Distributed Detection in Sensor Networks with Packet Losses and Finite Capacity Links *

Venkatesh Saligrama   Murat Alanyali   Onur Savas †

Abstract

We consider the problem of classifying among a set of M hypotheses via distributed noisy sensors. Sensor measurements are conditionally independent given the true hypothesis. The sensors can collaborate over a communication network and the task is to arrive at a consensus about the event after exchanging messages. We apply a variant of belief propagation as a strategy for collaboration to arrive at a solution to the distributed classification problem. We show that the message evolution can be re-formulated as the evolution of a linear dynamical system, which is primarily characterized by network connectivity. We show that a consensus to the centralized MAP estimate can almost always reached by the sensors for any arbitrary network. We then extend these results in several directions. First, we demonstrate that these results continue to hold with quantization of the messages, which is appealing from the point of view of finite bit rates supportable between links. We then demonstrate robustness against packet losses, which implies that optimal decisions can be achieved with asynchronous transmissions as well. Next, we present an account of energy requirements for distributed detection and demonstrate significant improvement over conventional decentralized detection. Finally, extensions to distributed estimation are described.

1 Introduction

Recent advances in computing and communication technologies provide impetus for deploying massive networks of tiny sensors capable of measuring, processing and exchanging data over a wireless medium. In typical

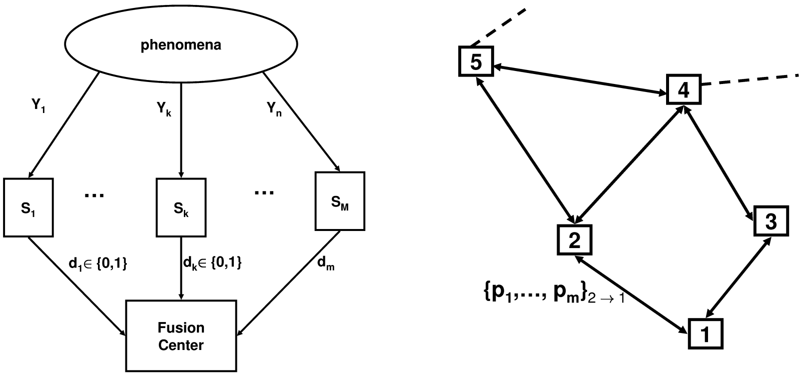
*This research was supported by ONR Young Investigator Award N00014-02-100362, PECASE Award and NSF CAREER Program under grants ANI-0238397,CCF-0430983, and CNS-0435353.

†The authors are affiliated with the Department of Electrical and Computer Engineering at Boston University. E-mail: {srv,alanyali,savas}@bu.edu.

applications energy limitation of individual sensors is a primary bottleneck as it entails further constraints in communication bandwidth, reliability and connectivity. Information processing models that account for such limitations have recently received much attention within the networking, signal processing and information-theory communities. In this paper we address this issue from a distributed viewpoint and consider a collection of sensors observing a single phenomena through noisy measurements. The sensors can only collaborate through a network defined by a connectivity graph in which messages may be subject to quantization or random losses. The task is to exchange messages in order to arrive at a consensus that reflects the classification of the event by a hypothetical node that has access to all observations and observation models.

The general question of dealing with distributed data in the context of detection has been an active topic of research (see [1, 2, 5, 9, 12–14, 17, 21–24] and references therein). Previously proposed techniques can be broadly categorized into two groups: The fusion-centric approach assumes that each sensor has a communication link to a data fusion center as shown in Fig. 1(a). Quantization of sensor data in this model was addressed by [13, 21], effects of power constraints on noisy communication channels were considered in [6, 7], and data compression issues were studied by [3] from an information-theoretic perspective. The ad-hoc approach, on the other hand, involves no designated fusion center but focuses on establishing consensus within the network via message exchanges. This approach is arguably more suitable to address energy issues in large-scale networks and also appears to have robustness advantages. Early work [4] establishes that consensus is achieved if messages are conditional expectations adopted to local measurements and messages, however the agreement itself is in general sensitive to the relative timing of messages, and computing the conditional expectations is not practically appealing. Message specifications and rigid messaging schedules that lead to consensus on optimal decisions were given in [20] for the special case of a completely connected communication topology.

We adopt the ad-hoc model and develop a data-oriented communication strategy in order to overcome the alluded issues. Related work for distributed optimization in sensor networks have been proposed recently in [10, 16] for specific types of network topologies. Here we pursue a more general objective of developing a truly ad-hoc, asynchronous, energy efficient detection theory for arbitrary network topologies. Our problem focuses on deriving conditions for arriving at a consensus at all the sensors and situations where the consensus is the centralized MAP estimate. A natural idea for collaboration is to exchange a vector of individual sensor beliefs (probabilities) for different hypothesis between linked sensors at any instant of time. This idea is formalized in the "so called" belief propagation (BP) algorithm [15] and preliminary results on their application to the detection problem is described in a number of our papers [1, 23, 24]. A description is shown in Figure 1(b) where sensor



(a) Decentralized Detection (b) Distributed Detection

Figure 1: Two distict schemes for detection in sensor networks.

nodes send a vector of likelihoods for each hypothesis at any instant of time. These likelihoods are dynamically updated based on information received by the sensor in the past. In this setup, we neither have a fusion center nor does each sensor need to know models for adjacent sensors. Nevertheless, BP is known to work generally for non-loopy network topologies. Furthermore, on account of finite link capacity, it is unclear as to how to deal with attendant effects of quantization. We deal with the first issue in Section 4 by first classifying loopy graph topologies for which the standard BP does converge to the MAP consensus. However, these turn out to be limited and motivates us to consider variants of BP algorithm and we show in Section 5 that for the class of problems where all sensors are engaged in the same classification task, consensus can indeed be attained through such modifications. We further show that if network topology is known then consensus on the exact posterior distribution can also be realized. We next deal with implications of finite link capacities in Section 6 by employing a novel robustness perspective. By showing that our algorithm is robust to perturbations of messages we are able to quantify explicitly the size of quantization before performance degrades. Robustness of the algorithm against random packet losses is established in Section 6.2. In Section 7, we compare the energy requirements of decentralized and distributed detections schemes on a uniform grid and show exponential improvement in energy scaling over the fusion-center approach. Finally, in Section 8 extensions to distributed estimation are described.

In summary the main advantages of the proposed scheme are as follows: **(A)** The sensor network can operate in a completely asynchronous fashion, i.e., the algorithm as well as the outcomes do not depend on when a message is transmitted. **(B)** Second, each sensor node in the networks does not have knowledge of sensing models for other sensors. This implies that the algorithm works irrespective of knowing "who is sending what." **(C)** The algorithm always converges to the optimal MAP estimate.

2 Setup

We consider MAP estimation in M-ary hypothesis testing problems with conditionally independent observations. The observation vector is denoted by $Y = (Y_v : v \in V)$, where throughout the paper V represents a set of sensors and Y_v represents the measurement taken by sensor $v \in V$. For definiteness we focus on the case when Y has a continuous distribution, but conclusions of the paper apply verbatim for discrete observations when densities are interpreted as probabilities.

Let $\{H_1, H_2, \cdots, H_M\}$ be a collection of M hypotheses with prior distribution π_o. The conditional probability density function of the observation vector Y under each hypothesis $H_i,\ i = 1, 2, \cdots, M$, is denoted by $f_i(\cdot)$. We shall assume that observations are conditionally independent given the true hypothesis. Specifically, for each realization $y = (y_v : v \in V)$ of observation vector Y

$$f_i(y) = \prod_{v \in V} f_{i,v}(y_v), \tag{1}$$

for marginal densities $f_{i,v}(\cdot)$. Let π denote the posterior distribution of the true hypothesis given the observations. Namely,

$$\pi(H_i) = K\pi_o(H_i) \prod_{v \in V} f_{i,v}(y_v), \qquad i = 1, 2, \cdots, M, \tag{2}$$

where K is a normalization constant that does not depend on i. Here both π and K depend on y but this dependence is suppressed in the notation for convenience. Note, for example, that an observation model where $Y_v = g_i(v, \eta_v)$ under hypothesis H_i conforms to the described setting if $g_i(\cdot) : i = 1, 2, \cdots, M$, are deterministic functions and $\eta_v : v \in V$ are conditionally independent random variables given the true hypothesis.

Hypothesis H_i is a MAP estimate if

$$i \in \arg\max_j \left\{ \pi_o(H_j) \prod_{v \in V} f_{j,v}(y_v) \right\}.$$

We shall consider distributed identification of a MAP estimate in cases when a single decision maker having access to all observations is not available. More specifically, we focus on distributed algorithms in which each sensor collaborates with other sensors and thereby forms an estimate of the posterior distribution. This collaboration is limited by a communication network structure represented by a directed graph $G = (V, E)$, which is assumed to be strongly connected in order to avoid trivialities. The vertices V of graph G correspond to sensors and an ordered pair (v', v) of vertices belongs to the edge set E if there exists a communication link from sensor v' to sensor v. A sensor is called a *neighbor* of sensor v if it has a communication link to v. We denote the set of neighbors of v by $N(v)$. That is,

$$N(v) = \{v' \in V : (v', v) \in E\}, \quad v \in V.$$

3 Collaborative Framework

In this section we provide a collaborative framework for the distributed MAP estimation problem based on local message passing algorithms known as belief propagation (BP) [15, 26]. Since BP is generically formulated in terms of Markov random fields (MRFs), we start here with a brief discussion of MRFs.

A random vector $X = (X_v : v \in V)$ is an MRF with respect to an undirected graph $\bar{G} = (V, \bar{E})$ if it possesses certain conditional independence properties defined with respect to neighborhood relations in $\bar{G}$. Namely, for each $v \in V$, given the neighbors $X_{v'} : (v, v') \in \bar{E}$ of v with respect to $\bar{G}$, the random variable X_v is conditionally independent of the remaining entries of X. The Hammersley-Clifford theorem [15, 26] establishes under mild technical conditions that a random vector X is a MRF if and only if its distribution has a product form whose factors are associated with only the cliques of $\bar{G}$. Of particular relevance to our discussion, X is a MRF if for positive mappings $\phi_v : v \in V$, and $\psi_e : e \in \bar{E}$

$$Prob(X_v = x_v,\ v \in V) \propto \prod_{v \in V} \phi_v(x_v) \prod_{e=(v,v') \in \bar{E}} \psi_e(x_v, x_{v'}), \tag{3}$$

for each possible realization $(x_v : v \in V)$. The mappings ϕ_v and ψ_e are referred to as *node* and *edge potentials* respectively. The graphical model $\bar{G}$ thereby provides a platform to represent pairwise correlations in X via edge potentials. In this setting, iterative BP algorithms aim to locally compute marginal distributions of X by passing messages between neighboring nodes along the edges of the graphical model $\bar{G}$. Before specifying BP in further detail, we give next an interpretation of the posterior distribution π in Equation (2) in terms of MRFs.

Consider now a random vector $X = (X_v : v \in V)$ where each marginal X_v takes values in the set $\{H_1, H_2, \cdots, H_M\}$. Let $\bar{G} = (V, \bar{E})$ be an arbitrary graph and the distribution of X be of the form (3) with potentials

$$\phi_v(H_i) = \pi_o(H_i)^{1/|V|} f_{i,v}(y_v), \qquad v \in V, \tag{4}$$

$$\psi_e(H_i, H_j) = \delta(H_i, H_j), \qquad e \in \bar{E}, \tag{5}$$

where $\delta(\cdot, \cdot)$ is the standard Kronecker delta function. If $\bar{G}$ is connected then it is easy to verify that all marginal distributions of X are identical, and furthermore they are equal to the posterior distribution π. Note that due to the arbitrariness of the graphical model $\bar{G}$, there is enormous flexibility in choosing the edges that BP messages traverse. In particular $\bar{G}$ can be chosen to coincide with the communication network graph, $G = (V, E)$. Hence the communication model does not have to bear any relationship to the underlying statistical model.

The main theme in BP from a distributed detection perspective is that each sensor node transmits a vector whose ith component is related to a local estimate for hypothesis H_i at that node. For two nodes v, v' such that $(v, v') \in E$, we denote by $m_k^i(v', v)$ the ith component of the kth message that is transmitted from node v' to node v. In this section we specialize to Pearl's sum-product algorithm [15] for BP. In the mechanics of this algorithm, sensor node v' composes this message by computing the product of most recently received messages pertaining to each hypothesis (excluding message from v), and averaging this product across all hypothesis with adequate weighing to reflect correlations between the different hypothesis. Specifically,

$$m_k^i(v', v) = \sum_{j=1}^{M} \phi_{v'}(H_j)\psi_{(v',v)}(H_i, H_j) \prod_{\hat{v} \in N(v')-\{v\}} m_{k-1}^j(\hat{v}, v').$$

is the message at kth time instant from node v' to node v about hypothesis H_i. On account of the specific potentials (4)–(5), this construction reduces to

$$m_0^i(v', v) = 1 \tag{6}$$

$$m_k^i(v', v) = \phi_{v'}(H_i) \prod_{\hat{v} \in N(v')-\{v\}} m_{k-1}^i(\hat{v}, v'), \tag{7}$$

along any edge $(v', v) \in E$, for each hypothesis H_i, and for each round $k \geq 1$. Messages are used by recipient nodes to compile their *beliefs*, which are estimates for the posterior distribution π and are defined as follows:

Definition 3.1 (Belief) The belief $\hat{\pi}_k^v = (\hat{\pi}_k^v(H_i) : i = 1, 2, \cdots, M)$ of node $v \in V$ at round k is the probability vector that satisfies

$$\hat{\pi}_k^v(H_i) = K\phi_v(H_i) \prod_{v' \in N(v)} m_k^i(v', v),$$

for some positive constant K that does not depend on i.

From the viewpoint of distributed system operation it is worthwhile to note that: (i) Each message is determined locally by the observation at the sensor and the prior messages received from neighboring sensors, (ii) message composition does not require global knowledge of sensor models, and (iii) the algorithm also entails a relaxed synchronization among sensors, as it can be implemented by programming each sensor to send out initial messages immediately and to send out its kth messages only after receiving $(k-1)$th messages from all of its neighbors.

If G is a singly-connected graph then well-known results [26] guarantee that each belief vector $\hat{\pi}_k^v,\ v \in V$, converges to the true posterior distribution π within a finite number of rounds. For general graphs and general potentials the sum-product algorithm is not expected to converge. Our focus is whether the scheme does indeed converge for the special information structure endowed by the classification problem. In exploring this possibility it will be convenient to transform the original problem into a linear dynamical system.

Towards this end, we identify each directed edge $e \in E$ by its source vertex $s(e)$ and its destination vertex $d(e)$ such that $e = (s(e), d(e))$. For each pair of edges $e, e' \in E$ let

$$a_{ee'} = \mathbf{1}\{s(e) = d(e') \text{ but } s(e') \neq d(e)\}. \tag{8}$$

Here and in rest of the paper $\mathbf{1}\{\cdot\}$ denotes the binary function whose value is 1 if its argument is correct, and 0 otherwise. In particular, $a_{ee'} = 1$ if and only if edge e' leads to the source of edge e and the ordered pair (e', e) does not form a directed cycle. For each hypothesis $i = 1, 2, \cdots, M$, let

$$u^i(v) \quad = \quad \log(\phi_v(H_i)), \qquad v \in V, \tag{9}$$

$$x_k^i(e) \quad = \quad \log(m_k^i(e)), \qquad e \in E. \tag{10}$$

where, we have denoted $m_k^i(s(e), d(e)) \equiv m_k^i(e)$ for notational simplicity. Taking the logarithm of both sides in equalities (6)–(7) leads to the linear system

$$x_k^i(e) = u^i(s(e)) + \sum_{e' \in E} a_{ee'} x_{k-1}^i(e'), \qquad x_0^i(e) = 0. \tag{11}$$

Define the vector $w^i = (w^i(e) : e \in E)$ by setting $w^i(e) = u^i(s(e))$, and define the binary matrix $A = [a_{ee'}]_{E \times E}$ so that equality (11) takes the vector form

$$x_k^i = w^i + A x_{k-1}^i, \quad x_0^i = 0. \tag{12}$$

Note that states of the dynamical system (12) have one entry per edge, rather than per sensor node. This is due to the mechanics of BP, which imposes that a node sends different messages over different edges. We will later consider modifications of BP that entail sending identical messages over all outgoing edges from a node, and thereby lead to reduced system dimensionality. We point out that the dynamical evolution in Equation (12) depends only on the graphical structure and not on the individual observations. In turn convergence properties of the algorithm is based on primarily the network topology, as we will see in the next section.

4 Consensus and Convergence

We next identify conditions under which BP leads to a consensus in the network in terms of convergence of individual beliefs. We consider the evolution of beliefs from a detection perspective and characterize topologies for which the consensus reflects a centralized MAP estimate. We adopt a definition of consensus that is substantially weak for the conventional objective of estimating the distribution, but is useful when we are only interested in achieving the MAP decision rule:

Definition 4.1 (Consensus) Given a subset S of hypotheses $\{H_1, H_2, \cdots, H_M\}$, the sensor network is said to asymptotically achieve *consensus on* S if S is the smallest set such that

$$\lim_{k\to\infty} \hat{\pi}_k^v(H_i) = 0 \quad \text{for all } H_i \notin S \text{ and all } v \in V \ .$$

If each member of S maximizes π then the sensor network is said to asymptotically achieve a *MAP-consensus*.

We first examine asymptotic properties of BP on graphs G for which the matrix A is primitive. For completeness the definition of primitivity is given next.

Definition 4.2 A nonnegative matrix A is said to be primitive if there exists a positive integer n such that all entries of A^n are strictly positive.

Before stating the main result note that the solution to the linear system (12) satisfies

$$x_k^i = \sum_{n=0}^{k-1} A^n w^i, \qquad k \geq 1. \tag{13}$$

Following results rely on the following straightforward observation, which is given here without proof:

Lemma 4.1 *Given positive integer n, $A^n = [a_{ee'}^n]_{E\times E}$ where $a_{ee'}^n$ is the number of directed paths of length $n+1$ that start with edge e', end with edge e, and that do not contain any 2-hop cycles.*

Suppose that A is primitive so that $A^n > 0$ for some integer $n > 0$. Since A is a binary matrix all entries of A^n are at least 1, which implies that $\lim_{k\to\infty} a_{e,e'}^k = \infty$. Therefore the spectral radius of A, denoted here by $\rho(A)$, is then strictly larger than 1. Let $(r_e : e \in E)$ and $(l_e : e \in E)$ be respectively a right and a left eigenvector of A corresponding to the eigenvalue $\rho(A)$, suitably normalized so that $r_e > 0,\ l_e > 0$ for $e \in E$ and $\sum_{e\in E} r_e l_e = 1$. We define the *weighted out-degree* $o(v)$ of each sensor node $v \in V$ as

$$o(v) \quad = \sum_{v':v\in N(v')} l_{(v,v')}.$$

Theorem 4.1 *If A is primitive then for each node $v \in V$*

$$\lim_{k\to\infty} \hat{\pi}_k^v(H_i) = 0$$

for each hypothesis $i \in \{1, 2, \cdots, M\}$ such that

$$\prod_{v\in V} \phi_v(H_i)^{o(v)} < \max_j \left\{ \prod_{v\in V} \phi_v(H_j)^{o(v)} \right\}.$$

Proof. Define the matrices $H = [h_{ee'}]_{E\times E}$ and $\Lambda_k,\ k \geq 1$, by setting $h_{ee'} = r_e l_{e'}$ for $e, e' \in E$ and $\Lambda_k = A^k \rho(A)^{-k} - H$. From standard results in Perron-Frobenius theory (see, for example, [11, Theorem 8.5.1]) which states that there exists positive constants C and $\gamma < 1$ such that

$$\|\Lambda_k\|_\infty \leq C\gamma^k. \tag{14}$$

Let $\alpha(k) = \sum_{n=0}^{k-1} \rho(A)^n$. By Equality (13) for each hypothesis $i = 1, 2, \cdots, M$

$$\lim_{k\to\infty} \frac{x_k^i}{\alpha(k)} \;=\; \lim_{k\to\infty} \sum_{n=0}^{k-1} \left(\frac{A^n}{\rho(A)^n}\right) \frac{\rho(A)^n}{\alpha(k)} w^i \;=\; Hw^i + \lim_{k\to\infty} \sum_{n=0}^{k-1} \Lambda_n \frac{\rho(A)^n}{\alpha(k)} w^i. \tag{15}$$

Note that for each l such that $0 \leq l \leq k-1$

$$\left\| \sum_{n=0}^{k-1} \Lambda_n \frac{\rho(A)^n}{\alpha(k)} w^i \right\| \;\leq\; \sum_{n=0}^{l} \|\Lambda_n w^i\| \frac{\rho(A)^n}{\alpha(k)} + \sum_{n=l+1}^{k-1} \|\Lambda_n w^i\| \frac{\rho(A)^n}{\alpha(k)} \;\leq\; C\|w^i\| \sum_{n=0}^{l} \frac{\rho(A)^n}{\alpha(k)} + C\|w^i\| \sum_{n=l+1}^{\infty} \gamma^n, \tag{16}$$

where, in the last inequality, we have used inequality (14) together with the fact that $\rho(A)^n/\alpha(k) \leq 1$ owing to $\rho(A) > 1$. One can thus choose l large enough so that the last term on the rhs of (16) is arbitrarily small, and for a given value of l one can choose k large enough to make the first term therein arbitrarily small. In turn by (15)

$$\lim_{k\to\infty} \frac{x_k^i}{\alpha(k)} = Hw^i,$$

which further implies via the definition of H that for each edge $e \in E$

$$\lim_{k\to\infty} \frac{x_k^i(e)}{\alpha(k)} \;=\; r_e \sum_{e'\in E} l_{e'} w^i(e') \;=\; r_e \sum_{v'\in V} o(v') u^i(v). \tag{17}$$

By Definitions 3.1 and (10)

$$\hat{\pi}_k^v(H_i) \;=\; K' \phi_v(H_i) \exp\left(\sum_{v'\in N(v)} x_k^i(v', v) \right)$$

for some positive constant K' that does not depend on i; in turn for any two hypotheses H_i, H_j

$$\lim_{k\to\infty} \frac{1}{\alpha(k)} \log \frac{\hat{\pi}_k^v(H_i)}{\hat{\pi}_k^v(H_j)} = \lim_{k\to\infty} \frac{1}{\alpha(k)} \left(\sum_{v'\in N(v)} x_k^i(v',v) - \sum_{v'\in N(v)} x_k^j(v',v) \right)$$

since $\lim_{k\to\infty} \alpha(k) = \infty$. Therefore by equality (17) and definition (9) of $u^i(v)$,

$$\lim_{k\to\infty} \frac{1}{\alpha(k)} \log \frac{\hat{\pi}_k^v(H_i)}{\hat{\pi}_k^v(H_j)} = \sum_{v'\in V} o(v') \log \phi_{v'}(H_i) - \sum_{v'\in V} o(v') \log \phi_{v'}(H_j). \tag{18}$$

The theorem now follows because $\lim_{k\to\infty} \hat{\pi}_k^v(H_i)/\hat{\pi}_k^v(H_j) = 0$ unless the rhs of (18) is non-negative. □

Note that if A is primitive Theorem 4.1 guarantees that MAP-consensus is asymptotically achieved provided that all nodes of G have the same weighted out-degree. This property is satisfied by regular graphs, as formally stated as a corollary next, as well as the large random graphs that are considered in the following section.

Corollary 4.1 *Let $G = (V, E)$ be a finite connected d-regular graph ($d > 2$), i.e., a connected graph where each vertex has degree d. Furthermore, let the matrix, A, as defined in Equation 8 be primitive. Then a MAP-consensus is achieved with the BP algorithm.*

While Definition 4.2 does not suggest a practical way of testing for primitivity, an alternate definition helps in some cases. More specifically, for $e \in E$, let $\{n_1, n_2, \cdots\}$ be the set of integers n such that $a_{ee}^n > 0$. Then A is primitive if and only if the greatest common divisor of $\{n_1, n_2, \cdots\}$ is 1 (see, for example, [11, Theorem 8.5.3] for a proof). In particular A is primitive if it has a strictly positive diagonal entry. We will appeal to this observation in Section 5 where we introduce variants of BP that achieve MAP-consensus on arbitrary graphs.

We consider next two interesting topologies that violate the primitivity condition in Theorem 4.1. An alternate definition ([11, Definition 8.5.0 and Theorem 8.5.2]) of primitivity implies that if A is not primitive then either it is reducible, or it has multiple eigenvectors of modulus $\rho(A)$. Theorem 4.2 below concerns the case when G has a tree structure and in turn A is nilpotent [1]. Theorem 4.3 concerns ring graphs wherein $\rho(A) = 1$ and A has again multiple maximal eigenvalues. The theorems establish that MAP-consensus is achieved in both cases. However in graphs for which A has multiple eigenvalues of modulus $\rho(A)$ and $\rho(A) > 1$, such as $L \times L$ tori with odd L, $\rho(A)^{-n}A^n$ oscillates persistently and consensus should not be expected in general.

In the scope of the following two theorems, $[G]$ represents the undirected graph obtained as follows: $[G]$ has the same node set as G, and an unordered pair $[v, v']$ of nodes is an edge of $[G]$ if $(v, v') \in E$ and $(v', v) \in E$.

[1] It is well-known that on trees BP leads to the true posterior distributions even for general Markov fields.

Theorem 4.2 *(Trees) Suppose that communication links are bidirectional so that if $(v, v') \in E$ then $(v', v) \in E$ also. If $[G]$ is a tree then the network asymptotically achieves MAP-consensus with BP algorithm.*

Proof. If $[G]$ is a tree, then the matrix A defined by Equation (8) is nilpotent since $A^n = 0$ for all integers n larger than the diameter of the tree. In turn by Equality (13) the messages converge within a number of steps no larger than the diameter. To identify the limit, note that by Lemma 4.1

$$\sum_{n=0}^{\infty} a_{ee'}^n = \mathbf{1}\{\text{there exists a directed path that starts with edge } e', \text{ ends with edge } e, \text{ and contains no 2-hop loops}\}$$

for $e, e' \in E$. Since $[G]$ is a tree, Equality (13) leads to

$$\lim_{k\to\infty} x_k^i(e) = \sum_{v\in V} \mathbf{1}\{\text{dist}(v, s(e)) < \text{dist}(v, d(e))\} u^i(v), \qquad e \in E,\ i = 1, 2, \cdots, M,$$

where $\text{dist}(v, v')$ represents the length of the unique path between vertices $v, v' \in V$. By Definition 3.1 the limit of the estimate $\hat{\pi}_k^v(i)$ at each sensor $v \in V$ is equal to the posterior distribution π. □

Theorem 4.3 *(Rings) Suppose that communication links are bidirectional so that if $(v, v') \in E$ then $(v', v) \in E$ also. If $[G]$ is a ring then the network asymptotically achieves MAP-consensus with BP algorithm.*

Proof. If $[G]$ is a ring then $A^n = A^{n+|V|}$ for $n \geq 1$; in turn Equality (13) leads to the vector equalities

$$\lim_{k\to\infty} \frac{x_k^i}{k} = \frac{1}{|V|} \sum_{n=0}^{|V|-1} A^n w^i, \qquad i = 1, 2, \cdots, M.$$

Under the hypothesis of the theorem $\sum_{n=0}^{|V|-1} a_{ee'}^n = 1$ for all edges $e, e' \in E$ that have a common orientation (that is, clockwise or counter-clockwise) and that $\sum_{n=0}^{|V|-1} a_{ee'}^n = 0$ otherwise. Therefore

$$\lim_{k\to\infty} \frac{x_k^i(e)}{k} = \frac{1}{|V|} \sum_{v\in V} u^i(v), \quad e \in E. \tag{19}$$

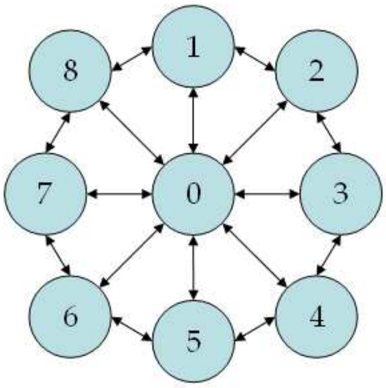
Since

$$\lim_{k\to\infty} \frac{1}{k} \log \frac{\hat{\pi}_k^v(H_i)}{\hat{\pi}_k^v(H_j)} = \lim_{k\to\infty} \frac{1}{k} \left(\sum_{v'\in N(v)} x_k^i(v', v) - \sum_{v'\in N(v)} x_k^j(v', v) \right), \qquad i, j = 1, 2, \cdots, M,$$

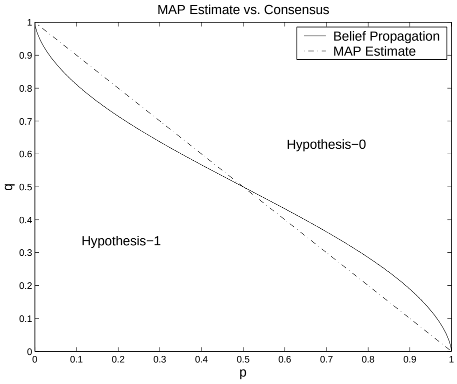
due to Definitions 3.1 and (10), Equation (19) implies that

$$\lim_{k\to\infty} \frac{1}{k} \log \frac{\hat{\pi}_k^v(H_i)}{\hat{\pi}_k^v(H_j)} = \frac{2}{|V|} \left(\sum_{v'\in V} u^i(v') - \sum_{v'\in v} u^j(v') \right). \tag{20}$$

The conclusion of the theorem now follows because $\sum_{v'\in V} u^i(v') = \log(\pi(H_i))$ due to definitions (9), (4), (2), and in turn by Equation (20) $\lim_{k\to\infty} \hat{\pi}_k^v(H_i)/\hat{\pi}_k^v(H_j) = 0$ unless $\pi(H_i) \geq \pi(H_j)$. □



(a) A 9-node irregular communication graph



(b) Decision regions for the MAP estimate (dashed) and the final consensus of BP (solid).

Figure 2: Illustration of how asymmetric graphs bias the consensus decision away from the optimal.

Although the above theorems establish MAP-consensus in the sense of Definition 4.1, convergence of beliefs to a single hypothesis may not emerge. If multiple MAP estimate exists the maximum posteriors at the different sensors oscillate between the different hypothesis within the MAP set (see [23] for an example).

For general topologies the consensus identified by Theorem 4.1 depends on the individual observations as well as the weighted out-degrees of the nodes at which each observation is taken. In the general case the consensus reflects right MAP estimates for certain values of observations, and wrong choices for others. We close this section with an illustration of this situation.

Example: Consider binary hypothesis testing in a 9-sensor network under the communication structure represented by the graph of Figure 2(a). Each edge in the graph represents two directed edges in opposite directions. Maximal eigenvectors of A no longer have equal weights corresponding to each edge, in particular the weighted out-degrees $o(1), o(2), \cdots, o(8)$ are equal and $o(0)/o(1) = 1.5091$. Suppose that the observations $(y_v : v \in V)$ translate to node potentials $\phi_0 = [q, 1-q]$, $\phi_1 = [p, 1-p]$ and $\phi_v = [0.5, 0.5]$ for $v = 2, 3, \cdots, 8$, where $p, q \in [0, 1]$. Figure 2(b) illustrates, for different values of p and q, the true MAP estimate and the final consensus due to BP as identified by Theorem 4.1. Note that this consensus reflects a flawed estimate if (p, q) lies in the area between the solid and dashed lines.

4.1 Random Graphs

In this section we describe distributed hypothesis testing for random graphs, $G = (V, E)$, which are constructed as follows. The vertices V of the graph, G, are associated with N sensor nodes uniformly distributed in a square *unit* area denoted by the region $\mathcal{Z}$ centered at zero as shown in the Fig. 3(a). The edges of the graph correspond to communication connectivity between the sensor nodes. In particular two vertices $v_1,\, v_2$, are connected, i.e., $(v_1,\, v_2) \in E$ if the Euclidean distance between the vertices is smaller than, R, the communication connectivity radius. In addition, for the sake of mathematical simplicity, we consider a periodic extension of the graph to avoid dealing with boundary conditions. In particular two vertices, $v_1,\, v_2$, whose planar coordinates are respectively, $(v_1^x, v_1^y),\, (v_2^x,\, v_2^y)$, are connected by a link if

$$\min \left(\left\| \begin{array}{c} v_1^x - v_2^x \\ v_1^y - v_2^y \end{array} \right\|, \left\| \begin{array}{c} v_1^x - (1 - v_2^x) \\ v_1^y - v_2^y \end{array} \right\|, \left\| \begin{array}{c} v_1^x - v_2^x \\ v_1^y - (1 - v_2^y) \end{array} \right\|, \left\| \begin{array}{c} v_1^x - (1 - v_2^x) \\ v_1^y - (1 - v_2^y) \end{array} \right\| \right) \leq R \tag{21}$$

One can expect from the above construction that for an appropriate connectivity radius, the graph, G is not only connected but that each vertex has the same number of neighbors on an average. We can then expect to derive a result similar to Corollary 4.1. The minimum radius of connectivity, R, required is of the order of $\frac{\log N}{\sqrt{N}}$ to ensure graphical connectivity of N uniformly distributed sensor nodes in a *unit* area [18]. However, for this minimum radius the variance in the degree for each node (i.e. number of neighbors for each node) is large. To ensure a near constant degree with high probability we need a slightly larger radius of connectivity, i.e., $R = \frac{2 \log^{3/2} N}{\sqrt{N}}$. This connectivity radius, R, not only ensures primitivity of the A matrix as described by Equation (8) but also ensures that the number of neighbors for each node is approximately constant. We collect these results in the following lemma, whose proof is provided in the appendix.

Lemma 4.2 *Consider the random graph, $G = (V, E)$, described above with connectivity radius, $R = \frac{2 \log^{3/2} N}{\sqrt{N}}$. It follows with high probability that: (a) The number of neighbors for each node is in the range $4\pi \log^3(N)(1 \pm \frac{1}{\log(N)})$; (b) The A matrix as defined by Equation 8 is primitive; (c) consensus is achieved at all the nodes.*

Proof. See the appendix for proof of the first two statements. The last statement follows by direct application of Theorem 4.1.

MAP consensus is harder to establish even though the choice of connectivity radius ensures approximate regularity and sets up parallels with Corollary 4.1. To understand this point observe from Equation (17) that the consensus is the hypothesis, j, that maximizes the weighted sum of log-likelihood probability at each sensor. The weights depend on the left eigenvector, l, of the A matrix. Consequently, MAP consensus is realized only when the components of the left eigenvector are identical (or nearly identical), a situation achieved for regular graphs as in Corollary 4.1. If one were to view random graphs as a random perturbation of regular graphs the resulting eigenvector perturbation is not guaranteed to be small in general especially when the difference between first and second eigenvalue can be arbitrarily small[2]. To address this point we consider two different strategies: (A) d-nearest neighbor graph: The graph is formed by connecting the d-nearest neighbors for each node. (B) Approximate Belief Propagation: Here message passing algorithm is modified so that MAP consensus is achieved and the approximation converges to the exact BP message passing scheme as the number of sensors approach infinity. This latter strategy is related to the schemes discussed in the following section, where MAP achieving consensus algorithms for arbitrary graphs are described.

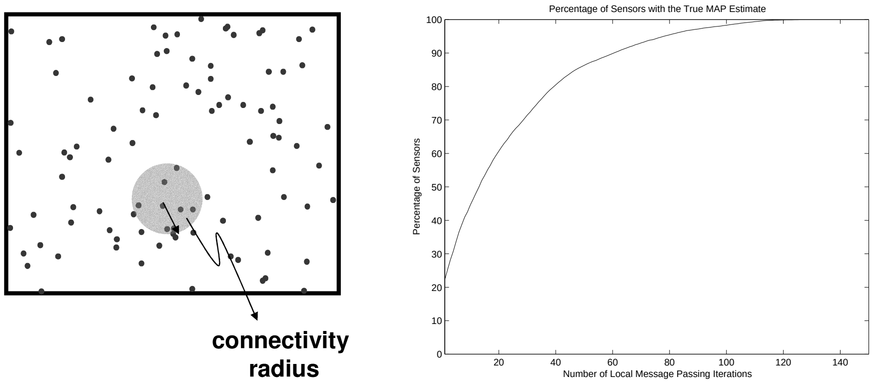
For d-nearest neighbor graph MAP consensus is guaranteed by Corollary 4.1. The main issue here is the asymmetric connectivity, i.e., different nodes have different connectivity patterns and not all nodes within a given connectivity radius may be connected. This issue can be addressed by choosing a sufficiently large communication connectivity radius. Indeed Lemma 4.2 asserts that if $d = \lfloor 4\pi \log^3 N \rfloor$ then with high probability these nodes are contained within a radius $R = 2\frac{\log^{3/2} N}{\sqrt{N}}$. Furthermore, asymptotically there are no more than $\lfloor 4\pi \log^3 N \rfloor$ nodes within this radius.

For the second strategy consider the following approximate BP algorithm, specifically,

$$\tilde{x}_k^i = w^i + A(I+\Delta)^{-1}\tilde{x}_{k-1}^i, \quad \tilde{x}_0^i = 0 \tag{22}$$

where, $\Delta = \text{diag}[\delta_{ii}]$ is a suitably chosen diagonal matrix with, $|\delta_{ii}| \leq 1/\log(N)$ and such that $A(I+\Delta)$ has equal column sums. Before establishing the latter point observe that $\delta_{ii} \overset{N\longrightarrow\infty}{\longrightarrow} 0$ and so the approximate BP indeed converges to the exact BP in the limit. Furthermore, if $A(I+\Delta)$ is primitive and has equal column sums the corresponding left eigenvector has identical components. Since weighting does not destroy primitivity it follows from an argument identical to Theorem 4.1 (and also described in the following section) that $\tilde{x}_k^i/\alpha(k)$ converges to the sum log-likelihood function (modulo a proportionality constant). Therefore, MAP consensus is achievable. To establish that such a weighting guarantees equal column sums with high probability note that from the proof of Lemma 4.2 it follows that each column of A has $\log^3 N(1 \pm 1/\log N)$ non-zero entries with high probability. Therefore, by choosing a number $|\delta_{jj}| \leq 1/\log N$ the sum of jth column of A can be made to take the value $4\pi \log^3 N$.

[2]eigenvector perturbations are inversely proportional to eigenvalue differences in the unperturbed matrix [11]



(a) Randomly distributed sensors with constant radius of communication connectivity

(b) Rate of convergence to consensus

Figure 3: Simulation of Distributed Detection for a random graph with 400 nodes and 10 different hypothesis.

5 Modified BP Algorithms

While BP lends itself to analysis and guarantees MAP-consensus in certain cases, it has a number of drawbacks: (a) Convergence and consensus under BP depends on primitivity of a transition matrix A that is determined by the communication graph G, (b) the consensus in the network need not be a MAP-consensus unless the graph G possesses fairly strict regularity properties, (c) Actual values of transmitted messages tend to 0 as the algorithm progresses, thereby leading to potential numerical instabilities, (d) BP algorithms require customizing information for a particular node, in that message from node v to node v' is a function of messages received from neighbors of v excluding v'. It would be simpler and energy efficient if one could fuse the received messages and broadcast a single message without the need for customization. In this section we address each issue and provide a progressive list of modified BP algorithms which efficiently lead to MAP-consensus in general graphs.

We specify below each modification via a linear system, similar to (11) for BP, but use the symbol z_k^i for the state of this system in order to avoid confusion with the BP algorithm.

Relaxing the primitivity requirement: Self-loops. Consider a message passing algorithm akin to BP but which is represented by the linear system on $\mathbf{R}^E$

$$z_k^i = w^i + (I + A) z_{k-1}^i, \qquad z_0^i = 0, \tag{23}$$

for each hypothesis $i = 1, 2, \cdots, M$. Here I is the identity matrix of appropriate size, and $z^i_k(e)$ denotes the kth message transmitted along the directed edge $e \in E$. In particular, for $(v, v') \in E$,

$$z^i_k(v, v') = u^i(v) + z^i_{k-1}(v, v') + \sum_{\hat{v} \in N(v)} z^i_{k-1}(\hat{v}, v).$$

The algorithm thus differs from BP due to the second term on the rhs, namely a replica of the previous message sent along the same edge. In the graphical representation of communication, this amounts to a BP algorithm when G is modified so that each node $v \in V$ has a self-loop, that is an edge from v to itself. Clearly, this does not entail any modification in the physical communication infrastructure. Notice that the matrix $(I + A)$ is primitive regardless of A, and thus Theorem 4.1 applies.

Relaxing the regularity requirement: Weighted Self-loops. The main issue suggested by Theorem 4.1 is that the consensus achieved via BP is in general based on a weighted log-likelihood ratio; therefore BP may lead to incorrect MAP estimates if the graph G is unbalanced in the sense that the weighted out-degree varies from node to node. The graph G can in fact be balanced by appropriate message passing, namely by introducing self-loops in G as in (23) in a way to increase the influence of poorly connected nodes on the evolution of beliefs. We introduce two modifications both of which lead to the desired solution: Here the number of self-loops for a particular node depends on the degree of that node. Specifically,

$$z^i_k = w^i + (L + A) z^i_{k-1}, \qquad z^i_0 = 0, \tag{24}$$

for some diagonal matrix $L = \text{diag}[l_{ee}]_{E \times E}$ where $l_{ee} = d_{max} - d_e$, where d_e is the number of edges e' such that $a_{e'e} = 1$ and $d_{max} = \max_{e \in E} d_e$. It then follows that column sums of the matrix $L + A$ are all identical; therefore the coefficients of the dominant left eigenvector of $L + A$ are all equal. Straightforward adaptation of Theorem 4.1 now implies that the network asymptotically achieves MAP-consensus. Alternatively, in cases where the maximum degree is not known a priori at each node, the same effect can be obtained by choosing $q_{ee} = d - d_e$ where d is any integer that is known to be larger than $d_{\max}$.

Dealing with numerical instabilities: Normalization. Consider now another variation of BP in which each node normalizes the messages it transmits by the number of messages it receives. The composition of messages also differs from BP in the inclusion of log-potentials u^i. Specifically, we consider message passing identified by the linear system

$$z^i_k = (I + A) D z^i_{k-1}, \qquad z^h_0 = w^i, \tag{25}$$

where, $D = \text{diag}[d_{ee}]_{E\times E}$ is a diagonal matrix with $d_{ee} = (1 + d_e)^{-1}$ for $e \in E$. Note that, as opposed to (24), this message passing algorithm does not require sensor nodes to know any global aspects of the graph topology, such as d_{max} or any upper bound to it. Furthermore the spectral radius of $(I + A)D$ is one. Consequently, log-messages $z_k^i(e)$ are bounded.

To see the convergence and consensus aspects of this algorithm, note first that the matrix $(I+A)D$ is primitive. Columns of $(I+A)D$ are probability vectors; in particular they have identical sums and therefore the coefficients of the dominant left eigenvector of $(I+A)D$ are all equal. Hence $\lim_{k\to\infty}((I+A)D)^k$ exists and the limit matrix has identical columns. In turn for each edge $e \in E$

$$\lim_{k\to\infty} z_k^i(e) \;=\; \lim_{k\to\infty} ((I + A)D)^k w^i \;=\; c(e) \sum_{v\in V} u^i(v)$$

for some positive constant $c(e)$. This implies that for large enough values of k, the vector $(z_k^i(e) : i = 1, 2, \cdots, M)$ and the posterior distribution π have common modes. The algorithm (25) therefore achieves MAP-consensus provided that $(z_k^i(e) : i = 1, 2, \cdots, M)$ is interpreted as the belief of the source node $s(e) \in V$ at time k. Note also that if the matrix A is locally available so that the constant $c(e)$ can be computed, then the exact posterior distribution can be reconstructed via

$$\lim_{k\to\infty} \frac{\exp(z_k^i(e)/c(e))}{\sum_{j=1}^{M} \exp(z_k^j(e)/c(e))} = \pi(H_i).$$

Reducing transmissions: Broadcast. We consider next another variation of BP based on the theme of nodes broadcasting the same message to all recipients. This differs from BP as it does not require compiling and transmitting separate message to separate sensor nodes. The message passing algorithm can then be represented with a linear system on a reduced state space. More specifically we consider the the following system on $\mathbf{R}^{|}V|$: Let the matrix $\bar{A} = [\bar{a}_{vv'}]_{V\times V}$ be defined by setting $\bar{a}_{vv'} = 1$ if $(v', v) \in E$ and $\bar{a}_{vv'} = 0$ otherwise. For each node $v \in V$ let d_v be the number of edges emanating from v, and define the diagonal matrix $\bar{D} = \text{diag}[\bar{d}_{vv}]_{V\times V}$ by setting $\bar{d}_{vv} = (1 + d_v)^{-1}$ for $v \in V$. We consider the message passing algorithm represented by

$$z_k^i = (I + \bar{A})\bar{D} z_{k-1}^i, \qquad z_0^i = u^i. \tag{26}$$

Here $z_k^i(v)$ denotes ith component of the message broadcast by node $v \in V$ in round k of the algorithm. Let us redefine the belief vector of node v at time k as $(z_k^i(v) : i = 1, 2, \cdots, M)$. Since $(I + \bar{A})\bar{D}$ is a primitive matrix whose transpose is a stochastic matrix, the analysis of algorithm (25) can be mimicked to see that the algorithm (26) achieves MAP-consensus. In Section 6.2 we will establish a stronger version of this result under the assumption of random message losses.

We summarize the conclusions of this sections in the following theorem, which is stated without proof:

Theorem 5.1 *For an arbitrary strongly-connected directed graph G, each of the message passing algorithms (24)–(26) achieves MAP-consensus.*

6 Robustness Issues

6.1 Detection with Finite-Link Capacity

We deal with finite link capacity by taking a robustness perspective. Specifically, we study impact of message quantization on the MAP estimate. The messages (rather than their logarithms) are quantized with a constant-ratio quantizer, i.e., the quantization levels, $(p_0, p_1, \ldots, p_q, \ldots)$, satisfies $p_{j+1}/p_j = 1+\gamma$. Mathematically, the quantizer, $Q(\cdot)$ is described by, $Q(m^i) = \text{argmin}\{p_j \geq m^i\}$ and alternatively can be described by the following uncertain system:

$$Q(m^i) = (1+\delta)m^i, \text{ for some } |\delta| \leq \gamma.$$

In the logarithmic domain this translates to:

$$z_k^i = Az_{k-1}^i + \log(1+\delta)\mathbf{1} + w^i, \;\; z_0^i = 0 \tag{27}$$

where $\mathbf{1}$ is a column vector of all ones. Based on the results of the previous section, we assume that A^T is a primitive stochastic matrix without loss of generality. Our task now reduces to quantifying the maximum allowable quantization level, γ, so that a MAP-consensus can be achieved. Let $\gamma_{|V|}(y)$ denote the maximum admissible quantization level for a network of $|V|$ identical sensors for a given realization of sensor observations, $Y = y$. The assumption of identical sensors implies that the distributions for the observations depend only on the hypothesis and is independent of the sensor. Consequently, $f_{i,v}(y_v) = g_i(y_v)$, with $g_i(\cdot)$ a family of probability distributions indexed by hypothesis, i. We then have the following result.

Theorem 6.1 *For quantized observations MAP consensus is guaranteed if $\gamma_{|V|}(y)$ satisfies:*

$$\log(1+\gamma_{|V|}(y)) < \min_{j \neq i_{MAP}} \left| \frac{1}{|V|} \sum_{v \in V} \log \frac{g_{i_{MAP}}(y_v)}{g_j(y_v)} \right| \tag{28}$$

where i_{MAP} denotes the MAP hypothesis. Further, as $|V| \to \infty$ the required quantization level $\gamma_{|V|}(y)$ converges to a constant $\gamma_* < \min_{(j \neq i)} \exp\left(D(g_i \| g_j)\right) - 1$, where $D(\cdot\|\cdot)$ is the Kullback-Leibler distance.

Proof. Proceeding along the lines of Theorem 4.1 we note that for any two hypothesis, i and j, we have,

$$\frac{z_n^i}{n} - \frac{z_n^j}{n} = \sum_{k=0}^{n-1} \frac{A^k}{n}(w^i - w^j) + \sum_{k=0}^{n-1} \frac{A^j}{n}(\log(1+\delta_{k-j}^i) - \log(1+\delta_{n-k}^j))$$

where without of loss of generality we have assumed that the components of the right eigenvector is equal to one. Now, if hypothesis, $i = i_{MAP}$, is the MAP hypothesis then it is more likely than j. We have $\sum_{v\in V} u^i(v) - \sum_{v\in V} u^j(v) \geq 0$. We need to ensure that the difference in the beliefs at each node is larger than zero as well, i.e., $\liminf_{n\to\infty} \frac{z_n^i}{n}(v) - \frac{z_n^j}{n}(v) > 0$. Now, proceeding along the same lines as Theorem 4.1 and observing that the maximal eigenvalue is unity and the corresponding left eigenvector has identical components, we obtain

$$\liminf_{n\to\infty} \frac{z_n^i}{n}(v) - \frac{z_n^j}{n}(v) = \liminf_{n\to\infty} \left(\sum_{v\in V} (u^i(v) - u^j(v)) + e_v^T \sum_{k=0}^{n-1} \frac{A^k}{n} \left(\log(1+\delta_{n-k}^i) - \log(1+\delta_{n-k}^j) \right) \right)$$

where, e_v is a unit column vector whose vth component is equal to one and all other components are equal to zero. To ensure that quantization does not modify the MAP estimate we need:

$$\sum_{v\in V} (u^i(v) - u^j(v)) \geq \limsup_{n\to\infty} \sup_{\delta_0, \delta_1, \dots, \delta_{n-1}} \sum_{k=0}^{n-1} \frac{e_v^T A^k}{n} \log(1+\delta_{k-j}) = |V| \log(1+\gamma_V(y))$$

where, the second equality follows from the fact that $\sup_{b(\cdot)\leq b_{max}} |\sum_k a(k)b(t-k)| = b_{max} \sum_k |a(k)|$ and from primitivity and stochasticity of A. This establishes the first statement. Now dividing by $|V|$ we see that the expression on the left is exactly the average sum of i.i.d. random variables. Now let l be the true hypothesis, which is not necessarily the MAP estimate. By appealing to the strong law of large numbers we have that:

$$\left| \frac{1}{|V|} \sum_{v\in V} \log \frac{g_{i_{MAP}}(y_v)}{g_j(y_v)} \right| \overset{|V|\to\infty}{\longrightarrow} |D(g_l \| g_{i_{MAP}}) - D(g_l \| g_j)| \overset{a}{=} D(g_l \| g_j)$$

where, the equality (a) follows from the fact that $i_{MAP} \longrightarrow l$ almost surely as $|V| \to \infty$. Since the inequality must hold for all $j \neq i$ the result follows. □

Consequently, it follows that if the distributions corresponding to each hypothesis are well-separated, the number of quantization-levels is also a constant in the log-message domain. This follows from Equation (27), which amounts to discretizing log-messages into a constant number of bits equal to $\log_2(B/\log(1+\gamma))$ bits per message (where B is the maximum value of any log-message). The boundedness assumption follows from the results in Section 4, where we showed that the log-messages are bounded.

6.2 Packet Losses and Asynchronous Operation

In this section we relax the assumption that one message is transmitted along each communication link at each round of the algorithm. Our aim here is to account for the following two effects: First, messages may be corrupted and lost due to imperfections in point-to-point communication. Although link layer protocols would provide some relief against this issue, robustness of network operation against message losses needs to be addressed, especially if the physical communication medium is wireless. Secondly, one can imagine situations where some sensors operate on a slower time-scale than others, thereby slowing down the network under the lock-step message-passing algorithm outlined in Section 3. This limitation may be overcome if each sensor contributes to the collaborative effort at its own time-scale. In both cases described above the network operation is asynchronous in the sense that not all links are necessarily active at each round of the algorithm. We next establish the attendant effects of this generalization in stochastic setting.

We consider the broadcast operation of Section 5 in the case when the connectivity of the network is time-varying. Namely, evolution of the messages is represented as

$$z_{k+1}^i = (I + A_k)D_k z_k^i, \qquad z_0^i = u^i, \tag{29}$$

where $A_k = [a_{vv'}(k)]_{V\times V}$ is a binary matrix and $D_k = \text{diag}[d_{vv}(k)]_{V\times V}$ is a diagonal matrix with

$$d_{vv}(k) = \left(1 + \sum_{v'\in V} a_{v'v}(k)\right)^{-1},$$

so that in particular columns of $(I + A_k)D_k$ are probability vectors. The matrix A_k identifies sensor nodes that communicate in round k. Namely, for any two sensors v, v', the corresponding entry of of A_k is defined as

$$a_{vv'}(k) = \begin{cases} 1 & \text{if node } v \text{ receives message of node } v' \text{ at round } k \\ 0 & \text{otherwise.} \end{cases}$$

We shall say that link $(v', v) \in E$ is functional at round k if $a_{vv'}(k) = 1$. The system (29) then describes the evolution of local beliefs when each transmitted message is normalized by the number of outgoing functional links (i.e., the number of receivers of the message) in the same round. We point out that such an algorithm is consistent with the currently employed wireless protocols.

Theorem 6.2 *Suppose that the matrices $(A_k : k \geq 1)$ are iid, and that $E[A_1]$ is irreducible. Then for each $v \in V$ there exists a random sequence $(\gamma_k(v) : k \geq 1)$ such that*

$$\lim_{k\to\infty} \frac{z_k^i(v)}{\gamma_k(v)} = \sum_{v'\in V} u^i(v'), \qquad \textit{almost surely,}$$

for $i = 1, 2, \cdots, M$. In particular for large enough values of k, the vector $(z_k^i(v) : i = 1, 2, \cdots, M)$ and the posterior distribution π have common modes.

We prove the theorem via an adaptation of the techniques in [25] for asymptotic analysis of stochastic-matrix products. We start with auxiliary results.

Given a square matrix $P = [p_{nm}]$ define

$$\lambda(P) = 1 - \min_{n_1, n_2} \sum_m \min(p_{n_1 m}, p_{n_2 m}).$$

Let $B_k = [(I + A_k)D_k]^T$, so that B_k is a stochastic matrix and

$$z_k^i = (B_1 B_2 \cdots B_k)^T u^i, \quad k \geq 1. \tag{30}$$

Lemma 6.1 *Under the hypothesis of Theorem 6.2, for each $\epsilon > 0$ there exists $k(\epsilon)$ such that for $k \geq k(\epsilon)$*

$$P\left(\lambda\left(B_{k_o+1} B_{k_o+2} \cdots B_{k_o+k(\epsilon)}\right) < 1\right) > 1 - \epsilon, \quad k_o \geq 0.$$

Proof. It suffices to show that for large enough k all entries of the matrix product

$$B_{k_o+1} B_{k_o+2} \cdots B_{k_o+k}$$

are positive with probability at least $1 - \epsilon$. By definition of B_ks, entries of this product are positive if and only if all entries in

$$(I + A_{k_o+1})^T (I + A_{k_o+2})^T \cdots (I + A_{k_o+k})^T \tag{31}$$

are positive. The $(v, v')^{th}$ entry in the product (31) is positive if and only if a hypothetical message that originates at node v in round k_o can reach node v' by round $k_o + k$ by traversing a functional link in each round. Note that a self-looping link is always functional due to the identity matrix contained in each factor of (31). Let $q(v, v')$ be the probability that link (v, v') is functional at a round, so that without loss of generality $E = \{(v, v') : q(v, v') > 0\}$, and let $(\xi(v, v') : (v, v') \in E)$ be independent geometric random variables where $\xi(v, v')$ has parameter $q(v, v')$. Since $E[A_1]$ is irreducible by hypothesis, the time to reach any node from any other node via functional links is stochastically dominated by $\sum_{(v,v') \in E} \xi(v, v')$. Define the random variable κ as

$$\kappa = \min\left\{k : (I + A_{k_o+1})^T (I + A_{k_o+2})^T \cdots (I + A_{k_o+k})^T \text{ has positive entries}\right\}.$$

Since there are $|V|^2$ node pairs, κ is stochastically dominated by $|V|^2 \sum_{(v,v') \in E} \xi(v, v')$. Let μ be the mean of this latter variable so that

$$P\left(\lambda\left(B_{k_o+1} B_{k_o+2} \cdots B_{k_o+k}\right) < 1\right) \;\geq\; 1 - P(\kappa > k) \;\geq\; 1 - \frac{\mu}{k},$$

where last inequality is an application of Markov's inequality. The lemma follows by choosing $k(\epsilon) = \mu/\epsilon$. □

Corollary 6.1 *Since each B_k takes values from a finite set, there exists a positive number $d < 1$ such that $\lambda(B_{k_o+1}B_{k_o+2}\cdots B_{k_o+k(\epsilon)}) < d$ whenever $\lambda(B_{k_o+1}B_{k_o+2}\cdots B_{k_o+k(\epsilon)}) < 1$, for $k_o \geq 0$.*

For a square matrix $P = [p_{nm}]$ define

$$\delta(P) = \max_{m} \max_{n_1,n_2} |p_{n_1 m} - p_{n_2 m}|. \tag{32}$$

The following lemma is a recitation of [25, Lemma 2]:

Lemma 6.2 *For $k \geq 1$*

$$\delta\left(B_k B_{k-1}\cdots B_1\right) \leq \prod_{l=1}^{k} \lambda(B_l).$$

Proof of Theorem 6.2. Fix $\sigma, \epsilon > 0$ and $k > k(\epsilon)$. Appeal to Lemma 6.2 to write

$$\delta(B_1B_2\cdots B_k) \leq \prod_{l=1}^{k \ mod \ k(\epsilon)} \lambda(B_l) \prod_{l=1}^{\lfloor k/k(\epsilon)\rfloor} \lambda\left(B_{k-lk(\epsilon)+1}B_{k-lk(\epsilon)+2}\cdots B_{k-(l-1)k(\epsilon)}\right).$$

Since each factor of the product on the left hand side is at most 1, Corollary 6.1 implies that the product is larger than σ only if there are more than $\lfloor \log_d(\sigma)\rfloor$ values of l with

$$\lambda\left(B_{k-lk(\epsilon)+1}B_{k-lk(\epsilon)+2}\cdots B_{k-(l-1)k(\epsilon)}\right) > d.$$

Lemma 6.1 now implies that for $k > k(\epsilon)\lfloor \log_d(\sigma)\rfloor$

$$P(\delta(B_1B_2\cdots B_k) > \sigma) \ \leq \ \sum_{l=1}^{\lfloor \log_d(\sigma)\rfloor} \begin{pmatrix} \lfloor k/k(\epsilon)\rfloor \\ l \end{pmatrix} (1-\epsilon)^{\lfloor k/k(\epsilon)\rfloor - l}\epsilon^l \ \leq \ (1-\epsilon)^{k/k(\epsilon)} k^{\lfloor \log_d(\sigma)\rfloor} c,$$

where c does not depend on k. The left hand side is thus summable in k; in turn

$$\limsup_{k\to\infty} \delta(B_1B_2\cdots B_k) \leq \sigma, \quad \text{almost surely,}$$

due to the Borel-Cantelli Lemma. Arbitrariness of σ implies that $\delta(B_1B_2\cdots B_k)$ converges, hence by definition (32) the rows of the product $B_1B_2\cdots B_k$ almost surely become identical (though they do not necessarily settle to a fixed vector). In light of equality (30) the theorem follows by identifying $\gamma_k(v)$ with the vth entry of an arbitrary row of $B_1B_2\cdots B_k$. □

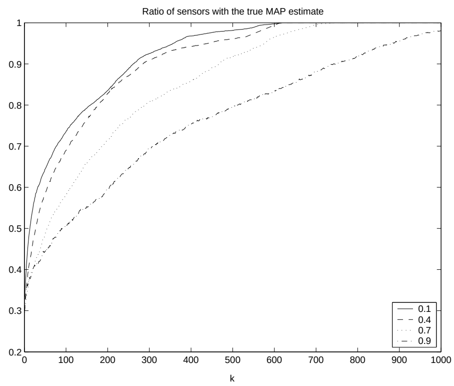


Figure 4: Convergence rate of consensus for different packet loss probabilities; Y-axis denotes percentage of sensors that have achieved consensus; X-axis denotes time index

Remark 6.1 *Note that the proof of Theorem 6.2 relies only on Lemma 6.1; hence the conclusion of the theorem holds under much more relaxed assumptions on the statistics of* $(A_k : k \geq 1)$. *From a deterministic perspective, it is not difficult to see that this conclusion holds if each link is functional infinitely often, provided that the communication graph is irreducible and aperiodic.*

Remark 6.2 *The algorithm requires neither measurement nor estimation of packet loss probabilities.*

To illustrate the effect of packet losses we have performed a number of numerical experiments. A typical simulation of the average convergence is illustrated in Figure 4 for $N = 400$ sensors placed on a uniform grid where any two nodes that are separated by minimum internode distance are connected with a bi-directional communication link. Here p denotes the message loss probability on each link, and losses are taken to occur independently for different links and at different times. From simulations it appears that the convergence rate degrades gracefully and we do not see appreciable differences even for large packet loss probabilities.

7 Energy Budgets

In this section we derive energy budgets for decentralized and distributed detection schemes for a sensor network deployed on a square grid. The results are listed in Table 1. The objective of this section is to provide preliminary baseline comparisons between the two schemes and a deeper study will be the subject of future work. Nevertheless, there are fundamental differences between the two approaches. Decentralized detection attempts to achieve good false alarm and detection performance averaged across all realizations of sensor observations,

Scheme	Energy (Joules/Node)
Decentralized Broadcast	$\mathcal{O}(N^{\alpha/2} d_0^{\alpha} E_b)$
Decentralized Multihop	$\mathcal{O}(\sqrt{N} d_0^{\alpha} E_b)$
Belief Propagation (grid)	$\mathcal{O}(N \log N d_0^{\alpha} E_b)$
Hierarchical BP	$\mathcal{O}(\log(N) d_0^{\alpha} E_b)$

Table 1: Energy scaling for different schemes; d_0 is the internode distance; E_b is the energy required to transmit 1-bit over a unit distance; α is the attenuation coefficient for transmit energy in free space

while distributed approach attempts to achieve centralized MAP consensus for every observed realization.

The setup of the problem is as follows. N nodes (vertices) of the graph are located at coordinates $(jd_0,\, kd_0)$, $j, k = -\lfloor\sqrt{N}/2\rfloor, \lfloor\sqrt{N}/2\rfloor - 1, \ldots,\, -1,\, 0,\, 1,\, \ldots,\, \lfloor\sqrt{N}/2\rfloor - 1$, on a uniform square planar grid. The edges of the graph are formed by connecting any two nodes that are at a distance d_0 from each other. The transmit energy is assumed to attenuate with distance d as $d^{-\alpha}$. This implies that if E_b is the joules/bit required for reliable decoding over a unit distance then $E_b d^{\alpha}$ is the corresponding energy required for distance d.

Decentralized Case: We analyze decentralized detection for point-to-point and multi-hop communication schemes with the fusion center located at the center of the square area. In the former scheme each node communicates its local decision directly to the fusion center. In the second scheme each node relays its local decision to a neighboring node, which in turn forwards that information in the direction of the fusion center. For the point-to-point scheme, the average energy consumption/node, $E_{ave} = \frac{E_b}{N} \sum_k d_v^4$, where, d_v is the distance from the vth node to the fusion center. A lower bound can be computed by observing that the set of nodes $(jd_0, kd_0),\ -\lfloor\sqrt{N}/2\rfloor \le k \le \lfloor\sqrt{N}/2\rfloor$ for each fixed value of j contains $\lfloor\sqrt{N}\rfloor$ nodes at a distance no less than $|j|d_0$. Therefore,

$$E_{ave} = \frac{E_b}{N} \sum_{v \in V} d_v^4 \ge \frac{E_b}{N} \sum_{k=-\lfloor\sqrt{N}/2\rfloor}^{\lfloor\sqrt{N}/2\rfloor} \lfloor\sqrt{N}\rfloor (kd_0)^{\alpha} \equiv \mathcal{O}(N^{\alpha/2} E_b d_0^{\alpha})$$

For the multi-hop scheme we appeal to the max-flow min-cut theorem. Consider a cut that separates the set of nodes into S, $\bar{S} = V - S$ with $S = (jd_0, kd_0),\ j \ge j_0 > 0,\ -\lfloor\sqrt{N}/2\rfloor \le k \le \lfloor\sqrt{N}/2\rfloor$. The number of bits passing the cut from S towards the fusion center located in $\bar{S}$ is equal to the number of nodes in the cut set, which is equal to $(\lfloor\sqrt{N}/2\rfloor - j_0)\sqrt{N}$. Now this traffic must be supported by $\lfloor\sqrt{N}\rfloor$ nodes that are located at the boundary of the cut set and the total energy required is $E_b(\lfloor\sqrt{N}/2\rfloor - j_0)\sqrt{N} d_0^{\alpha}$. Summing this energy over all

j_0 and normalizing with respect to the number of nodes gives us:

$$E_{ave} = \frac{E_b}{N} \sum_{j_0=1}^{\sqrt{N}} \lfloor \sqrt{N} \rfloor (\lfloor \sqrt{N}/2 \rfloor - j_0) d_0^\alpha \equiv \mathcal{O}(\sqrt{N} E_b d_0^\alpha)$$

Distributed Case: Our task here is to determine the amount of energy expended per node in realizing a MAP consensus. First, the results of Section 6 indicate that the distinction between messages and bits is unnecessary since each message corresponds to a constant number of bits (irrespective of the number of nodes). Therefore, we are left to determine the number of messages transmitted per node to achieve a MAP consensus. We derive an upper bound by drawing upon results from mixing rate of markov chains. First, from [8], it follows that for a stochastic matrix $\tilde{A} = (I + A)D$ with A being a symmetric incidence matrix (as in Section 5):

$$\|x^T \tilde{A}^k - \mathbf{1}^T\|_\infty \leq \epsilon, \ \forall \ k \geq \frac{1}{1 - |\lambda_2|} \log \frac{N}{\epsilon}$$

where x is any positive vector that sums to unity and λ_2 is the second largest eigenvalue in magnitude. We are left to derive an upper bound for the second eigenvalue. For the uniform square grid these bounds follow from well-known results from graph theory [8, 19]. First, the square grid graph can be decomposed in to a product[3], $G_1 \times G_2$ of line graphs G_1 and G_2. Next, eigenvalues of the adjacency matrix of the grid graph are the sum of the eigenvalues of G_1 and G_2 respectively. For our situation the graph G_1 (and G_2) correspond to line graphs, i.e., vertices at locations, $kd_0, \ -\lfloor \sqrt{N}/2 \rfloor \leq k \leq \lfloor \sqrt{N}/2 \rfloor$ and edges connecting any two nodes at a distance d_0. The eigenvalues of the line graph are $(1 - \cos(\pi k/\sqrt{N-1})), \ k = 0, 1, 2, \ldots, \sqrt{N-1}$. Consequently, the second eigenvalue of the (normalized) adjacency matrix, $\tilde{A}$, approaches 1 as $1 - C/N$, where C is some constant. This implies that the energy budget required for achieving the MAP-consensus scales as $\mathcal{O}(N \log N d_0^\alpha E_b)$ Joules/node.

Hierarchical: The above result forms the basis for the hierarchical scheme. In this scheme sensor nodes are organized hierarchically.In the lowest layer small clusters of sensors achieve their respective consensus within the cluster. A cluster head is selected within each cluster and subsets of cluster heads form higher layer networks and achieve consensus. Next, a cluster head for each higher level network can form the next layer and so on. We derive the average energy requirements for such a scheme. Suppose C identical clusters each of size L achieve consensus through local message passing with each message expending energy E. The total energy, E_{tot} is

$$E_{tot} = \#clusters \times \#nodes/cluster \times \#messages/cluster \times Energy/message = (C)(L)(L \log L)(E)$$

[3] (v, w_1) and (v, w_2), (v_1, w) and (v_2, w) are edges in the product graph if $w_1, \ w_2$ and $v_1, \ v_2$ are adjacent in G_2 and G_1 respectively

The average energy, $E_{ave} = E_{tot}/\#nodes = (L\log L)(E)$. To apply this idea to the hierarchical setup we proceed as follows: The first layer is partitioned into $C_1 = N/L$ uniform square clusters of L nodes. In general the kth layer contains $C_k = C_{k-1}/L$ clusters of L nodes. For N nodes there are a maximum of $k_{max} = \log N/\log L$ hierarchical layers. Now, suppose E_k is the energy required per message transmission at the kth layer then the total energy, E_{tot}^k expended upto the kth layer is:

$$E_{tot}^k = E_{tot}^{k-1} + (C_k)(L)(L\log L)(E_k),\ 1 \leq k \leq k_{max}$$

We are left to compute the energy/message E_k expended at the kth layer. To compute this we note that if the inter-node distance is d_0 for the first layer, then it is $\sqrt{L}d_0$ in the second layer. In general, for the kth layer the inter-node distance, $d_k = \sqrt{L}d_{k-1}$. We next consider two possible scenarios: (A) Known cluster head locations: Here intermediate nodes between any two cluster heads serve as relays and forward messages from one cluster head to the other; (B) Unknown cluster head locations: Here the intermediate nodes still serve as relays but diffuse messages to the next cluster head, which then updates the message and retransmits. In scenario (A) the energy expended/node/message at the kth layer is $E_k = \sqrt{L}E_{k-1}$ since $\sqrt{L}$ multi-hops at the $k-1$th layer are required to complete one kth layer message transmission. For scenario (B) the corresponding relation is $E_k = LE_{k-1}$, which follows from the fact that the number of hops to reach the nearest kth layer clusterhead takes $\sqrt{L}$ hops of (k-1)th layer. In this time a maximum of L nodes at the (k-1)th layer have been used as relays for each kth layer node. Substituting these facts and noting that $E_1 = E_b d_0^\alpha$ we obtain for scenario (A) that $E_{ave} = \mathcal{O}(L\log L E_b d^\alpha)$ Joules/node, while for scenario (B) $E_{ave} = \mathcal{O}(LE_b d_0^\alpha \log N)$ Joules/node.

8 Continuous Parameter Estimation

The techniques of Section 6.2 can be extended to continuous parameter estimation via standard approximation techniques. Namely, let Z be a continuous random variable on $\mathbf{R}^n$ for some integer n, and let the observations $(Y_v : v \in V)$ be conditionally independent given Z. If Z has finite mean, then given $\epsilon > 0$ there exists a positive integer $M(\epsilon)$, a finite partition $\{R_i(\epsilon) : i = 1, 2, \cdots, M(\epsilon)\}$ of $\mathbf{R}^n$, and constants $\{\beta_i(\epsilon) : i = 1, 2, \cdots, M(\epsilon)\}$ such that

$$\|Z - Z_\epsilon\| < \epsilon; \text{ where } Z_\epsilon = \sum_{i=1}^{M(\epsilon)} \beta_i(\epsilon)\mathbf{1}\{Z \in R_i(\epsilon)\}.$$

When each event $\{Z \in R_i(\epsilon)\}$ is interpreted as a separate hypothesis H_i, the distributed algorithm of Section 6.2 can be employed to identify the event with the largest posterior probability, and hence the (centralized) MAP estimate of Z can be approximated up to a desired accuracy.

It is worthwhile to consider Gaussian estimation problems in more detail, since distributed algorithms that entail no approximation errors can be identified for such cases. Towards that end, suppose further that Z and $(Y_v : v \in V)$ are jointly Gaussian, and define, for each $v \in V$, the locally computable quantities

$$\mu_v = E[Z|Y_v = y_v], \quad \sigma_v = E[(Z - \mu_v)^2|Y_v = y_v], \quad \xi_v = \sigma_v^{-1}\mu_v.$$

Note that the conditional independence assumption implies that the centralized MAP estimate of Z has the form

$$E[Z|Y_v = y_v : v \in V] = \left(\sum_{v' \in V} \sigma_{v'}^{-1}\right)^{-1} \sum_{v' \in V} \xi_{v'}.$$

Consider now a message passing algorithm that involves two types of messages represented by the two decoupled linear systems such that

$$\begin{aligned} x_k &= (I + A_k)D_k x_{k-1}, \quad x_0(v) = \xi_v : v \in V \\ s_k &= (I + A_k)D_k s_{k-1}, \quad s_0(v) = \sigma_v^{-1} : v \in V, \end{aligned}$$

where $(I + A_k)D_k$ is as defined in Section 6.2. Theorem 6.2 now implies that

$$\lim_{k \to \infty} \frac{x_k(v')}{s_k(v')} = E[Z|Y_v = y_v : v \in V]$$

for each sensor node $v' \in V$.

9 Conclusion

We have considered the scenario of N distributed noisy sensors observing a single event. The sensors are distributed and can only exchange messages through a network. The sensor network is modelled by means of a graph, which captures the connectivity of different sensor nodes in the network. The task is to arrive at a consensus about the event after exchanging such messages. The paper focuses on characterizing the fundamental conditions required to reach a consensus. The novelty of the paper lies in applying belief propagation as a message passing strategy to solve a distributed hypothesis testing problem for a pre-specified network connectivity. We show that the message evolution can be re-formulated as the evolution of a linear dynamical system, which is primarily characterized by network connectivity. Next a family of modified algorithms are considered. These algorithms converge to a MAP-consensus irrespective of graph topology and are robust to random link failures and finite link capacities. Energy scaling laws are then derived, which compare favorably with respect to conventional decentralized detection schemes. Finally a natural extension to distributed estimation is also presented.

10 Appendix

Proof of Lemma 4.2

Consider the region $\mathcal{Z}$ of unit area in which N nodes are uniformly distributed with edges between any two nodes if Equation 21 is satisfied. Suppose $\mathcal{C}_j$ is a circle of radius $R = \frac{\alpha(\log N)^{3/2}}{\sqrt{N}}$ around the node j. The edge connectivity matrix as defined in Equation 8 is denoted by A_N and the accompanying graph by $\Gamma(A_N)$. We are interested in the asymptotic properties as $N \to \infty$.

Proof that number of links approaches a constant: We introduce the random variable $X_k^j \in \{0,\, 1\}$ to indicate whether or not node, k, is within the radius R of node j. The sum $S_N = \sum_{k=1}^N X_k^j$ is the total number of nodes that are linked to node j. It follows from the uniform distribution that, $p = \text{Prob}\{X_k^j = 1\} = \text{Vol}(\mathcal{C}_j) = 4\pi \log^3 N/N$. Therefore, $E(\sum_{k=1}^N X_k^j) = pN$. It follows from Chernoff bound that,

$$\text{Prob}\left\{\left|\sum_{k=1}^N X_k^j - pN\right| \geq \epsilon pN\right\} \leq e^{-\epsilon^2 pN/3} \Longrightarrow \text{Prob}\left\{\left|\sum_{k=1}^N X_k^j - pN\right| \geq \frac{pN}{\log N}\right\} \leq \frac{1}{N^{\alpha^2}}$$

where we have chosen $\epsilon = 1/\log N$ in the latter expression. We can repeat this argument for N nodes in the network.

$$\text{Prob}\left\{\max_{1\leq j\leq N}\left|\sum_{k=1}^N X_k^j - pN\right| \geq \frac{pN}{\log N}\right\} \leq \frac{1}{N^{\alpha^2-1}}$$

The upper bound converges for $\alpha^2 > 2$. By a direct application of Borel-Cantelli lemma it follows that,

$$\frac{1}{pN}\sum_{k=1}^N X_k^j \longrightarrow 1, \text{ almost surely} \Longrightarrow \#\text{links/node} = \log^3 N\left(1 + \frac{C}{\log N}\right),\ |C| \leq 1, \text{ w.h.p} \tag{33}$$

Proof of primitivity: This follows from Theorem [8.5.3] in [11], which states the following: Suppose A is an irreducible and non-negative matrix associated with the directed graph, $\Gamma(A)$. Let $L_j = [k_1^j,\, k_2^j, \ldots]$ be the set of all path lengths that start at node j and end at j. The matrix, A, is primitive if the greatest common divisor of path lengths is equal to one for every j. Irreducibility can be established through strong connectedness (see Theorem [6.2.24] in [11]) of the induced graph. Strong connectedness requires that for any pair of nodes, $i,\, j$ there is a directed path of finite length. To establish strong connectedness of the induced graph $\Gamma(A)$ we let $e_1 = (s_1,\, d_1)$, $e_m = (s_m,\, d_m)$ be any two edges. Since the underlying graph G is connected, it follows that there is a directed path from node d_1 to node d_m. Suppose, this path contains a directed cycle on $\Gamma(A)$, i.e., the path contains the sequence of edges $e,\, e'$, which form a directed cycle. If $e \neq e_1$ it is always possible to obtain a modified path that does not include this cycle (simply delete e' from the path). If not, consider circles, $\mathcal{C}_1,\, \mathcal{C}_2$ of radii $R/2,\, R$ centered around nodes $s_1,\, d_1$ respectively. Consider any node, j, other than s_1 in the intersection of these circles (which exist with high probability). Replace e' by the directed edge (d_1, j) and augment with the directed edge (j, s_1). The new path now formed is a feasible directed path and establishes strong connectedness and therefore irreducibility. Primitivity follows from the fact that the intersection of circles contain multiple nodes with high probability. Therefore, paths of even and odd lengths exist with high probability.

References

[1] M. Alanyali, S. Venkatesh, O. Savas, and S. Aeron. Distributed bayesian hypothesis testing in sensor networks. In *American Control Conference*, Boston, MA, July 2004.

[2] S. Appadwedula, V. V. Veeravalli, and D. L. Jones. Robust and locally-optimum decentralized detection with censoring sensors. In *5th Int. Conf. on Information Fusion*, 2002.

[3] T. Berger, Z. Zhang, and H. Vishwanathan. The ceo problem. *IEEE Transactions on Information Theory*, May 1996.

[4] V. Borkar and P. Varaiya. Asymptotic agreement in distributed estimation. *IEEE Transactions on Automatic Control*, 27, 1982.

[5] J. F. Chamberland and V. V. Veeravalli. Decentralized detection in sensor networks. *IEEE Transactions on Signal Processing*, 2003.

[6] J.F. Chamberland and V. Veeravalli. How dense should a sensor network be for detection applications? *preprint*, page http://tesla.csl.uiuc.edu/ vvv/papers/journal/dependent.pdf.

[7] J.F. Chamberland and V. Veeravalli. Asymptotic results for decentralized detection in power constrained wireless sensor networks. *IEEE Special Issue on Wireless Sensor Networks*, pages 1007–1015, August 2004.

[8] F. R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

[9] Editor F. A. Sadjadi. *Selected Papers on Sensor and Data Fusion*. SPIE Milestone Series, 1996.

[10] A. O. Hero and D. Blatt. Sensor network source localization via projection onto convex sets(pocs). In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, PA, March 2005.

[11] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[12] T. Kasetkasem and P. K. Varshney. Communication structure planning for multisensor detection systems. *IEE Proc. Radar, Sonar Navig.*, 48:2–8, 2001.

[13] M. Longo, T. D. Lookabaugh, and Robert M. Gray. Quantization for decentralized hypothesis testing under communication constraints. *IEEE Trans. on Inform. Theory*, 36:241–255, 1990.

[14] N. Patwari and A. Hero. Reducing transmissions from wireless sensors in distributed detection networks using hierarchical censoring. In *ICASSP*. IEEE, 2003.

[15] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Interference*. Morgan-Kauffman, 1988.

[16] M. G. Rabbat and R. D. Nowak. Distributed optimization in sensor networks. In *Third International Symposium on Information Processing in Sensor Networks*, Berkeley, CA, April 2004.

[17] C. Rago, P. Willett, and Y. Bar-Shalom. Censoring sensors: a low-communication-rate scheme for distributed detection. *IEEE Trans. Aerosp. Electron. Syst.*, 32:554–568, 1996.

[18] L. Santalo. *Integral Geometry and Geometric Probability*. Cambridge University Press, 2004.

[19] D. Spielman. Applied extremal combinatorics. *MIT 18.325 Lecture Notes*, Spring 1996.

[20] P.F. Swaszek and P. Willett. Parley as an approach to distributed detection. *IEEE Transactions on Aerospace and Electronic Systems*, pages 447–457, January 1995.

[21] J. N. Tsitsiklis. Decentralized detection. *in Advances in Statistical Signal Processing, H. V. Poor and J. B. Thomas Eds*, 2, 1993.

[22] P. K. Varshney. *Distributed Detection and Data Fusion*. Springer, 1997.

[23] S. Venkatesh and M. Alanyali. M-ary hypothesis testing in sensor networks. In *Conference on Information Sciences and Systems*, Princeton, NJ, February 2004.

[24] S. Venkatesh, M. Alanyali, O. Savas, and S. Aeron. Classification in sensor networks. In *International Symposium on Information Theory*, Chicago, IL, June 2004.

[25] J. Wolfowitz. Products of indecomposable, aperiodic, stochastic matrices. *Proceedings of the American Mathematical Society*, 15:733–737, August 1963.

[26] J. S. Yedidida, W. T. Freeman, and Y. Weiss. Generalized belief propagation. *NIPS 13*, 2001.