# Asynchronous Distributed Detection in Sensor Networks *

Venkatesh Saligrama   Murat Alanyali   Onur Savas †

**Abstract**

We consider the problem of classifying among a set of $M$ hypothesis with $N$ distributed noisy sensors. The $N$ sensors can collaborate over a finite link-capacity network. The task is to arrive at a consensus about the event after exchanging such messages. In contrast to the conventional decentralized detection approach, wherein the bit rates for each link is explicitly constrained, our approach is based on a high-rate limit perspective. We apply a variant of belief propagation—to account for finite link-capacity—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 finite bit rates supportable between links. We then demonstrate robustness against packet losses, which implies that optimal decisions can achieved with asynchronous transmissions as well. Next, we present energy scaling laws for distributed detection and demonstrate significant improvement over conventional decentralized detection. Finally, extensions to distributed estimation are described.

## 1 Introduction

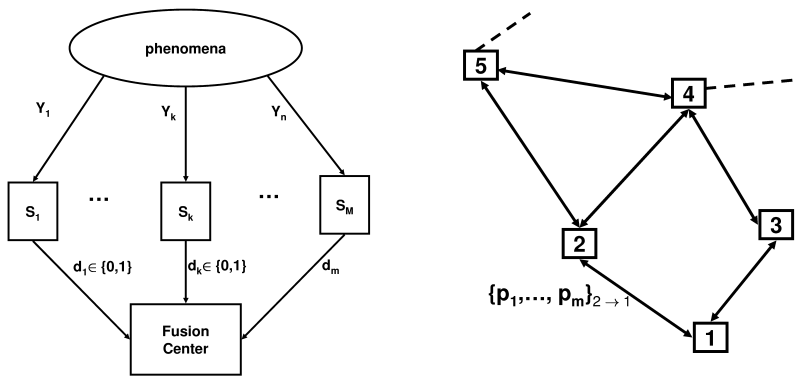
Recent advances in sensor and computing technologies provide impetus for deploying wireless sensor networks—a network of massively distributed tiny devices capable of sensing, processing and exchanging data over a wireless medium.

*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.

In this paper we focus on the scenario of $N$ distributed noisy sensors observing a single phenomena. The sensors can only collaborate through a network defined by a connectivity graph. The task is to arrive at a consensus about the event after exchanging such messages. Fundamental problems arise when data is distributed and centralized solutions are no longer feasible due to time/rate/energy constraints.

The general question of dealing with distributed data in the context of detection has been an active topic of research(see [1–3, 6, 7, 10–13, 16, 19, 23–26] and references therein). Much of this research focuses on a fusion centric approach with $N$ sensors having communication links to a data fusion center as shown in Fig. 1(a). Here, the research is focused on capacity constrained networks. Research has addressed quantization of sensor




(a) Decentralized Detection (b) Distributed Detection

Figure 1: Various schemes for detection in sensor networks.

data [11] and exploiting source correlation [15] to reduce sensor bit rate. In particular cases, it has been shown that for a N-sensor network with a capacity constraint of N bits per unit time, having each sensor send one bit is optimal [3]. In general, the data from each sensor is compressed on to a message taking values over a finite alphabet. The objective is to find optimal fusion rule and the associated decision rules at each of the individual sensors to minimize the error probability. The principle drawbacks of the approach are well-known and has been documented in [23, 24]. We point out some of these here for the sake of exposition: (**A**) It can be shown that (if the hypothesis were conditionally independent) the decision rules at each sensor reduce to a likelihood ratio threshold test. Nevertheless, the decision rules are coupled in that thresholds have to be solved jointly for all the sensors. This not only has computational implications but also assumes centralized knowledge of the sensor models. (**B**) The network topologies for which these results hold are restrictive. (**C**) Considering all energy costs in an energy budget as in [18] shows that sending one bit of data consumes only marginally less energy than sending many bits. It has been argued [2, 10, 13, 19] that in energy-limited wireless sensor networks a more appropriate goal is to minimize the probability that sensors must transmit. Nevertheless, a hierarchical structure with a single fusion center is still used. (**D**) Moreover, energy for communicating over large distances is significantly more than that required to communicate over small distances. Finally, a fusion centric approach has the disadvantage of a single point of failure.

To overcome these issues we develop a data-centric as opposed to decision-centric communication strategy. Related work for distributed optimization in sensor networks have been proposed recently in [8, 17] for specific types of network topologies. In this paper 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 [14] and preliminary results on their application to the detection problem is described in a number of our papers [1, 25, 26]. A description is shown in Figure 1(b) where sensor nodes send a vector of likelihoods for each hypothesis at any instant of time. These likelihoods can be dynamically updated based on information received by the sensor in the past. Evidently, the algorithm overcomes the centralization issue alluded to earlier. 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, a situation that is quite restrictive and difficult to impose in a sensor network. 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 4.2 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 prove that this consensus is not only the MAP estimate but also that the exact posterior distribution can be realized. We next deal with the finite link capacity in Section 5 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. Next we show that our algorithms are robust to random packet losses in Section 5.2. In Section 6 energy scaling laws in wireless environments for uniform grid as well as random networks are derived. The results show exponential improvement in energy scaling over the conventional fusion center approach. Finally, in Section 7 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. **(D)** There is no single point of failure as is the case for the fusion center approach.

## 2 Setup

We consider MAP estimation in M-ary hypothesis testing with hypotheses $\mathcal{H} = \{H_1, H_2, \cdots, H_M\}$ and prior distribution $\pi_o(\cdot)$. Estimation is based on a random vector $Y = (Y_v : v \in V)$ of observations that belong to an arbitrary abstract space. Throughout the paper $V$ is interpreted as a set of sensors, and $y_v$ is interpreted as realization of the measurement taken by sensor $v \in V$. Let $f_m$ be the conditional probability density function of $Y$ under hypothesis $H_m$ for $m = 1, 2, \cdots, M$. We shall assume that observations are conditionally independent given the true hypothesis; namely,

$$f_m(y) = \prod_{v \in V} f_m^v(y_v), \quad y = (y_v : v \in V) \in \mathbf{R}^V \tag{1}$$

for marginal densities $f_m^v$. Let $\pi$ denote the posterior distribution of the true hypothesis given that $Y_v = y_v$ for $v \in V$, which is identified uniquely by the relation

$$\pi(H_m) \propto \pi_o(H_m) \prod_{v \in V} f_m^v(y_v), \quad m = 1, 2, \cdots, M. \tag{2}$$

In particular $H_{m^*}$ is a MAP estimate if

$$m^* \in \arg\max_m \left\{ \pi_o(H_m) \prod_{v \in V} f_m^v(y_v) \right\}.$$

We concentrate on distributed applications in which a single decision maker having access to all observations $(Y_v : v \in V)$ is not available. Instead, each sensor can collaborate with other sensors and form an estimate of the posterior distribution. The collaboration is limited by a communication network structure represented by a weighted, strongly connected digraph $G = (V, E)$. The vertices $V$ of this graph 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 $v$. Sensor $v'$ is referred to as a *neighbor* of sensor $v$ if $(v', v) \in E$. Let $N(v)$ denote the set of neighbors of sensor $v$ so that

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

**Relationship to Markov Random Fields** In motivating the collaborative framework of the paper, it will prove useful to interpret the posterior distribution $\pi$ in terms of Markov random fields (MRFs). A random vector $X = (X_v : v \in V)$ is a MRF with respect to an undirected graph $\tilde{G} = (V, \tilde{E})$ if its marginals admit certain consistency conditions defined relative to the neighborhood relations of $\tilde{G}$ [14, 28]. In particular, if all combinations of possible marginals have positive probability, then by the Hammersley-Clifford theorem there exists positive mappings $\phi_v : v \in V$, and $\psi_e : e \in \tilde{E}$ such that

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

for each realization $(x_v : v \in V)$ of $X$. In broad terms, node potentials $\phi_v$ account for likeliness of marginal values whereas edge potentials $\psi_e$ account for pairwise correlations in $X$. Let $\tilde{G}$ be an arbitrary connected graph spanning the nodes $V$, and consider the distribution (3) of $X$ with

$$\phi_v(H_m) = (\pi_o(H_m))^{1/V} f_m^v(y_v), \quad v \in V, \tag{4}$$

$$\psi_e(H_j, H_m) = \delta(H_j, H_m), \quad e \in E, \tag{5}$$

where $\delta(\cdot, \cdot)$ is the standard Kronecker delta function. It is easy to verify that due to connectivity of $\tilde{G}$ marginal distributions of $X$ are identical, and furthermore they equal to $\pi$. Although most combinations of marginal values of $X$ have zero probability, this issue will not lead to complications in applying efficient algorithms that compute marginal distributions of MRFs (such as belief propagation [14, 28]) in order to arrive at estimates of the underlying hypothesis based on $\pi$ here. Furthermore, in the context of the detection problem we have enormous flexibility in choosing the edges. Specifically, any arbitrary graph, i.e., arbitrary choice of edges associated with edge potentials as in Equation 5 accounts for the informational aspects of the problem. *This implies that the MRF model, $\tilde{G} = (V, \tilde{E})$, can be chosen to coincide with the communication network graph, $G = (V, E)$. This justifies the use of BP as a message passing algorithm*

# 3 Collaborative Framework

The so called BP algorithm [14, 28] provides a framework for collaboration between nodes of a graph to compute marginal distributions of a MRF. The main idea from a detection perspective is that each sensor node, $v$, transmits a vector whose $m$th component is related to a local estimate for hypothesis $H_m$ at node $v$. This overcomes the centralization issue underlying conventional decentralized detection, where decisions are transmitted. In the mechanics of the algorithm, at time step $k$ each sensor node $v' \in V$ forwards a message, $m_k^{(v',v)}(h), h \in \mathcal{H}$, to sensor node $v$ via the communication infrastructure represented by the digraph $G$. More specifically, sensor node

$v'$ computes the product of most recently received messages pertaining to each hypothesis $h'$ (excluding message from $v$), and averages this product across all hypothesis with adequate weighing to reflect correlations between the hypothesis $h$ and $h'$. On account of the specific potentials (4)–(5) these messages are given by

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

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

along any edge $(v, v') \in E$, for each hypothesis $h \in \mathcal{H}$, and round $k \geq 0$. Messages are used by recipient nodes to compile their *beliefs*, which are estimates for the posterior distribution $\pi$ defined as follows:

**Definition 3.1** (Belief) The belief $(\hat{\pi}_k^v(h) : h \in \mathcal{H})$ of node $v \in V$ at round $k$ is a probability vector uniquely identified by the relation

$$\hat{\pi}_k^v(h) \propto \pi_o(h)\phi_v(h) \prod_{v'\in N(v)} m_k^{(v',v)}(h). \tag{8}$$

¿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) Ccomposition of the messages 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 $k$th messages only after receiving $(k-1)$th messages from all of its neighbors.

If $G$ is a singly-connected graph then well-known results [28] on Pearl's sum-product algorithm guarantee that each belief $\hat{\pi}_k^v,\ v \in V$, converges to the true posterior distribution $\pi$ 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 structures endowed by the classification problem. To explore this strategy we first transform the original problem into a linear dynamical system.

We identify each edge $e \in E$ by its source vertex $s(e)$ and its destination vertex $d(e)$ so that $e = (s(e), d(e))$. Therefore, each edge, $e$, can be associated with neighboring edges, $I_e$, incident on it and the set of edges, $O_e$, that it is incident on, i.e.,

$$I_e = \{e' \in E \;\;|\;\; d(e') = s(e)\} \tag{9}$$

$$O_e = \{e' \in E \;\;|\;\; s(e') = d(e)\}$$

For each pair of edges $e, e' \in E$ let

$$a_{e,e'} = \delta(s(e), d(e'))(1 - \delta(s(e'), d(e))). \tag{10}$$

Note that $a_{e,e'} = 1$ if and only if edge $e' \in I_e$ and the ordered pair $(e', e)$ is not a directed cycle. For each hypothesis $h \in \mathcal{H}$ let

$$\begin{aligned} u^h(v) &= \log(\phi_v(h)), \quad v \in V \\ x_k^h(e) &= \log(m_k^e(h)), \quad e \in E. \end{aligned}$$

Taking the logarithm of both sides in equalities (7) leads to the linear system

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

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

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

We note that the dynamical evolution in Equation 12 depends only on the graphical structure and not on the individual observations. This key insight as we will see in the next section results in consensus among different sensors based primarily on the network topology. Finally, we will show how to achieve the correct MAP estimates for arbitrary connected graphs.

## 4 Consensus and Convergence

In this section we derive results for reaching a consensus based on our analysis in the previous section. We lift the restriction on link capacity limits and discuss decentralized determination of MAP estimates. In a subsequent section we will discuss methods for achieving a MAP consensus with finite-link capacities and lossy links. We start with a formal definition of consensus in the present context.

**Definition 4.1** (MAP Consensus) Sensor node $v \in V$ eventually succeeds in MAP estimation if

$$\limsup_{k \to \infty} \hat{\pi}_k^v(h) = 0 \tag{13}$$

for all $h \in \mathcal{H}$ such that $h \notin \arg\max_{h'} \pi(h')$. The sensor network is said to asymptotically achieve a MAP-consensus if each sensor eventually succeeds in MAP estimation. Sensor network is said to achieve a consensus if Equation 13 holds for the same subset of hypotheses $h \in \mathcal{H}$, which are not necessarily MAP estimates, at all nodes $v \in V$.

We point out that the notion of consensus is substantially weaker than the conventional objective of estimating the distribution. In Section 4.2 we define a stronger notion of convergence while discussing modifications to the BP algorithm. The weaker notion is useful when we are only interested in achieving the MAP decision rule and as it turns out the message passing algorithm can guarantee a MAP consensus for particular graph topologies. In the following we will state results for different graphical structures and provide main outlines for the proof for these results. First, note that the solution to the linear system satisfies:

$$x_k^h = \sum_{j=0}^{k-1} A^j u^h; \;\; k \geq 1 \tag{14}$$

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

**Lemma 4.1** *The matrix $A^j = [a^j_{e,e'}]_{E \times E}$ where $a^j_{e,e'}$ is the number of directed paths of length $j$ edges that start with edge $e'$, end with edge $e$, and that do not have any 2-hop cycles.*

Based on the above discussions we have the following results. The first theorem concerns the case when the graph $G$ is a tree and the subsequent result deals with ring graphs. It is well-known that in this case BP leads to the true posterior distributions even for general Markov fields. However, the proof in our context makes use of the special structure of the transition matrix $A$ defined in Equation 10.

**Theorem 4.1** *(Trees) If $G$ is a tree then the network asymptotically achieves MAP-consensus with BP algorithm.*

**Proof.** If $G$ is a tree, then $A$ defined as in Equation 10 is nilpotent since $A^j = 0$ for all integers $j$ larger than the diameter of the tree. Equality (14) therefore indicates that the messages are guaranteed to converge within a number of steps no larger than the diameter. Note that for $e, e' \in E$

$$\sum_{j=0}^{\infty} a^j_{e,e'} = \begin{cases} 1 & \text{if there exists a simple directed path} \\ & \text{in } G \text{ with source edge } e' \text{ and destination edge } e \\ 0 & \text{else,} \end{cases}$$

hence equality (14) leads to

$$\lim_{k \to \infty} x_k^h(e) = \sum_{v \in V} \mathbf{1}\{\text{dist}(v, s(e)) < \text{ dist}(v, d(e))\} u^h(v)$$

for $e \in E$, where $\text{dist}(v, v')$ represents the length of the unique path between vertices $v, v' \in V$. It now follows by equality (8) that the limit of the estimate $\hat{\pi}_k^v(h)$ at each sensor $v \in V$ is equal to the posterior distribution (2).

□

**Theorem 4.2** *(Rings) If $G$ is a ring then the network asymptotically achieves MAP-consensus with BP algorithm.*

**Proof.** If $G$ is a simple cycle then for $e, e' \in E$ the sequence $(a^j_{e,e'} : j = 0, 1, 2 \cdots)$ has period $|V|$. In particular $A^j = A^{j+|V|}$ and thus $A$ is idempotent. Equality (14) then leads to

$$\lim_{k\to\infty} \frac{x^h_k(e)}{k} = \frac{1}{|V|} \sum_{j=0}^{|V|-1} A^j u^h.$$

It is not difficult to see that $\sum_{j=0}^{|V|-1} a^j_{e,e'} = 1$ for all edges $e, e' \in E$ that have a common orientation (that is, clockwise or counter-clockwise) and that $\sum_{j=0}^{|V|-1} a^j_{e,e'} = 0$ otherwise; in turn

$$\lim_{k\to\infty} \frac{x^h_k(e)}{k} = \frac{1}{|V|} \sum_{v\in V} u^h(v), \quad e \in E. \tag{15}$$

Since

$$\hat{\pi}^v_k(h) \ \propto \ \phi_v(h) \exp\left( \sum_{v'\in N(v)} x^h_k(v', v) \right),$$

it follows via (15) that for any two hypotheses $h, h' \in \mathcal{H}$

$$\lim_{k\to\infty} \frac{1}{k} \log \frac{\hat{\pi}^v_k(h)}{\hat{\pi}^v_k(h')} = \frac{2}{|V|} \left( \sum_{v'\in V} u^h(v') - \sum_{v'\in v} u^{h'}(v'). \right)$$

The conclusion of the theorem now follows since

$$\sum_{v'\in V} u^h(v') \ = \ \log \left( \prod_{v'\in V} \phi_{v'}(h) \right) \ = \ \log(\pi(h)), \quad h \in \mathcal{H},$$

by definitions (2) and (4). □

Although consensus is achieved in the sense of Definition 4.1, convergence of beliefs may not emerge. This happens only when a unique MAP estimate does not exist, and in this case mode of the belief at each sensor may oscillate among maximizers of the posterior distribution $\pi$ (see [25] for an example).

## 4.1 Regular Graphs

For general graphs with arbitrary cycles it turns out that the dynamical evolution does not converge to the correct likelihood ratios. Nevertheless, it turns out that for d-regular graphs and random graphs MAP consensus can be guaranteed. First, we need the notion of a primitive matrix.

**Definition 4.2** A matrix $A$ is said to be primitive if $A^m > 0$ for some positive integer $m$.

We now have the following theorem:

**Theorem 4.3** *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 10 be primitive. Then a MAP consensus is achieved with the BP algorithm.*

**Proof.** Since, $d > 2$ it follows that the spectral radius, $\rho(A)$, is larger than $1$. From standard results in Perron-Frobenius theory [9] it follows that,

$$\lim_{k\to\infty} (A/\rho(A))^k = W \tag{16}$$

where, $W$, is a rank-one matrix formed from the left and right eigenvector corresponding to the maximal eigenvalue. For simplicity, define $\alpha(k) = \sum_{j=0}^{k} \rho(A)^j$. By equality (14)

$$\lim_{k\to\infty} \frac{x_k^h}{\alpha(k)} = \lim_{k\to\infty} \sum_{j=0}^{k-1} \left(\frac{A^j}{\rho(A)^j}\right) \frac{\rho(A)^j}{\alpha(k)} u^h = \lim_{k\to\infty} \sum_{j=0}^{k-1} (W + \epsilon_j) \frac{\rho(A)^j}{\alpha(k)} u^h,$$

where, in the last equation, we have used Equation 16 to obtain a real-valued matrix sequence, $\epsilon = \{\epsilon_j\}$ that vanishes as $j \to \infty$ and that satisfies $\|\epsilon\|_\infty \le C_0 < \infty$. This implies that for some $l > 0$,

$$\frac{x_k^h}{\alpha(k)} = W u^h + \sum_{j=k-l}^{k-1} \epsilon_j \frac{\rho(A)^j}{\alpha(k)} u^h + \sum_{j=0}^{k-1-l} \epsilon_j \frac{\rho(A)^j}{\alpha(k)} u^h$$

Therefore, we have

$$\left| \frac{x_k^h}{\alpha(k)} - W u^h \right| \le \max_{k-l \le j \le k-1} \|\epsilon_j u^h\|_\infty + C_0 \rho(A)^{-l} \|u^h\|_\infty$$

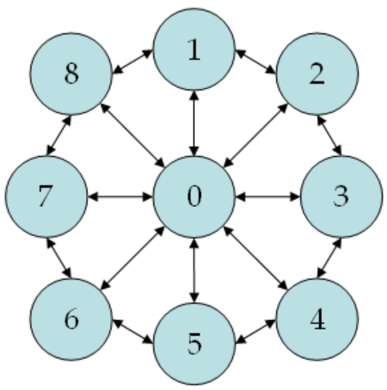
where, for the first term of the RHS we have used the fact that $\rho(A)^j/\alpha(k) \le 1$ and for the second term we have used $\rho(A)^{k-l}/\alpha(k) \le \rho(A)^{-l}$. Consequently, for any $\epsilon > 0$ there exists a sufficiently large number $l$ such that the limiting difference $\limsup_k \left| \frac{x_k^h}{\alpha(k)} - W u^h \right| \le \epsilon$. Now since $\epsilon$ is arbitrary the result follows. To establish the fact that the consensus is a MAP estimate we note that for a regular graph the unique left eigenvector (upto a constant multiplication) for $W$ is a column vector of ones. Therefore, it follows that,

$$\lim_{k\to\infty} \frac{x_k^h(e)}{\alpha(k)} \propto \sum_{v\in V} u^h(v), \quad e \in E. \tag{17}$$

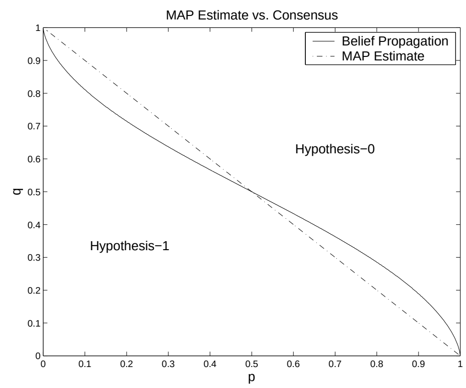
The result now follows along the lines of Theorem 4.2 (from Equation 15). $\square$

The theorem does not hold for general graphs, i.e., sensors do reach a consensus but the estimate is not guaranteed to be a MAP estimate. To see this consider the following example:

**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. The eigenvectors no longer have equal weights corresponding to each edge. Suppose that the observations $(x_v : v \in V)$ translate to node potentials $\phi_0 = [q, 1-q],\ \phi_1 = [p, 1-p]$ and $\phi_v = [0.5, 0.5]$ for $k = 2, 3, 4, 5, 6$, where $p, q \in [0, 1]$. Figure 2(b) illustrates the true MAP estimate and the final consensus due to BP for different values of $p$ and $q$. Note that the consensus is determined to a larger extent by the value of $q$ rather than the value of $p$. Note also that the consensus reflects a flawed estimate if $(p, q)$ lies in the area between the solid and dashed lines.




(a) A 9-node irregular communication graph

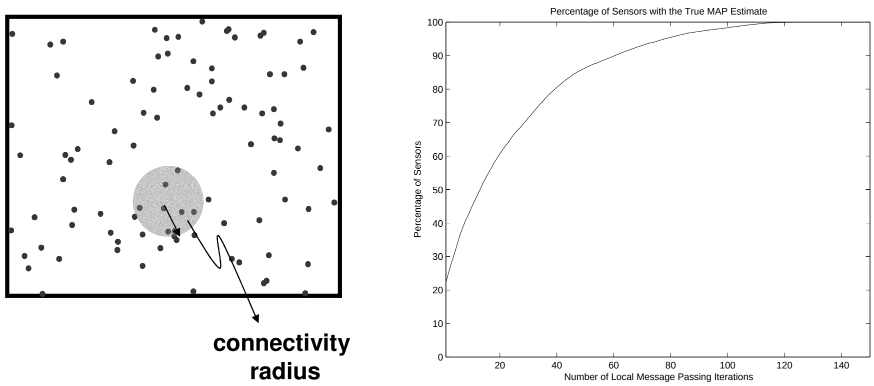



(b) Decision regions for the MAP estimate (dashed) and the final consensus of BP (solid).

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

We now extend these results to random graphs. The random graph is constructed in the following manner. $N$ sensor nodes are uniformly distributed in a square unit area denoted by the region $\mathcal{Z}$ centered at zero as shown in the Fig. 3(a). Unlike deterministic regular graphs discussed earlier, the communication connectivity radius needs to be chosen carefully. This is to ensure that the random graph is still regular with high probability. It is well-known [20] that the minimum communication radius, $R$, is of the order of $\frac{\log N}{\sqrt{N}}$ to ensure graphical connectivity of $N$ uniformly distributed sensor nodes. However, for this minimum radius the degree (i.e. number of neighbors for each node) is highly variable. To ensure a constant degree with high probability we need a slightly larger radius of connectivity, i.e., $R = \frac{(2\log N)^{3/2}}{\sqrt{N}}$. For the sake of mathematical simplicity we consider a periodic extension of the graph so that issues related to dealing with boundary nodes does not arise. For this situation the edges are formed by linking any two vertices that are at a distance smaller than $R$ in the original graph or in the extension. In particular two nodes, whose planar coordinates are respectively, $(a_1, b_1)$, $(a_2,\, b_2)$, are connected by a link if

$$\min\left((a_1 - \tilde{a}_2)^2 + (b_1 - \tilde{b}_2)^2\right)^{1/2} \leq R;\ \tilde{a}_2 = (a_2)\text{mod}(1),\ \tilde{b}_2 = (b_2)\text{mod}(1)) \tag{18}$$




(a) Randomly distributed sensors with constant connectivity radius in an unknown plume.

(b) Detection Probability Rate with 400 Sensors with 10 different Hypothesis.

Figure 3: Classification by Sensor Networks.

**Theorem 4.4** *Consider the random graph setup above. A MAP consensus is achieved by an approximate BP algorithm, which converges to the BP algorithm as $|V| \to \infty$.*

**Proof.** We provide a brief outline here and refer the details to the appendix.

(A) We show that the in-degree and out-degree for each node as defined in Equation 9 is asymptotically equal to $(\log(n))^3$ almost surely. Next the primitivity of the matrix $A$ with high probability is established. This implies that that there is a rank-one matrix $W$ such that, $\lim_{k\to\infty}\left(\frac{A}{\rho(A)}\right)^k = W$.

(B) We consider an approximate BP algorithm, specifically,

$$\tilde{x}_k^h = u^h + A(I+\Delta)^{-1}\tilde{x}_{k-1}^h, \quad \tilde{x}_0^h = 0 \tag{19}$$

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. The approximate message passing scheme corresponds to weighting each message in inverse proportion to its out-degree before transmission.

(C) The weighting does not destroy primitivity and therefore we have

$$\lim_{k\to\infty}\left(\frac{A(I+\Delta)^{-1}}{\rho(A(I+\Delta)^{-1})}\right)^k = \tilde{W}$$

where the left eigenvector of $W$ is a vector of all ones. Now the argument follows along the lines of Theorem 4.3 and for a suitable normalization $\alpha(k)$, we have

$$\lim_{k\to\infty} \frac{\tilde{x}_k^h(e)}{\alpha(k)} \propto \sum_{v\in V} u^h(v), \quad e \in E.$$

for some $C_0 > 0$ with high probability. The MAP consensus can now be established along the lines of Theorem 4.3. Now, as $N \to \infty$ the weighting converges to zero, which is equivalent to the BP algorithm. $\square$

The convergence properties for a typical random graph are illustrated in Figure 3(b). Nevertheless, the drawback of the BP algorithms are fivefold: (a) The estimates for general graphs are not guaranteed to be MAP estimates; (b) Message updates at the various nodes need to follow a certain order; (c) the scheme is not necessarily robust to link failures; (d) BP algorithms require customizing information for a particular node, in that message from node $J$ to node $K$ is a function of messages received from neighbors of $J$ excluding $K$. It would be simpler and energy efficient if one could fuse the received messages at each node and just broadcast them without the need for customization. (e) Equation 17 holds out the possibility that consensus to the exact posterior distribution is achievable at least for regular graphs. Indeed, it follows from Equation 17 that,

$$\frac{1}{\alpha(k)} \log m_k^h(v,v') \propto \sum_{v\in V} \log f_h^v(y) = \log \prod_{v\in V} f_h^v(y) \Longrightarrow \frac{1}{\alpha(k)|N(v)|} \sum_{v'\in N(v)} x_k^h(v',v) \propto \log \prod_{v\in V} f_h^v(y)$$

where, the proportionality does not depend on the hypothesis. The last expression is the normalized sum of log-messages from the neighbors of each node. This differs from the exact MAP log-likelihood function by the log-prior distribution $\log(\pi_o(h))$. This justifies consideration of a modified belief estimator. Specifically, we define the log-belief estimate, $\log \tilde{\pi}_k^v(h)$, as

$$\log \tilde{\pi}_k^v(h) = \log(\pi_o(h)) + \eta(k) \sum_{v'\in N(v)} x_k^h(e') \tag{20}$$

This also motivates the concept of a strong consensus.

**Definition 4.3** (Strong Consensus) Sensor nodes $v \in V$ are said to achieve a strong consensus if the belief estimate, $\tilde{\pi}_k^v(h)$, converges to the same distribution (not necessarily the correct posterior distribution). We say that a sensor node $v \in V$ converges in distribution if,

$$\lim_{k\to\infty} \tilde{\pi}_k^v(h) = \text{Prob}\,\{h|(Y_v : v \in V)\}\ \forall\ h \in \mathcal{H}.$$

Furthermore the sensor network is said to asymptotically converge in distribution if each sensor converges to the same posterior distribution.

## 4.2 Modified BP Algorithms

The preceding arguments lead us to consider convergence and strong consensus for general graphs. Consensus and convergence appear to hinge on whether or not the transition matrix, $A$, is primitive. For connected graphs this can be accomplished by modifying the message passing algorithm as follows:

$$z_k^h = u^h + (I + A)z_{k-1}^h$$

where, we have used the symbol $z_k^h$ as the state in place of $x_k^h$ to avoid confusion with the BP algorithm. It then follows that the matrix $(I + A)$ is primitive and Theorem[8.5.1] of [9] applies and it follows that:

**Proposition 4.1** *Consider a connected graph G=(V,E) and the message passing algorithm given above along with the belief estimate of Equation 20. It follows that the messages converge to a strong consensus, which is equal to a weighted log-likelihood, i.e.,*

$$\log \tilde{\pi}_k^v(h) \longrightarrow \log(\pi_0(h)) + \sum_{v\in V} l_v \log f_h^v(y)$$

*where, the weights, $l_v$, are coefficients of the left eigenvector of $W = \lim_{m\to\infty}(\rho(A)^{-1}A)^m$.*

**Proof.** The proof follows by direct substitution and is omitted. □

The main issue as the above proposition suggests is that the consensus achieved is in general a weighted log-likelihood ratio, i.e., unbalanced graphs lead to incorrect posterior distributions. Therefore, the question arises as to how to balance the messages for an arbitrary graph. We explore two possibilities both of which lead to the desired solution:

**Self Loops:** Here the message passing scheme described by Equation 12 is modified by creating virtual self-loops in the communication graph. The number of self-loops for a particular node is equal to the difference between the maximum degree of the graph and the degree of a particular node, i.e.,

$$z_k^h = u^h + (D + A)z_{k-1}^h,\ z_0^h = 0$$

Alternatively, from standard linear systems theory it follows that the following scheme may be used as well:

$$z_k^h = (D + A)z_{k-1}^h,\ z_0^h = u^h$$

In the above schemes, $D = \text{diag}[d_{ee}]$, is a diagonal matrix with $d_{ee} = d_{max} - d_e$, where $d_e = \sum_{e'\in E} a_{e'e}$ is the in-degree of sensor node $s(e) \in V$, and $d_{\max} = \max\{d_e : e \in E\}$. With this modification it follows that the column sums are all identical. Therefore, the coefficients of the left eigenvector are all equal. Consequently, the messages, $z_k^h$, converge to the correct posterior distribution. The main drawback of this idea is that the maximum degree must be known a priori at each node. The problem can be overcome by choosing $d_{ee} = d^* - d_e$ where $d$ is any number that is known to be larger than $d_{\max}$. An alternative strategy is explored next.

**Normalization:** Here each transmitter node normalizes its message for each hypothesis by the number of different messages it receives, i.e., we have

$$z_k^h = (I + A)Dz_{k-1}^h, \; z_0^h = u^h, \tag{21}$$

where $D = [d_{ee}]$ is a diagonal matrix with $d_{ee} = 1/d_e$, where $d_e$ is again the in-degree of sensor node $s(e)$. With this modification, it follows that the conditions for Theorem[8.5.1] of [9] are satisfied with column sums being identical. Therefore, the coefficients for the left eigenvector are all equal. The advantage of this strategy is that message passing algorithm does not need to know the graph topology. Another aspect of this strategy is that the maximum eigenvalue of the transition matrix is one. Consequently, the messages are all bounded, and potential numerical instabilities are thereby avoided. We will point out a third aspect of this scheme when we deal with link failures in Section 5.2.

**Broadcast:** Each transmitter node broadcasts the same message to all its neighbors. This is different from earlier strategies because in the BP algorithm message from node $J$ to node $K$ is based on messages received from all the neighboring nodes excluding $K$. In contrast in the current scheme, the message update is undirected and combines messages from all the adjacent nodes, which is then normalized as before and broadcast. Evolution of messages admits a linear representation as before, though with reduced dimensionality. That is,

$$z_k^h = (I + A)Dz_{k-1}^h, \; z_0^h = u^h, \tag{22}$$

where $z_k^h = (z_k^h(v) : v \in V)$ (where $(z_k^h(v) : h \in \mathcal{H})$ is the message broadcast by node $v$ at round $k$) and $u^h = (u^h(v) : v \in V)$ are $V$-dimensional column vectors, $A = [a_{v,v'}]_{V \times V}$ such that $a_{v,v'} = \delta((v, v') \in E)$, and $D$ is a diagonal matrix that transforms columns of $A$ into probability vectors. Again, by arguments described above it follows that the beliefs converge to the centralized posterior distribution.

Furthermore, all of these strategies are consistent with the currently employed wireless protocols. In summary we have the following theorem, which is stated without proof:

**Theorem 4.5** *For an arbitrary graph topology the message passing scheme of Equations 22, 20 converge to the centralized posterior distribution at all nodes.*

# 5 Robustness Issues

## 5.1 Detection with Finite-Link Capacity

To deal with finite link capacity we take a robustness viewpoint. In other words, the problem we pose is to determine how quantization of the likelihood vector impacts the MAP estimate. Suppose, each message (note that this is not the log-message) is quantized with a logarithmic quantizer, $Q(\cdot)$ given by,

$$Q(z) = (1+\delta)z, \text{ for some } |\delta| \leq \gamma$$

where, $\gamma$ is the maximum allowable quantization. In the logarithmic domain this translates to:

$$z_k^h = A z_{k-1}^h + \log(1+\delta)\mathbf{1} + u^h, \; x_0^h = u^h$$

where $\mathbf{1}$ is a column vector of all ones. We can assume without loss of generality that $A^T$ is a stochastic matrix based on the results in the previous section. Our task now reduces to quantifying the maximum allowable quantization level, $\gamma$, so that a MAP consensus can be achieved. We then have the following result.

**Theorem 5.1** *The maximum quantization allowable (for sufficiently large number of identical sensors) such that consensus does not deviate from the MAP estimate is given by:*

$$|\gamma|^2 < \min_{h_1 \neq h_2, h_0} \exp\left(E_{h_0}\left(\frac{\log f_{h_1}}{\log f_{h_2}}\right)\right) = \min_{h_1 \neq h_2, h_0} |\exp\left(D(f_{h_0} \| f_{h_1}) - D(f_{h_0} \| f_{h_2})\right)|$$

*where, $D(\cdot\|\cdot)$ is the Kullback-Leibler distance and $E_{h_0}(\cdot)$ is the expectation under hypothesis $h_0$.*

**Proof.** Proceeding along the lines of Theorem 4.3 we note that,

$$\limsup_{k\to\infty} \left| \frac{z_k^h}{k} - \sum_{v \in V} u^h(v) \right| \leq \limsup_{k\to\infty} \left| \sum_{j=0}^{k-1} \frac{A^j}{k} \log(1+\delta_{k-j}) \right| \leq \limsup_{k\to\infty} \max_p \left| \sum_{q=0}^{V} \sum_{j=0}^{k-1} \frac{a_{pq}^j}{k} \right| \log(1+\gamma) \leq |V| \log(1+\gamma)$$

where, the third inequality follows from the fact that if $b(\cdot) \leq b_{max}$ then $|\sum_k a(k) b(t-k)| \leq b_{max} \sum_k |a(k)|$. The last inequality follows from primitivity and stochasticity of $A$. If the log-likelihood of $h$ is higher than $h'$ we require that,

$$\frac{1}{|V|} \sum_v (u^h(v) - u^{h'}(v)) > 0 \Longrightarrow \frac{z_k^h}{k} > \frac{z_k^{h'}}{k}$$

¿From the strong law of large numbers it follows that $\frac{1}{|V|} \sum_v (u^h(v) - u^{h'}(v))$ converges to the differences in the Kullback-Leibler distance [5] under the true hypothesis. This implies that we need

$$\log(1+\gamma) \leq \frac{1}{2} \min_{h_1 \neq h_2, h_0} |D(f_{h_0} \| f_{h_1}) - D(f_{h_0} \| f_{h_2})|$$

Consequently, it follows that if the distributions corresponding to each hypothesis are well-separated, the number of quantization-levels are also a constant. Therefore, log-messages can be discretized into 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). This follows from the results in Section 4, where we showed that the log-messages are bounded.

### 5.2 Packet Losses and Asynchronous Operation

In this section we relax the assumption that a 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. Namely, if each sensor waits for messages from all neighbors to compose their next message, then the network evolves at the time scale of the slowest sensor. 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. Here we address the attendant effects of this generalization in stochastic and deterministic settings.

Consider the broadcast operation of Section 4.2 in the case when the connectivity of the network is time-varying. Namely, evolution of the messages is represented as

$$z_{k+1}^h = (I + A_k) D_k z_k^h, \quad z_0^h = u^h, \tag{23}$$

where $A_k = [a_{ij}(k)]$ is a binary matrix and $D_k = [d_{ij}(k)]$ is a diagonal matrix with

$$d_{jj}(k) = \left(1 + \sum_i a_{ij}(k)\right)^{-1},$$

so that in particular columns of $(I + A_k)D_k$ are probability vectors. Given sensors $i, j$ we shall say that link $i \to j$ is *functional* in round $k$ if sensor $j$ receives a message from sensor $i$ in round $k$. Entries of $A_k$ are then interpreted as

$$a_{ij}(k) = I\{ \text{ link } j \to i \text{ is functional at round } k\}.$$

Hence the system (23) 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 5.2** *Suppose that the matrices* $(A_k : k \geq 1)$ *is are iid, and that* $E[A_1]$ *is irreducible. Then for* $v \in V$ *there exists a random sequence* $(\gamma_k(v) : k \geq 1)$ *such that*

$$\lim_{k\to\infty} \frac{z_k^h(v)}{\gamma_k(v)} = \sum_{v'} u^h(v'), \quad \textit{almost surely.}$$

*In particular for large values of* $k$*, the vector* $(z_k^h(v) : h \in \mathcal{H})$ *and the posterior distribution* $\pi$ *have common modes.*

We prove the theorem via an adaptation of the techniques in [27] for asymptotic analysis of stochastic-matrix products. We start with auxiliary results.

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

$$\lambda(P) = 1 - \min_{i_1, i_2} \sum_j \min(p_{i_1 j}, p_{i_2 j}).$$

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

$$z_k^h = (B_1 B_2 \cdots B_k)^T u^h, \quad k \geq 1. \tag{24}$$

**Lemma 5.1** *Under the hypothesis of Theorem 5.2, 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 $B_k$s, entries of this product are positive if and only if all entries in

$$(I + A_{k_o+1})^T (I + A_{k_o+2})^T \cdots (I + A_{k_o+k})^T \tag{25}$$

are positive. The $(i, j)^{th}$ entry in the product (25) is positive if and only if a hypothetical message that originates at node $i$ in round $k_o$ can reach node $j$ 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 (25). Let $q(i, j)$ be the probability that link $i \to j$ is functional at a round, so that without loss of generality $E = \{(i, j) : q(i, j) > 0\}$, and let $(\xi(i, j) : (i, j) \in E)$ be independent geometric random variables where $\xi(i, j)$ has parameter $q(i, j)$.

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 $\sum_{(i,j)\in E} \xi(i,j)$. Define the random variable $\kappa$ 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, $\kappa$ is stochastically dominated by $|V|^2 \sum_{(i,j)\in E} \xi(i,j)$. Let $\mu$ 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 follows by an application of Markov's inequality. The lemma follows by choosing $k(\epsilon) = \mu/\epsilon$. $\square$

**Corollary 5.1** *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 square matrix $P = [p_{ij}]$ define

$$\delta(P) = \max_j \max_{i_1,i_2} |p_{i_1 j} - p_{i_2 j}|. \tag{26}$$

The following lemma is a recitation of [27, Lemma 2]:

**Lemma 5.2** *For $k \geq 1$*

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

**Proof of Theorem 5.2.** Fix $\sigma, \epsilon > 0$ and $k > k(\epsilon)$. Appeal to Lemma 5.2 to write

$$\delta(B_1B_2\cdots B_k) \leq \prod_{i=1}^{k \bmod k(\epsilon)} \lambda(B_i) \prod_{j=1}^{\lfloor k/k(\epsilon)\rfloor} \lambda\left(B_{k-jk(\epsilon)+1}B_{k-jk(\epsilon)+2}\cdots B_{k-(j-1)k(\epsilon)}\right).$$

Since each factor of the product on the left hand side is at most 1, Corollary 5.1 implies that the product is larger than $\sigma$ only if there are more than $\lfloor \log_d(\sigma)\rfloor$ values of $j$ with

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

Lemma 5.1 now implies that for $k > k(\epsilon)\lfloor \log_d(\sigma)\rfloor$

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

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

$$\limsup_{k\to\infty} \delta(B_1 B_2 \cdots B_k) \le \sigma, \quad \text{almost surely,}$$

due to the Borel-Cantelli Lemma. Arbitrariness of $\sigma$ implies that $\delta(B_1 B_2 \cdots B_k)$ converges, hence by definition (26) rows of the product $B_1 B_2 \cdots B_k$ almost surely become identical (though they do not necessarily settle to a fixed vector). In light of equality (24) the theorem follows by identifying $\gamma_k(v)$ with the $v$th entry of an arbitrary row of $B_1 B_2 \cdots B_k$. $\square$

**Remark 5.1** *Note that the proof of Theorem 5.2 relies only on Lemma 5.1; hence the conclusion of the theorem holds under much more relaxed assumptions on the statistics of* $(A_k : k \ge 1)$. *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 5.2** *The algorithm does not require estimation of packet loss probabilities.*

In practice estimating convergence rates is extremely difficult and is known to be NP hard. To illustrate the effect of packet losses we have performed a number of experiments. From simulations it appears that the convergence rate degrades gracefully and we do not see appreciable differences even for sufficiently large packet losses. A typical simulation of the average convergence is illustrated in Figure 4 for $N = 400$ sensors placed on a uniform grid with connectivity between any two nodes separated by minimum internode distance. At each round each link is assumed to be functional with probability $p$, independent of other links and other rounds.

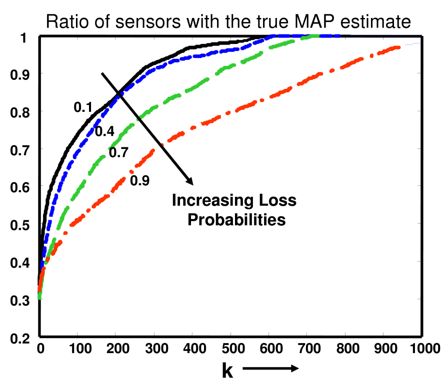



Figure 4: Convergence rate of consensus for different packet loss probabilities; Y-axis denotes percentage of sensors that have achieved consensus; X-axis denotes time index.

| Scheme | Energy (Joules/Node) |
|---|---|
| Decentralized Broadcast | $\mathcal{O}(N^{3/2}d_0^4E_b)$ |
| Decentralized Multihop | $\mathcal{O}(N^{1/2}d_0^4E_b)$ |
| Belief Propagation (grid) | $\mathcal{O}(N^{1/2}d_0^4E_b)$ |
| Hierarchical BP | $\mathcal{O}(\log(N)d_0^4E_b)$ |

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.

# 6 Scaling Laws

The scheme developed above is based on refinement of information at each time step and at each sensor location. In effect, we differentiate between data and information, and successively blend refinement and transportation of data in the course of the algorithm. This motivates comparison of decentralized detection with the BP approach developed in the paper. Comparisons between conventional decentralized detection scheme and the BP scheme appears in the Table 1. We consider four possible schemes for sensors on a uniform two-dimensional grid with connectivity between any two nodes separated by minimum internode distance. The first two schemes use conventional decentralized detection with a fusion center located approximately at the center. The last two schemes employ some version of distributed algorithms presented here.

We can handle two cases: (a) $N$ sensor nodes distributed uniformly over a square area; (b) $N$ nodes on a uniform square grid (lattice). For brevity we only consider the latter case here. Each node is separated by some minimum distance, $d_0$ with a power attenuation that scales with distance $d$ as $d^{-4}$ implying that if $E_b$ is the joules/bit required for reliable decoding over a unit distance then $E_b d^4$ is the corresponding energy required for reliable transmission for distance equal to $d$.

For the decentralized scheme we designate an arbitrary node as the fusion center. We have two possible communication schemes that can be applied. The first scheme is a point-to-point scheme wherein each node communicates its local decision directly to the fusion center. The second scheme relies on multi-hop, wherein each node relays its local decision to a neighboring node, which in turn forwards that information in the direction of the fusion center. For both of these schemes we have the following result.

**Proposition 6.1** *The average energy required for decentralized detection under the point-to-point scheme scales as $N^{3/2}E_b$ joules/bit/node while energy for the multi-hop scheme scales as $\sqrt{N}E_b$ joules/bit/node.*

**Proof.** For simplicity, we only consider a square area with nodes on a uniform grid organized into vertical and horizontal rows with the center node as the fusion center as shown in Figure 5. For the point-to-point scheme, it

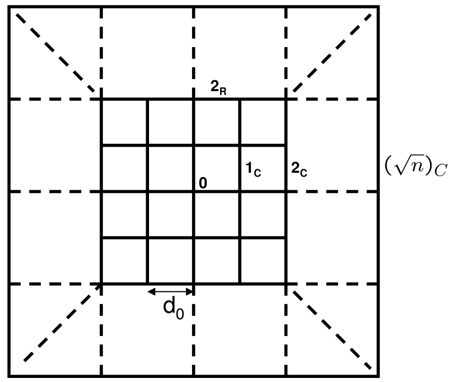



Figure 5: Sensors on a Uniform grid; $0,\ k_C,\ k_R$ denotes the fusion center, kth column and row respectively.

follows that the average energy consumption, $E_{ave}$, is given by:

$$E_{ave} = \frac{E_b}{N}\sum_k d_v^4$$

where, $d_v$ is the distance from the vth node to the fusion center. Now, a uniform square grid with minimum distance separation of $d_0$ requires a dimension of length and width equal to $d_0\sqrt{N}$. A lower bound for energy is computed as:

$$E_{ave} = \frac{E_b}{N}\sum_{v\in V} d_v^4 \geq \frac{E_b}{N}\sum_k^{\sqrt{N}}(\sqrt{N})(\sqrt{k}d_0)^4 = \mathcal{O}(N^2)E_b$$

where, the second equality follows from the fact that $\sqrt{N}$ nodes are located along any row and the minimum distance from the kth row to the fusion center is $kd_0$. The energy requirements for the multi-hop scheme can be computed by a straightforward application of max-flow min-cut theorem. Consider a cut along the kth horizontal row from the fusion center. The number of bits passing the cut towards the fusion center is equal to the number of nodes in the cut set, which is equal to $(\sqrt{N}-k)\sqrt{N}$. Now this traffic must be supported by $\sqrt{N}$ nodes that are at the boundary of the cut set and the total energy required is $E_b\sqrt{N}(\sqrt{N}-k)d_0^4$. Summing this energy over all cuts and normalizing with respect to the number of nodes gives us:

$$E_{ave} = \frac{E_b}{N}\sum_{k=1}^{\sqrt{N}}\sqrt{N}(\sqrt{N}-k)d_0^4 \approx \mathcal{O}(\sqrt{N})$$

Next we consider the BP approach. It follows from [9, 21] that an exponential convergence rate can be readily established, i.e.,

$$\|x(k) - \sum_{j=1}^{N} u^h(j)\|_\infty \leq C_0 r^k$$

Here $r$ is the ratio between the second and first largest eigenvalues of the matrix $\tilde{A}$ described in Equation 21. This leads to the conclusion that for a fixed minimum K-L distance between any two hypothesis, it requires $\mathcal{O}(\log(1/r))$ messages to realize the MAP estimate. Nevertheless, this is misleading because the constant term $C_0$ depends on graph as well [21]. To get a better idea of the convergence rate we compute upper and lower bounds. The diameter of the communication network quantifies a lower bound on the time it takes for the proposed scheme to converge to the consensus at each node as observed below.

**Lemma 6.1** *Suppose the network graph has a diameter equal to $D$. The energy required for the consensus scheme is lower bounded by $D$ joules/bit/node.*

**Proof.** The proof follows from the fact that $A^k$ has zero entries so long as $k < D$. This is because the diameter characterizes the minimum graphical distance between any two arbitrary nodes and it takes at least $D$ hops to guarantee that a message from any node reaches another arbitrary node.

**Remark 6.1** *For identical sensors, by appealing to large deviations theory [5], it is not difficult to show that with $\lfloor \log(1/\epsilon) \rfloor$ number of messages it is sufficient to ensure that with high probability (taken with respect to the observations) each sensor deviates less than $\epsilon$ from the centralized MAP estimate.*

We are after a more general result, i.e., the minimum time to reach a consensus irrespective of the specific realization of the observations. To derive an upper bound we first collect several results. First, from [4], it follows that for a stochastic matrix $P$ we have,

$$\|P^t x - \pi\|_\infty \leq \epsilon, \ \forall \ t \geq \frac{1}{1 - |\lambda_2|} \log \frac{N}{\epsilon}$$

where $\pi$ is the equilibrium distribution and $\lambda_2$ is the second largest eigenvalue in magnitude. In our context we are given a stochastic matrix, $\tilde{A} = (I + A)D$ (where the columns sum to one) and we need to derive bounds on its second largest eigenvalue.

An upperbound can be readily derived for grid graphs, which follows from [22]. A grid graph is defined as follows. Suppose, $G_1$ and $G_2$ are graphs so that $(v, w_1)$ and $(v, w_2)$, $(v_1, w)$ and $(v_2, w)$ are adjacent in $G_1 \times G_2$ if and only if $w_1$, $w_2$ are adjacent in $G_2$, and $G_1$ respectively. The eigenvalues are described in the following proposition, which we state from [22]:

**Proposition 6.2** *Consider a grid graph as described above formed with line graphs $G_1$ and $G_2$. It then follows that the eigenvalues of the adjacency matrix of the grid graph are the sum of the eigenvalues of $G_1$ and $G_2$ respectively.*

The proposition can be directly applied to compute the second eigenvalue. Suppose $G_1$ and $G_2$ are each formed with $\sqrt{N}$ nodes along a straight line with a separation distance equal to $d_0$. Furthermore, let the communication connectivity radius be equal to $d_0$. Then the eigenvalues of the adjacency matrix 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 the following energy scaling for grid graphs:

**Proposition 6.3** *For grid graphs the average energy required for achieving the MAP consensus for a hypothesis testing problem with two hypothesis scales as $\mathcal{O}(N)d_0^4 E_b$ Joules/node.*

We point out that to establish this result we have used the fact that the conversion from messages to bits does not significantly increase the energy requirements. This follows from the quantization results in Section 5, where we showed there for $M$ hypothesis it only takes $\mathcal{O}(M)$ bits to encode messages for each node.

A similar result can also be established for random graphs described in Theorem 4.4 except that the results now hold with high probability. These results imply that the energy requirements for achieving consensus is worse than that required for multi-hop decentralized detection and significantly better than the point-to-point decentralized detection scenario. However, it should be pointed out that the consensus based approach achieves the optimal MAP consensus irrespective of the actual realization along with significant robustness to link losses.

Our remedy relies on seeking a MAP estimate at a few nodes as opposed to all of the nodes. In this scheme sensor nodes are organized hierarchically as shown in Figure 6. In the lowest layer small clusters of sensors achieve their respective consensus within the cluster. A cluster head within each cluster forms part of a higher level network over a larger area. Consensus in this larger area is then attained. This strategy can be repeated over several layers with significant energy savings. We point out that the algorithms employed in this setup are ad-hoc and lack knowledge of node locations, which is consistent with the underlying philosophy outlined in the rest of the paper.

Now consider $N$ nodes on a grid graph. Partition into $P = N/L$ clusters each of size $L$. Suppose $E_m$ is the energy required to transmit each message to a neighboring node. To achieve a consensus among each of the clusters the total energy, $E(P, L, E_m)$, required is equal to the product of the number of clusters times the number of nodes in the cluster times the number of messages, $M$, per node (to achieve consensus) times energy, $E_m$, required for each message. Now the number of messages required to achieve a consensus in a cluster is of order of the size of the cluster. Therefore,

$$E_0 = E(P, L, E_m) = PL(L)(E_m) = NLd_0^4 E_b.$$

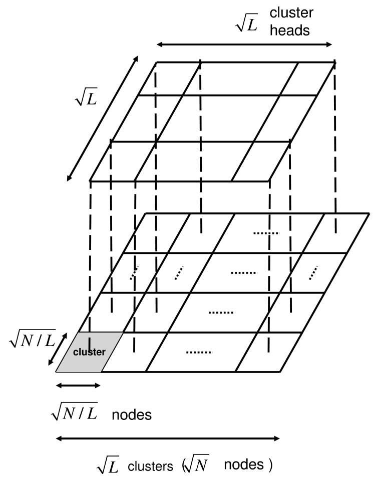



Figure 6: Illustration of Hierarchical Scheme; Sensor nodes organized in to clusters of size $L$; Each cluster has a clusterhead which forms part of a higher level network

Now at the next layer $P$ nodes collaborate to achieve a consensus. The smallest distance between any two nodes in the $P$ cluster is now $\sqrt{L}d_0$. Therefore, if a point-to-point scheme is used the energy/message/node, $E_m = L^2 d_0^4 E_b$. The total energy is then $E(1, P, E_m) = P(P\sqrt{L}^4 d_0^4 E_b)$. However, for a multi-hop scheme (using intermediate nodes as relays) the energy, $E_m = \sqrt{L}d_0^4 E_b$ and so the total energy is $E_1 = E(1, P, E_0) = P(P\sqrt{L}d_0^4 E_b)$. Therefore, for $P = L = \sqrt{N}$ the average energy/node required for the multi-hop case scales as $E_{ave} = (E_0 + E_1)/N \equiv \mathcal{O}(\sqrt{N})$, which is equal to the average energy required for multi-hop decentralized scheme. This idea can be extended to obtain the following setup. The first layer is given by $N/L$ partitions of size $L$ clusters. The second layer is partitioned into $(N/L)/L = N/L^2$ partitions of size $L$ clusters (formed with cluster heads for each cluster in the first layer). This process is then repeated until all the nodes are exhausted.

The main drawback of the optimal multi-hop scheme is that it uses knowledge of node locations. Therefore, to get around this issue a suboptimal scheme can be considered. Here at the kth layer the cluster heads from the (k-1)th broadcast information to the neighboring nodes (which serve as relays). These neighboring nodes in turn broadcast the received information to their neighbors and so on it follows. Ultimately, this information is received by the nearest clusterheads, which then update their likelihoods/messages and re-broadcast the new information. The energy required for this scheme can be computed as follows. As described earlier, in the second layer the distance between any two clusterheads is no more than, $d = \sqrt{L}d_0$. Therefore, if a message is broadcast and relayed by adjacent nodes it is guaranteed to reach the nearest clusterhead after $\sqrt{L}$ time steps. In this time a maximum of $L$ nodes have also broadcast the same message. This follows from the fact that there are at most $L$ nodes in the circle of radius $\sqrt{L}d_0$. We now have the following theorem:

**Theorem 6.1** *For the hierarchical setup described above the average energy scales as* $\mathcal{O}(L\log(N))$ *Joules/node.*

The proof follows by direct computation. First, we note that if the minimum distance is $d_0$ for the first layer, then the minimum distance is $\sqrt{L}d_0$ in the second layer. In general, for the kth layer the minimum distance, $d_k = \sqrt{L}d_{k-1}$. Therefore, the energy and the average energy, $E^k$, $E^k_{ave}$, respectively (assuming a multihop strategy) used upto the kth layer is:

$$E^k = E^{k-1} + (N/L^k)L^2\left(\frac{d_k}{d_0}\right)^2 d_0^4 E_b,\ \ E^k_{ave} = E^{k-1}_{ave} + (1/L^k)L^2\left(\frac{d_k}{d_0}\right)^2 d_0^4 E_b.$$

Now the maximum value for $k$ is bounded by $k_{max} = \log(N)/\log(L)$. Therefore upon direct computation we get that the average energy scales as $\mathcal{O}(L\log(N)/\log(L))$ Joules/node to achieve a MAP estimate. For a constant $L$ this leads to exponential improvement in energy scaling.

# 7 Distributed Estimation

The techniques of Section 5.2 can be extended to continuous parameter estimation via standard approximation techniques. Namely, let $Z$ be a continuous random variable on $R^k$, 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 finite partition $\{R_m(\epsilon)\}$ of $R^k$ and constants $\{\gamma_m(\epsilon)\}$ such that

$$\|Z - Z_\epsilon\| < \epsilon;\ \text{ where }\ Z_\epsilon = \sum_m \gamma_m(\epsilon) I\{Z \in R_m(\epsilon)\}.$$

When each event $I\{Z \in R_m(\epsilon)\}$ is interpreted as a separate hypothesis, the distributed algorithm of Section 5.2 can be employed to compute exact posterior probabilities of these events, in turn the centralized posterior distribution of $Z$ can be approximated up to a desired accuracy.

It is worthwhile to consider in more detail special cases in which the constituent variables of the problem are jointly Gaussian, since distributed algorithms that entail no approximation errors can be identified for such cases. In that case, define, for each $v$, the locally computable quantities

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

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

$$E[Z|Y_v = y_v, v \in V] = \left(\sum_v \sigma_v^{-1}\right)^{-1} \sum_v \xi_v.$$

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

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

where $(I + A_k)D_k$ is as defined in Section 5.2, so that by Theorem 5.2

$$
\lim_{k \to \infty} \frac{x_k}{s_k} = E[Z | Y_v = y_v, v \in V].
$$

## 8 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.

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

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

[5] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley Series in Communications, 1991.

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

[7] M. Gastpar and M. Vetterli. Scaling laws for homogeneous sensor networks. *In 41st Annual Allerton Conf. on Commun., Control and Comp., (Monticello, IL, USA)*, 2003.

[8] 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.

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

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

[11] 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.

[12] R. Nowak and U. Mitra. Boundary estimation in sensor networks. In *2nd International Workshop on Information Processing in Sensor Networks*, Palo Alto, CA, April 2003.

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

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

[15] S. S. Pradhan, J. Kusuma, and K. Ramchandran. Distributed compression in a dense microsensor network. *IEEE Signal Processing Mag.*, pages 51–60, 2002.

[16] M. G. Rabbat and R. D. Nowak. Decentralized source localization and tracking. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, Canada, July 2004.

[17] 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.

[18] V. Raghunathan, C. Schurgers, S. Park, and M. B. Srivastava. Energy-aware wireless microsensor networks. *IEEE Signal Processing Mag.*, pages 40–50, 2002.

[19] 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.

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

[21] E. Seneta. *Non-negative Matrices and Markov Chains*. Springer Verlag, 1973.

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

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

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

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

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

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

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

# 9 Appendix

**Proof of Theorem 4.4**

Consider the region $\mathcal{Z}$ of unit area in which $N$ nodes are uniformly distributed with edges between any two nodes if Equation 18 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 10 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, $p = \text{Prob}\{X_k^j = 1\} = \text{Vol}(\mathcal{C}_j)$. 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.

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

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

$$\frac{1}{pN}\sum_{k=1}^{N} X_k^j \longrightarrow 1, \text{ almost surely} \Longrightarrow \text{node degree} = \log^3 N\left(1 + \frac{C}{\log N}\right),\ |C| \le 1, \text{ w.h.p} \tag{27}$$

**Proof of primitivity:** This follows from Theorem [8.5.3] in [9], 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 [9]) of the induced graph. The main complication is that the matrix $A$ as defined by Equation 10 is directed. In particular note that $a_{e,e'} = 0$ if $e$ and $e'$ forms a directed cycle. To establish strong connectedness of $\Gamma(A_N)$ we let $e_1 = (s_1,\, d_1),\ e_m = (s_m,\, d_m)$ be any two edges. From the node connectivity it follows that there is a directed path from node $d_1$ to node $d_m$. Suppose, this path contains a directed cycle, 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 can be constructed.

**Proof of Equal Column Sums:** Now since the matrix $A$ is primitive it follows that there exists a unique positive Perron eigenvalue, $\rho(A)$ and a positive eigenvector. Furthermore primitivity is not destroyed by pre-multiplication by $(I + \Delta)^{-1}$ as was done in Equation 19. It suffices to show that $\tilde{A} = A(I + \Delta)^{-1}$ has equal column sums. Notice that a column $j$ of $A$ has $\log^3 N(1 + C/\log N)$ non-zero entries from Equation 27. Therefore, by choosing $|\Delta| \le 1/\log N$ the columns can be made equal.