that reweighted BP does not provide significant improvement over LBP. When the connectivity range increases, i.e., more loops appear, we observe that: (i) ρ^* progressively decreases; (ii) the improvement brought by URW-BP(ρ^*) over LBP increases (see Fig. 1, second row); (iii) there is a wide range of values of ρ such that URW-BP(ρ) outperforms LBP.

Comparing results obtained under different correlation models, namely (a), (b) and (c), we notice that when the values of λ are very low (e.g., case b), the curve of KLD vs. ρ appears flat, with values of KLD close to 0 for all $\rho > 0.5$. In this case, in fact, the impact of loops is negligible, therefore LBP already provides good performance and no further optimization is really needed. On the other hand, especially from comparison of scenario (a) and (c), values of ρ^* vs. R do not change significantly for different correlation models. This fact suggests that it is possible to infer ρ^* just from the network topology or more precisely, as we will see in next section, from the average degree ($\overline{d}$).

4.3. Results for Larger Networks and Optimization of ρ

Results of extensive simulations performed in networks with larger numbers of nodes turn out to be similar to those observed in the example of $K = 4$. The curve of KLD vs ρ is in all cases exhibits a unique minimum in $(0, 1)$, with values moving from 1 towards 0 as the communication range increases. In addition, simulation results confirm the intuition that ρ^* is determined essentially by the average number of neighbors of each node, i.e., by the average degree $\overline{d}$, rather than by the total number of nodes K, or the connectivity range R, or the values of correlations λ_{ij}.

The plot in Fig. 2, for instance, is obtained by merging results from several simulations, considering 100 Monte Carlo runs for every value of K from 4 to 15, and normalized communication ranges from 0.25 to 1, all with correlation coefficients modeled according to scenario (a). Single curves ($K = 5, 8, 11, 14$) are also plotted as examples. With any pair $(K, R/L)$ corresponds a certain average degree, $\overline{d}$, and the value of ρ^* is then plotted as a function of $\overline{d}$. Simulation data are then compared to theoretical expression (11) found for symmetric graphs.

Results indicate that the above expression of ρ^* vs. $\overline{d}$ becomes increasingly accurate as $\overline{d} \to \infty$, and, in practice, it can be considered a good approximation for $\overline{d} > 3$. Note that a correct choice of ρ^* is needed especially when $\overline{d}$ is large, that is where the gap between URW-BP and LBP increases. In summary, URW-BP with ρ set according to (11) turns out to provide a significant performance improvement over LBP in all considered scenarios. Moreover, URW-BP avoids complex optimization procedures as in TRW-BP (Sec. 3.2) thus keeping complexity low, as in standard BP.

5. CONCLUSIONS

In this paper we have shown that a simple variation of BP, where all messages are weighted by a constant factor ρ, leads to substantial performance improvement in distributed inference problems in wireless networks. We studied in detail the case of binary hypothesis testing (applicable, for instance, to distributed detection problems) assuming an underlying Markov random field model.

The proposed method outperforms standard BP for a wide range of ρ, especially when variables' interactions are high. In addition, we showed that the optimal ρ can be well approximated by a simple function of the average node degree and computed in a distributed way by a consensus algorithm. Therefore, URW-BP does not result in a significant increase of complexity compared to traditional BP. This property makes it suitable for application in a variety of practical problems.

6. REFERENCES

[1] A. Dogandzic, B. Zhang, "Distributed Estimation and Detection for Sensor Networks Using Hidden Markov Random Field Models", *IEEE Trans. Sig. Proc.,* vol. 54, no. 8, Aug. 2006.

[2] B. S. Rao, H. Durrant-Whyte, "A Decentralized Bayesian Algorithm for Identification of Tracked Targets", *IEEE Trans. on Systems, Man, and Cybernetics,* vol. 23, no. 6, Nov./Dec. 1993.

[3] F. Penna, R. Garello, M. A. Spirito, "Distributed Inference of Channel Occupation Probabilities in Cognitive Networks via Message Passing", *IEEE DySPAN*, Singapore, Apr. 2010.

[4] J. Pearl, "Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference", Morgan Kaufman, San Mateo, CA, 1988.

[5] J. S. Yedidia, W. T. Freeman, Y. Weiss, "Generalized Belief Propagation", *Advances in Neural Information Processing Systems (NIPS)*, vol. 13, Dec. 2000.

[6] M. J. Wainwright, T. S. Jaakkola, A. S. Willsky, "Tree-reweighted belief propagation algorithms and approximate ML estimation by pseudo-moment matching", *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2003.

[7] M. J. Wainwright, T. S. Jaakkola, A. S. Willsky, "A New Class of Upper Bounds on the Log Partition Function", *IEEE Trans. on Information Theory,* vol. 51, no. 7, July 2005.

[8] K. Murphy, Y. Weiss, and M. Jordan, "Loopy belief propagation for approximate inference: an empirical study", *Conf. on Uncertainty in Artificial Intelligence*, 1999.

[9] R. Olfati-Saber, R.M. Murray, "Consensus problems in networks of agents with switching topology and time-delays", *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520-1533, Sept. 2004.

[10] V. Savic, H. Wymeersch, F. Penna, S. Zazo, "Optimized Edge Appearance Probability for Cooperative Localization based on Tree-Reweighted Nonparametric Belief Propagation", to be presented at *IEEE ICASSP*, May 2011.