Learning the Interference Graph of a Wireless Network

Jing Yang, *Member, IEEE,* Stark C. Draper, *Member, IEEE,* Robert Nowak, *Fellow, IEEE*

***Abstract*—A key challenge in wireless networking is the management of interference between transmissions. Identifying which transmitters interfere with each other is a crucial first step. Complicating the task is the fact that the topology of wireless networks changes with time, and so identification may need to be performed on a regular basis. Injecting probing traffic to assess interference can lead to unacceptable overhead, and so this paper focuses on interference estimation based on passive traffic monitoring. We concentrate on networks that use the CSMA/CA (Carrier Sense Multiple Access/Collision Avoidance) protocol, although our model is more general.**

We cast the task of estimating the interference environment as a graph learning problem. Nodes represent transmitters and edges represent the presence of interference between pairs of transmitters. We passively observe network traffic transmission patterns and collect information on transmission successes and failures. We establish bounds on the number of observations (each a snapshot of a network traffic pattern) required to identify the interference graph reliably with high probability.

Our main results are scaling laws telling us how the number of observations must grow in terms of the total number of nodes n in the network and the maximum number of interfering transmitters d per node (maximum node degree). The effects of hidden terminal interference (i.e., interference not detectable via carrier sensing) on the observation requirements are also quantified. We show that it is necessary and sufficient that the observation period grows like $d^2 \log n$, and we propose a practical algorithm that reliably identifies the graph from this length of observation. The observation requirements scale quite mildly with network size, and networks with sparse interference (small d) can be identified more rapidly. Computational experiments based on a realistic simulations of the traffic and protocol lend additional support to these conclusions.

***Index Terms*—Interference graph learning, CSMA/CA protocol, minimax lower bounds**

I. Introduction

Due to the broadcast nature of wireless communications, simultaneous transmissions in the same frequency band and time slot may interfere with each other, thereby limiting system throughput. Interference estimation is thus an essential part of wireless network operation. Knowledge of interference among nodes is an important input in many wireless network configuration tasks, such as channel assignment, transmit power control, and scheduling. A number of recent efforts have made significant progress towards the goal of real-time identification of the network interference environment. Some of the recent approaches (e.g., Interference maps [2] and Micro-probing [3]) inject traffic into the network to infer occurrences of interference. While such approaches can be quite accurate in determining interference, the overhead of making such "active" measurements in a large network limits their practicality. In addition, since network topology often changes over time, the periodic use of active probing methods to identify interference can place a significant burden on the network. Since the network topology changes at a relatively slow time scale, it is reasonable to assume that the interference among nodes stays constant over a period of time. We study an interference learning problem where the underlying interference pattern among the nodes stay fixed during the observation period. The theoretical results we obtain indicate the time scale over which the interference patterns can be identified and tracked.

Inspired by two passive WLAN (wireless local area network) monitoring approaches Jigsaw [4], [5] and WIT [6], a different approach to online interference estimation "Passive Interference Estimation" (PIE) is explored in [7], [8]. Rather than introducing active probe traffic, interference estimates are based on passively collected data. PIE estimates the interference in the network in an on-line fashion. All interference estimates are obtained by observing ongoing traffic and feedback information at transmitters. Experimental studies in small testbeds [8] show that PIE is quite promising, but very little is understood about how the method scales-up to large complex networks. That is the focus of this paper.

The main contribution of this paper is to quantify the required observation time as a function of network size and topological connectivity. This provides insight into the time scale over which network interference patterns can be identified, and tracked, in dynamically changing wireless networks. Our result is applicable to WLANs and other wireless communication systems and technologies operating under CSMA/CA-like protocols such as, e.g., Wireless Personal Area Networks (WPAN) [9] and ZigBee [10].

We formulate passive interference estimation as a statistical learning problem. Given an arbitrary WLAN that consists of n access points (APs) and a variable number of mobile clients, our goal is to recover the "interference" or "conflict" graph among these APs with as few measurements as possible. This graph encodes the interference relations between APs and other APs' clients. We study three versions of the problem. In the first version we study "direct" interference between APs where edges in the graph indicate that a pair of APs are within each other's carrier sensing range. Letting d be the maximum number of interfering APs per AP, we show that to identify

This work was supported in part by the Air Force Office of Scientific Research under grant FA9550-09-1-0140 and by the National Science Foundation under grant CCF 0963834. It was presented, in part, at the *IEEE Int. Symp. Inform. Theory*, Cambridge, MA, Jul. 2012 [1].

J. Yang, S. C. Draper and R. Nowak are with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, WI, 53706, USA. e-mail: {yangjing, sdraper, nowak}@ece.wisc.edu.

the conflict graph one must collect a number of measurements proportional to $d^2 \log n$. This is quite mild dependence on the network size n and indicates that interference graph inference is scalable to large networks and sparser patterns of interference are easier to identify than denser patterns. In the second version we quantify the effect of "hidden" terminal interference. This type of interference occurs when one AP interferes with another AP's clients, but the transmission of the interfering AP is not detectable by the other AP. In this case feedback on transmission successes and failures is required to estimate the graph. In the first two versions of the problem we assume the interference environment is static, i.e., time-invariant. In the third version we study both the direct and hidden interference problem when the channel state between pairs of devices can fluctuate randomly. This generalization captures effects such as deep fades (and poor channel state information) that can cause uncertainty as to whether a transmission failure is due to interference some other cause. For all versions of the problem we present easy-to-implement graph estimation algorithms. The algorithms are adaptive to d, in the sense that they do not require apriori knowledge of d. For the first two (static) versions we also develop lower bounds that demonstrate that the time-complexity attained by the algorithms cannot be improved upon by any other scheme.

We adopt the following set of notations. We use upper-case, e.g., X, and bold-face upper case, e.g., $\mathbf{X}$, to denote random variables and vectors, respectively. A vector without subscript consists of n elements, each corresponds to an AP, e.g., $\mathbf{X} := (X_1, X_2, \ldots, X_n)$. Sets and events are denoted with calligraphic font (e.g., $\mathcal{E}$). The cardinality of a finite set $\mathcal{V}$ is denoted as $|\mathcal{V}|$. A vector with a set as its subscript consists of the elements corresponding to the transmitters in the set, e.g., $\mathbf{Q}_\mathcal{C} := \{Q_c\}_{c\in\mathcal{C}}$.

The paper is organized as follows. In Section II we formulate the WLAN interference identification problem as a graph learning problem. We review the CSMA/CA protocol and propose a statistical model for a network using this protocol. In Section III we present our main results for the static setting in the form of matching upper and lower bounds (up to constant factors) on the observation requirements for reliable estimation of both direct and hidden interference graph. In Section IV, we generalize the statistical model to account for random fluctuations in the interference environment and present achievable results for "robust" graph inference. We present an experimental study in Section V that supports the theoretical analysis. The experiments are based on simulations of the traffic and protocol that incorporate more real-world effects than the model used to develop the theory. However, in the experiments we do observe the scaling behavior predicted by the theory. Concluding remarks are made in Section VI. Most proofs are deferred to the appendices.

II. Problem Formulation

In this section we present the problem setting. In Section II-A we present the important characteristics of the multiple-access protocol. In Section II-B we model the interference environment using a graph. In Section II-C we present the estimation problem and sketch our algorithms. Finally, in Section II-D we present the statistical model and assumptions that underlie our analysis.

A. CSMA/CA protocol and ACK/NACK mechanism

We assume the system uses a CSMA/CA-like protocol at the medium access control layer, e.g., [11]. The important characteristics of the protocol are the following. When the transmitter has a packet to send, it listens to the desired channel. If the channel is idle, it sends the packet. If the channel is busy, i.e., there exists an active transmitter within the listener's carrier sensing range, the transmitter holds its transmission until the current transmission ends. At the conclusion of that transmission the following back-off mechanism is invoked. The transmitter randomly chooses an integer number, τ, uniformly in the range $[0, W-1]$, and waits for τ time slots. The positive integer W represents the back-off window size. If the channel is idle at the end of the node's back-off period, the node transmits its packet; otherwise it repeats the waiting and back-off process until it find a free channel. Statistically, the random back-off mechanism allows every node equal access to the medium.

However, even if two APs are not within each other's carrier sensing range, the transmission from one AP may still corrupt the signal received at clients of the other. This is the so-called "hidden terminal" problem. To identify this type of interference further information is needed. We assume that an ACK/NACK mechanism is used. Specifically, we assume that whenever an AP successfully delivers a packet to its destination, an ACK is fed back to acknowledge the successful transmission. If the AP does not receive the ACK after a period of time, it assumes that the packet is lost. The ACK/NACK mechanism enables the APs to detect collisions however they occur.

B. Interference Graph

We use a graph $G = (\mathcal{V}, \mathcal{E})$ to represent the interference among APs in the network. The node set $\mathcal{V}$ represents the APs, and edge set $\mathcal{E}$ represents the pairwise interference level among APs. An example of such of graph is depicted in Figure 1.

We partition $\mathcal{E}$ into two subsets: *direct* interference $\mathcal{E}_D$ and *hidden* interference $\mathcal{E}_H$. Direct interference occurs when two APs are within each other's carrier sensing range. An example in Figure 1 is the AP pair $(4, 5)$. Under the assumption that the carrier sensing range is the same for every AP, the edges in $\mathcal{E}_D$ are *undirected*. However, as mentioned above, carrier sensing cannot resolve all of the collisions in the network. Hidden-terminal type interference is represented by the edges in $\mathcal{E}_H$. Such interference may be asymmetric and so the edges in $\mathcal{E}_H$ are *directed*. In Figure 1 the AP pairs $(1, 2)$ and $(3, 4)$ cannot detect each other. Yet, they can interfere with each other's clients since their carrier sensing ranges intersect. The effected clients would (roughly) lie in the intersection of the (roughly circular) carrier sensing ranges of the AP pair. Nodes 3 and 5 are are an example of an AP pair that does not interfere (either directly or indirectly) since their carrier sensing ranges do not overlap.

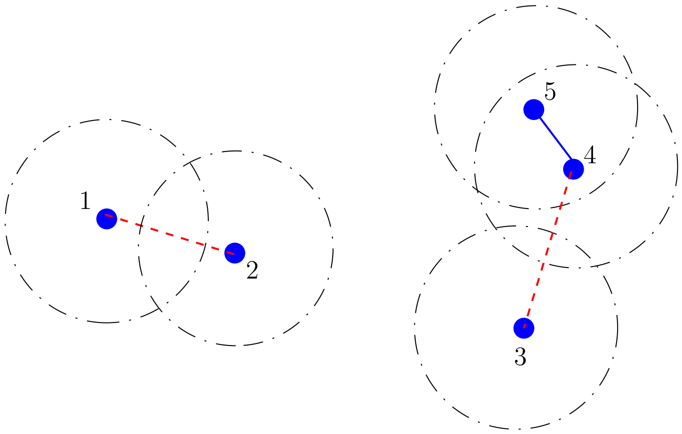

Fig. 1. System model with 5 APs. The edges connecting APs represent the interference between them. Solid line represents direct interference, dashed lines represent hidden interference. Each circle represents the carrier sensing range of the AP at its center. Clients that are located in the intersection of the carrier sensing regions of a pair of APs may be subject to interference.

One can further associate a probability or weight with each interference edge. This weight reflects the level (severity or persistence) of the interference. Our main interest is to identify the presence of interference. Once this is accomplished, estimating the severity of interference is straightforward. Thus, our emphasis is on determining the presence of edges and not their weights. The sparsity of the interference graph is parameterized by d, the maximum node degree. This parameter is a bound on the number of other nodes that interfere with any particular node.

C. Inference Problem and Algorithm

Our objective is to determine the edge set $\mathcal{E}$ based on passive observations. We assume the existence of a central controller that, over a period of time, collects the transmission status of all APs and information on transmission successes (ACKs). The collected information during this *observation period* is the dataset we use to estimate the interference graph. Both enterprise-type WLANs and proposed architectures for 4G systems that involve combinations of macro plus pico-cells have such central controllers.

In this section and in Section III we consider *static channel states*. Specifically, we assume that whether or not a pair of APs can sense each other is fixed (deterministic) throughout the period of observation. For hidden interference, we also assume that the channel between each AP and all clients of other APs is also static. Further, if a transmission failure occurs it must have been caused by a collision with a packet transmitted by one of the other active APs. Thus, for now we ignore other factors that contribute to transmission failures such as deep fades. In Sec. IV, we relax this assumption and take the randomness of the communication channels into consideration. We defer till then the formal definition of the system model with *random* channel states.

We now outline the basic ideas underlying our algorithms. First, consider the direct interference environment, characterized by $\mathcal{E}_D$. Note that due to the use of carrier sensing in CSMA/CA, if two APs are within each other's carrier sensing range, they will not transmit simultaneously. One can therefore infer that any pair of APs that transmit simultaneously during the observation period must not be able to hear each other. In other words, there is no *direct* interference between them. The algorithm starts with a full set of $\binom{|\mathcal{V}|}{2}$ candidate edges. Each time a simultaneous transmission is observed, the corresponding edge is removed. If the observation period is sufficiently long, all possible co-occurring transmission will be observed and the correct *direct* interference edge set $\mathcal{E}_D$ will be recovered.

Next, consider the cases of hidden interference characterized by $\mathcal{E}_H$. Estimating this set is more involved and requires the collected ACK information. When a collision is detected at an AP, it implies that at least one of the other APs transmitting at the same time is interfering. The subset of APs transmitting at that time instance becomes a candidate set of hidden interferes for that AP. For each collision detected by the AP, another candidate subset is formed. The true set of hidden interferers for that AP must intersect with all the candidate sets. When the observation period is sufficiently long, we show that the *minimum hitting set* [12] that intersects with *all* of the candidate subsets is the set of hidden interferers. The edge set $\mathcal{E}_H$ can thus be recovered.

Our approaches to both problems were inspired by the PIE algorithms proposed in [7], [8]. For the direct interference problem, our approach is similar to PIE. Both rely on simultaneous transmissions to infer the interference graph. Our results provide the theoretical analysis and scaling behavior characterization that complement the empirical results of [7], [8]. Our approach to the hidden interferences problem is quite different from [7], [8]. While PIE uses a correlation based approach we use the hitting set approach mentioned above which attains more accurate results.

D. Statistical Model with Static Channel States

The graph $G_D = (\mathcal{V}, \mathcal{E}_D)$ represents the carrier sensing relationships among the APs. Specifically, if AP i and AP j are within each other's carrier sensing range, there is an edge between i, j, denoted as (i, j). The existence of such an edge implies that AP i and AP j are close and the transmission of each can be observed by the other. An example in Figure 1 is the AP pair $(4, 5)$. We term this "direct" interference and G_D the direct interference graph. We define $\mathcal{N}_i$ to be the set of neighbors of AP i in $G_D = (\mathcal{V}, \mathcal{E}_D)$, and let $d_i = |\mathcal{N}_i|$.

Observations of the network activation status are taken at time epochs. We use $X_i(t) \in \{0, 1\}$ to denote the activation state of node i at time t: $X_i(t) = 1$ means that node i transmits; $X_i(t) = 0$ means that node i does not transmit. In general, $X_i(t)$ is determined by the traffic status and backoff times of the APs that compete for the same channel.

As mentioned, to infer the hidden interference we required information on transmission successes and failures. Define $Y_i(t) \in \{0, 1, \varnothing\}$ to be the feedback information information received at AP i at the end of session t. $Y_i(t) = 1$ means that an ACK is received at AP i, indicating that the transmission in session t is successful; $Y_i(t) = 0$ means that the transmission failed, caused by some simultaneous transmissions; $Y_i(t) = \varnothing$ means that node i does not transmit in that session, i.e., $X_i(t) = 0$.

We let a graph $G_H = (\mathcal{V}, \mathcal{E}_H)$ represent the hidden interference among APs that cannot hear each other. This interference depends on the locations of the clients associated with each AP and thus may not be symmetric. These edges are therefore *directed*. We define

$$p_{ij} = \mathbb{P}(Y_j(t) = 0 | X_i(t) = X_j(t) = 1, \mathbf{X}_{\mathcal{V}\backslash\{i,j\}}(t) = \mathbf{0}) \quad (1)$$

i.e., p_{ij} is the probability that, when i, j are isolated from the rest of the APs and $X_i(t) = X_j(t) = 1$, AP i interferes with AP j, causing transmission failure of AP j.

The p_{ij} capture the randomness in locations of the clients associated with each AP. An AP may interfere with only a subset of the clients of a neighboring AP. For example, in Fig. 1 AP 1 may interfere with a client of AP 2 that is halfway between the APs, but it likely will not interfere with a client on the far side of AP 2. Thus, which clients one AP interferes with depends on the location of the other AP's clients. The value of p_{ij} represents the proportion of AP j's clients that AP i interferes with. It can be interpreted as the probability that AP j communicates with a client located in the overlapped area of the carrier sensing ranges of APs i and j. We don't define p_{ij} for $(i, j) \in \mathcal{E}_d$ since such pairs of APs are within each other's carrier sensing range and thus never transmit simultaneously. This is the case for AP 4 and AP 5 in Fig. 1.

Define $\mathcal{S}_j = \{(i, j) \in \mathcal{E}_H \mid i \in \mathcal{V}\}$ to be the hidden interferer set for AP j, i.e., the set of APs with $p_{ij} > 0$. We let $s_j = |\mathcal{S}_j|$. We point out that in general $\forall \mathcal{S} \subseteq \mathcal{V}\backslash\{i, j\}$,

$$p_{ij} \leq \mathbb{P}(Y_j(t) = 0 | X_i(t) = 1, X_j(t) = 1, \mathbf{X}_{\mathcal{S}} = \mathbf{0}), \quad (2)$$

i.e., p_{ij} is *less* than the probability that a collision occurs at AP j when both APs i and j are transmitting. This is because the collision at AP j may be caused by an active AP other than i, and AP i may just happen to be transmitting at the same time.

The complete interference graph $G = (\mathcal{V}, \mathcal{E})$ consists of both direct interference graph and hidden interference graph, i.e., $\mathcal{E} = \mathcal{E}_D \cup \mathcal{E}_H$. We note that $\mathcal{E}_D \cap \mathcal{E}_H = \varnothing$. We now state some statistical assumptions important in the development of our analytic results.

Assumption 1

(0) *APs are synchronized and the time axis is partitioned into synchronized sessions where each session consists of a contention period followed by a transmission cycle. We use $t = 1, 2, \ldots$ to denote the indices of the sessions.*

(i) *The traffic status $Q_i(t)$ are i.i.d. Bernoulli random variables with common parameter p, where $0 < p < 1$, for all $i \in \mathcal{V}$ and all $t \in \mathbb{Z}^+$. In other words, the $Q_i(t)$ are independent across transmitters and sessions: $\mathbb{P}(Q_i(t) = 1) = p$ for all i and t. APs competes for the channel at the beginning of session t if and only if $Q_i(t)$.*

(ii) *The backoff time $T_i(t)$ are continuous i.i.d. random variables uniformly distributed over $[0, W]$ and are statistically independent for all $i \in \mathcal{V}$ and all $t \in \mathbb{Z}^+$. W is a fixed positive integer.*

(iii) *For all $(i, j) \in \mathcal{E}_H$, there exists a constant p_{min}, $0 < p_{\min} < 1$, s.t. $p_{ij} \geq p_{\min}$.*

(iv) *For all $i \in \mathcal{V}$, there exists an integer $d \geq 1$, s.t. $d_i \leq d - 1$.*

(v) *For all $j \in \mathcal{V}$, there exists $s \in \mathbb{Z}^+$, s.t. $s_j \leq s$.*

Assumption 1-(0) is essential for our analysis. In practical WLANs, APs competing for the same channel are "locally" synchronized because of the CSMA/CA protocol. Assumption 1-(0) approximates the original "locally" synchronous system as a synchronized one. Since APs far apart on the interference graph have a relatively light influence on each other's activation status, this assumption provides a good approximation of the original system and greatly simplifies our analysis. Assumption 1-(i) ignores the time dependency and coupling effect of the traffic queue status of APs. It is a good approximation of traffic for a system operating in light traffic regime. Assumption 1-(ii) that the $T_i(t)$s are continuous rather than discrete quantities (as they are in CSMA/CA protocol), implies that with probability one no two adjacent nodes in $G_D = (\mathcal{V}, \mathcal{E}_D)$ will have the same back-off time. This is a valid assumption when W is a large integer. The first two assumptions guarantee that the joint distribution of $X_i(t)$s and $Y_i(t)$s is independent and identical across t. Therefore, in the analysis hereafter we ignore the time index and focus on the distribution of X_is and Y_is in one observation. Assumption 1-(iii) defines the lower bound of significant interference. The final assumptions, (iv) and (v), model the fact that the interference graph will be sparse because of the widespread spatial distributions natural to the large-scale wireless networks of interest.

III. Main Results for Static Channels

In this section we present our main results for static channels. We break the overall problem into four subproblems. In Section III-A we provide an achievable upper bound on the number of observations required to infer the directed interference edge set $\mathcal{E}_D$. In Section III-B we present a matching lower bound on the number of observations required to infer $\mathcal{E}_D$. In Section III-C we provide an achievable upper bound on the number of observations required to infer the hidden interference edge set $\mathcal{E}_H$. Finally, in Section III-D we present a lower bound on the number of observations required to infer $\mathcal{E}_H$.

A. An Upper Bound for Inference of $G_D = (\mathcal{V}, \mathcal{E}_D)$

In this section, we formalize the algorithm sketched in Section II-C for estimating $G_D = (\mathcal{V}, \mathcal{E}_D)$ and present analysis results. In the static setting, for any $(i, j) \in \mathcal{E}_D$, APs i and j can always sense each other's transmissions. For any AP pair $(i, j) \notin \mathcal{E}_D$, the APs can never detect each other's transmissions.

Say that a sequence of transmission patterns $\mathbf{X}(1)$, $\mathbf{X}(2)$, $\ldots$, $\mathbf{X}(k)$ is observed. Due to the use of the CSMA/CA protocol and the continuous back-off time of Assumption 1-(ii), any pair of APs active in the same slot must not be able to hear each other. Thus, there is no edge in $G_D = (\mathcal{V}, \mathcal{E}_D)$. In other words, given an observation $\mathbf{X}$, for any i, j with $X_i = X_j = 1$, we know that $(i, j) \notin \mathcal{E}_D$.

Based on this observation, the algorithm starts at $t = 1$ with a fully connected graph connecting the n APs with $\binom{|\mathcal{V}|}{2} = n(n-1)/2$ edges. For each transmission pattern $\mathbf{X}$ observed, we remove all edges (i, j) s.t. $X_i = X_j = 1$. Our first result quantifies the number of observations k required to eliminate, with high probability, all edges not in $\mathcal{E}_D$, thereby recovering the underlying interference graph G_D. In Appendix A we show the following result:

Theorem 1 *Let $\delta > 0$, and let*

$$k \geq \frac{1}{-\log(1 - p^2/d^2)} \left(\log \binom{n}{2} + \log \frac{1}{\delta} \right).$$

Then, with probability at least $1-\delta$, the estimated interference graph $\hat{G}_D = (\mathcal{V}, \hat{\mathcal{E}}_D)$ is equal to G_D after k observations.

The idea of the proof is first to lower bound the probability that two nonadjacent APs i, j *never* transmit simultaneously in k observations. Then, by taking a union bound, an upper bound on the required k is obtained.

Remark: In the theorem the upper bound $k = O(d^2 \log n)$ when $p^2/d^2 \ll 1$. This is a natural regime where wireless networks become dense and the carrier sensing threshold is fixed. This order is the best we hope for in terms of p and d through passive observations. This is because if two non-interfering APs never transmit simultaneously, their behavior is the same as if they were within each other's carrier sensing range. Thus, we cannot determine whether or not there is an edge between them. Since each transmitter competes with its neighbors, the probability that it gets the channel has the order of $1/d$. So, roughly speaking, it takes about d^2 snapshots to observe two non-interfering APs active at the same time. Since there are about n^2 such pairs in the network, a union bound gives us the factor $\log n$.

B. A Minimax Lower Bound for Inference of $G_D = (\mathcal{V}, \mathcal{E}_D)$

We now provide a minimax lower bound on the number of observations needed to recover the direct interference graph $G_D = (\mathcal{V}, \mathcal{E}_D)$. Denoting the estimated graph as $\hat{G}_D = (\mathcal{V}, \hat{\mathcal{E}}_D)$ we prove the following result in Appendix B.

Theorem 2 *If $n \geq 3$, $d - 1 \leq (3n - \sqrt{n^2 + 16n})/4$, and*

$$k \leq \frac{\alpha (d-1)^2}{\left(2 + \frac{1}{1-p}\right)} \log n,$$

then, for any α, $0 < \alpha < 1/8$,

$$\min_{\hat{G}_D} \max_{G_D} \mathbb{P}(\hat{G}_D \neq G_D; G_D) \geq \frac{\sqrt{n}}{1 + \sqrt{n}} \left(1 - 2\alpha - \sqrt{\frac{2\alpha}{\log n}} \right).$$

The approach to this result is the following. We construct a set of M maximum-degree d graphs. We construct the set so that the graphs in the set are very similar to each other. This will make it hard to distinguish between them. For each graph in the set the statistical assumptions of Section II-D induce a distribution on the observed transmission patterns. Given k observed transmission patterns we consider the M-ary hypothesis test for the underlying graph. Since the size of the candidate graph set for the original estimation problem is much greater than M, this test is easier than the original problem. Therefore, a lower bound for this test will also lower bound the original estimation problem. Since for each graph we know the distribution of patterns, we can lower bound the probability of error for this hypothesis test using the Kullback-Leibler divergence between each pair of induced distributions. We prove that, for the collection of graphs we construct, the number of observations required to detect the correct underlying graph has the order of $\Omega(d^2 \log n)$.

Remark: We note that the order of the lower bound, $\Omega(d^2 \log n)$, has the same order as the upper bound in Theorem 1 when $p^2/d^2 \ll 1$. Therefore, the estimation method based on pairwise comparison is asymptotically optimal when $d \to +\infty$.

C. An Upper Bound for Inference of $G_H = (\mathcal{V}, \mathcal{E}_H)$

We now present result on inferring the hidden interference graph $G_H = (\mathcal{V}, \mathcal{E}_H)$ through observations of $\mathbf{X}(1), \ldots \mathbf{X}(k)$ and $\mathbf{Y}(1), \ldots \mathbf{Y}(k)$. When the transmission of AP j fails (indicated by the feedback $Y_j = 0$) the failure must have been caused by collision with a transmission from one of the active APs in $\mathcal{S}_j$. However, as there may be multiple hidden interferers, there may be no single AP that is always transmitting when $Y_j = 0$. Complicating the situation is the fact that an AP that transmits regularly when $Y_j = 1$ may not be a hidden interferer at all. This is because the transmission status of an AP can be highly positively correlated (because of CSMA/CA) with one or more hidden interferers. We first illustrate the types of statistical dependencies we need to address before presenting our algorithm.

First consider a scenario where a triplet of APs (i, j, k) lie in the direct interference range of AP l, i.e., i, j, and k are all are in $\mathcal{N}_l$. Further assume that there is no direct interference between i, j, and k. The transmission of any of these three APs will suppress the transmission of AP l and thus increase the (conditional) probability of transmission of the other APs in $\mathcal{N}_l$. Thus, the activation statuses of i, j, k are positively correlated. Now say that i, j are both hidden interferers for AP m. Then, even though transmissions of AP k may be correlated with transmission *failures* of AP m, due to the positive correlation statuses of i, j, k, AP k may not be a hidden interferer for node m. One possible scenario is that $\mathbb{P}(Y_m = 0 | X_k = 1, X_m = 1)$ may be even greater than $\mathbb{P}(Y_m = 0 | X_i = 1, X_m = 1)$ or $\mathbb{P}(Y_m = 0 | X_j = 1, X_m = 1)$. The upshot is that correlation-based approaches to determining hidden interferers, such those adopted in [7], [8], may not accurately distinguish true interferers from non-interferers. To address these issues we propose the following approach, based on *minimum hitting sets*.

First, given k observations, define

$$\mathcal{K}_j(k) = \{t \in \{1, 2, \ldots, k\} \mid Y_j(t) = 0\}$$

to be the sessions in which AP j's transmissions fail. For each

$t \in \mathcal{K}_j(t)$ define the set of candidate hidden interferers as

$$\mathcal{S}_j^t = \{i \in \mathcal{V} \mid i \neq j, X_i(t) = 1\}$$

and let

$$\hat{\mathcal{S}}_j(k) = \arg\min_{\mathcal{S} \subseteq \mathcal{V}} \{|\mathcal{S}| \mid \mathcal{S} \cap \mathcal{S}_j^t \neq \varnothing, \forall\, t \in \mathcal{K}_j(k)\},$$

where, if there are multiple minimizers, one is selected at random. The set $\hat{\mathcal{S}}_j(k)$ is a *minimum hitting set* of the candidate interferer sets $\{\mathcal{S}_j^t\}_{t \in \mathcal{K}_j(k)}$ per the following definition:

Definition 1 (Minimum Hitting Set) *Given a collection of subsets of some alphabet, a set which intersects all subsets in the collection in at least one element is called a "hitting set". A "minimum" hitting set is a hitting set of the smallest size.*

In Appendix C we prove the following lemma.

Lemma 1 $\lim_{k \to \infty} \mathbb{P}(\hat{\mathcal{S}}_j(k) = \mathcal{S}_j) = 1$.

Lemma 1 implies that when k is sufficiently large, identifying the minimum hitting set of the candidate interferer sets is equivalent to identifying the hidden interferer set of an AP. In the following, we drop the argument k in $\hat{\mathcal{S}}_j(k)$ for conciseness of notation and without fear of ambiguity.

Given k observations, our algorithm determines the minimum hitting set $\hat{\mathcal{S}}_j$ of $\{\mathcal{S}_j^t \mid t \in \mathcal{K}_j(k)\}$ for each $j \in \mathcal{V}$. By Lemma 1 we recover $G_H = (\mathcal{V}, \mathcal{E}_H)$ correctly, with high probability, for k sufficiently large. The following theorem provides an upper bound on the number of observations required so that $\hat{\mathcal{S}}_j = \mathcal{S}_j$ for every AP $j \in \mathcal{V}$ with high probability. Once the estimated minimum hitting set $\hat{\mathcal{S}}_j$ is obtained, the estimated hidden interference graph $\hat{G}_H$ is constructed by adding a directed edge from each AP in $\hat{\mathcal{S}}_j$ to AP j, i.e.,

$$\hat{\mathcal{E}}_H = \bigcup_{j \in \mathcal{V}} \left\{ (i,j) \mid i \in \hat{\mathcal{S}}_j \right\}.$$

The following theorem is proved in Appendix D.

Theorem 3 *Let* $\delta > 0$*, and let*

$$k \geq \frac{1}{-\log\left(1 - \frac{p^2(1-p)^s p_{\min}}{d^2}\right)} \left(\log(ns) + \log\frac{1}{\delta}\right)$$

Then, with probability at least $1-\delta$*,* $\hat{G}_H$ *equals* G_H*.*

The approach taken in the proof can be summarized as follows. For every AP j, we first upper bound the probability that the minimum hitting set obtained after k observations is not equal to the true minimum hitting set. This is equal to the probability that there exists at least one AP $i \in \mathcal{S}_j$ that is not included in $\hat{\mathcal{S}}_j$. By taking the union bound across all possible i and j, we obtain an upper bound on k.

1) Finding the minimum hitting set: In general, finding the minimum hitting set is NP-hard [12]. However, under the assumption that $s \ll n$, the minimum hitting set can be solved for in polynomial time. This is a regime appropriate to the large-scale wireless networks of interest in this paper. First we use the algorithm of Section III-A to identify G_D. Next, each AP in turn. For AP j test every subset of nonadjacent APs in G_D (and not including AP j) to determine whether it is a hitting set of $\{\mathcal{S}_j^t \mid t \in \mathcal{K}_j(k)\}$. Since the number of hidden interferers $s_j \leq s$, we start with the smallest possible hitting sets, i.e., $s_j = 1$. We increment the size of the testing subset by one, until a hitting set is achieved. In this way we find the minimum hitting set for the given $\{\mathcal{S}_j^t \mid t \in \mathcal{K}_j(k)\}$.

The worst situation for this incremental approach will be when s_j is as large as possible. By Assumption 1-(v) $s_j \leq s$. Recalling that $d_j - 1$ is the number of direct interferers (which can be eliminated from consideration), in this case the number of subsets we test is at most

$$\sum_{i=1}^{s} \binom{n-d_j}{i} = \beta_3 (n-d_j)^s$$

for some bounded constant β_3. In other words, the number of subsets we need to test is upper bounded by $O(n^s)$ [13].

2) A special scenario: $s = 1$*:* In certain situations hidden interferers are rare. For example, they are rare when the threshold for carrier sensing is set low which makes carrier sensing range large. It can be reasonable to assume in such settings that there is at most one hidden interferer for any AP, i.e., $s = 1$. In this subsection we show that in this case, the minimal hitting set algorithm presented above specializes to a correlation-based method.

Specifically, assume that the single interferer in G_H for AP j is AP i. Then, whenever $t \in \mathcal{K}_j(k)$, X_i must equal 1. Of course, transmitter i may not be the only active transmitter when $Y_j = 0$. However, conditioned on the event $Y_j = 0$, $\mathbb{P}(X_i = 1 | Y_j = 0) = 1$, which is strictly larger than $\mathbb{P}(X_l = 1 | Y_j = 0)$ for any other AP $l \neq j$. Therefore, thresholding $\mathbb{P}(X_l = 1 | Y_j = 0)$ reliably detects the true hidden interferers. Take k observations of $\mathbf{X}$ and $\mathbf{Y}$. For each AP j, we count the number of times that an AP is active for all $t \in \mathcal{K}_j(k)$. The AP with the highest count is the hidden interferer of AP j. This is equivalent to searching for the minimum hitting set, since the hidden interferer appears in every candidate interferer set. We can show that it is also equivalent to the correlation based approach in [8]. When $s > 1$, thresholding the empirical frequency of APs being active for $t \in \mathcal{K}_j(k)$ is different from the minimum hitting set method.

The hidden interference graph learning problem is related to the partially labeled multi-class classification problem discussed in [14]. In [14] a problem is studied wherein each object is given a candidate set of labels, only one of which is correct. In our problem, for each AP j, from the candidate hidden interferer set $\mathcal{S}_j^t$ we want to extract the indexes (labels) of the actual interferers. When $s = 1$, the only true hidden interferer is the correct "label" for AP j. In general, our problem differs from [14] in that there may be more than one true hidden interferers (correct labels) for an AP in our problem. Therefore, our framework generalizes the partially labeled multi-class classification problem, and implies broader applicability of our approach.

D. A Minimax Lower Bound for Inference of $G_H = (\mathcal{V}, \mathcal{E}_H)$

Finally, we provide a lower bound on the number of observations required to recover the underlying hidden interference graph $G_H = (\mathcal{V}, \mathcal{E}_H)$. In Appendix E we prove the following theorem.

Theorem 4 *If*

$$d \leq c_1 n, \quad s-1 \leq c_2 n, \quad 2c_1 + c_2 < 1 \tag{3}$$

and

$$k \leq -\frac{\log(2c_1(\frac{1}{2c_1+c_2}-1)n)}{\left(\frac{1-(1-p)^d}{d}\right)^2 (1-p)^{s-1}\log(1-p_{\min})}$$

then for any α, $0 < \alpha < 1/8$,

$$\min_{\hat{G}_H \in \mathcal{G}_s} \max_{G \in \mathcal{G}_d \times \mathcal{G}_s} \mathbb{P}(\hat{G}_H \neq G_H; G) \geq$$

$$\frac{\sqrt{2c_1(\frac{1}{2c_1+c_2}-1)n}}{1+\sqrt{2c_1(\frac{1}{2c_1+c_2}-1)n}}\left(1-2\alpha-\sqrt{\frac{2\alpha}{\log(2c_1(\frac{1}{2c_1+c_2}-1)n)}}\right).$$

The conditions of (3) ensure that the max degree d and bound on the number of hidden interferers s both scale at most linearly in n, subject to a constraint on the joint scaling. The approach to this result is the same as we followed for the proof of Theorem 2. We reduce the original problem to an M-ary hypothesis test and show that, asymptotically, the lower bound has the same order as the upper bound.

Remarks: Since the distribution of $\mathbf{Y}$ depends on the underlying direct interference graph as well as on the hidden interference graph, the lower bound is over all possible interference graphs G. As d increases, the lower bound in Theorem 4 is of order $\Omega\left(-\frac{d^2}{\log(1-p_{\min})(1-p)^{s-1}}\log n\right)$. Since $\log(1-p_{\min})$ can be approximated as $-p_{\min}$ when $p_{\min}$ is small, the lower bound is of the same order as the upper bound provided in Theorem 3. Therefore, the bounds are tight and our algorithm is optimal.

IV. Main Results for Random Channels

In the earlier sections we study a setting where the channel states between pairs of APs and between APs and clients are static. However, the time-varying nature of the wireless medium means that these modeling assumptions may be optimistic. The channel state between a pair of communication devices is often better modeled as random. One implication is that the sensed signal strength at an AP – or the interference level at a client's receiver – is also random. We now try to capture this randomness in our model. We modify our earlier model by introducing independent random noise into the carrier sensing channel between APs and into the communication channel between APs and clients. The inference algorithms must be adjusted to accommodate these additional sources of randomness.

A. System Model with Random Channel States

Because of the randomness of channel states between a pair of APs, two APs may not always sense each other, even though they are within each other's carrier sensing range. In order to capture this randomness, we define q_{ij} to be the probability that AP j *cannot* sense AP i when AP i is transmitting, e.g., the signal strength received from AP i is below the carrier sensing threshold at AP j. As before we assume reciprocity of channels between APs so $q_{ij} = q_{ji}$ for all $i, j \in \mathcal{V}$. We again categorize edges into direct and hidden interference sets, but the distinction is less hard than before. For every $(i,j) \in \mathcal{E}_D$ we assume

$$q_{ij} \leq q_{\text{low}},$$

for some $q_{\text{low}} \geq 0$. This assumption reduces to that of the static scenario of Sec. III if we set $q_{\text{low}} = 0$. On the other hand, for $(i,j) \in \mathcal{E}_H = \mathcal{E} \backslash \mathcal{E}_D$ we assume that

$$q_{ij} \geq q_{\text{up}},$$

for some $q_{\text{up}} \leq 1$; earlier we assumed that $q_{\text{up}} = 1$. Further, $0 \leq q_{\text{low}} < q_{\text{up}} \leq 1$ and both probabilities may scale with d, n.

The definition of p_{ij} remains the same as before. Recall that p_{ij} is the probability that AP i causes transmission failure of AP j, and $p_{ij} \geq p_{\min}$ for all $(i,j) \in \mathcal{E}_H$. In the static scenario, the p_{ij} mainly model the random locations of clients that communicate with APs. In the random channel state setting, the p_{ij} can also capture randomness in channel states. In Section IV-C we use this flexibility to model transmission failures due to deep fades. We define $p_{\emptyset j}$ to be the probability that the transmission from AP j fails due to poor channel conditions. While these failures are not caused by any active AP, this is not necessarily known and thus we need to incorporate this cause of failure into the interference set $\mathcal{S}_j$. Thus, we consider $\emptyset$ to be a "virtual" interferer and define the augmented set of interferers as $\mathcal{S}_j \cup \emptyset$. We will assume that $p_{\emptyset j}$ is independent of the status of all APs in the network.

B. Direct Interference Graph Inference

Since in the random channel state setting directly interfering APs may not always detect each other, we keep count of the number of co-occurring transmissions and require this number to exceed a threshold to declare existence of an edge. Assuming a length-k observation interval define

$$\hat{I}_{ij} = \frac{\sum_{t=1}^{k} \mathbf{1}\{X_i(t) = X_j(t) = 1\}}{k},$$

where $\mathbf{1}\{\cdot\}$ is the indicator function. If $\hat{I}_{ij} \geq \epsilon$ we declare that APs i and j *do not* interfere. Thus, to get the estimated graph $\hat{G}_D$ we start with a fully connected graph and remove all edges (i,j) such that $\hat{I}_{ij} \geq \epsilon$. As we see in Theorem 5, the value of ϵ depends on q_{up}, q_{low}, as well as p, d.

To derive our results we require the following assumption in addition to those of Assumption 1:

Assumption 2

(i) $q_{\text{up}} \geq c_3 d^2 q_{\text{low}}$, *where* $c_3 \geq 1$ *is a constant.*

Assumption 2 provides a sufficient condition to distinguish the edges in $\mathcal{E}_D$ from the non-edges. Intuitively, we expect to observe that a pair of APs not connected by an edge in $\mathcal{E}_D$ to transmit simultaneously more frequently than those connected by edges in $\mathcal{E}_D$. In the worst scenario, two APs i, j, $(i,j) \notin$

$\mathcal{E}_D$, have $d-1$ neighbors each in G_D, while another pair $(i', j') \in \mathcal{E}_D$ only have each other as the neighbor. Assumption 2 ensures that co-occurring transmissions of i, j occur more often than those of (i', j') even in the worst scenario.

In Appendix F we prove the following achievability result:

Theorem 5 *Under Assumptions 1 and 2, if*

$$\epsilon = \frac{p^2 q_{\mathrm{up}}(1-1/c_3)}{2d^2},$$

and

$$k \geq \frac{1}{\epsilon^2}\left(\log n + \frac{1}{2}\log\frac{1}{\delta}\right)$$

then, with probability at least $1-\delta$, the estimated interference graph $\hat{G}_D = (\mathcal{V}, \hat{\mathcal{E}}_D)$ is equal to G_D.

We note that the upper bound on the number of observations k for reliable estimation of the direct inference graph is $O(d^4 \log n)$. This is worse than the minimax lower bound in static setting. This is natural since randomness in communication channels corrupts the observations, making it more difficult to distinguish true interferers from non-interferers.

C. Hidden Interference Graph Inference

As discussed in the section introduction, we can now model transmission failures due to both deep fades and interference. The following assumption says that the former are not too likely. If deep fades were more likely to cause transmission failures than interference, it would be extremely hard to distinguish such failures from other hidden interferers (and would really perhaps not be very useful to determine the interference graph at all).

Assumption 3

(i) *We assume that $p_{\emptyset j}$ is upper bounded by $p_\emptyset$ for any j and further that $p_\emptyset < p_{\min}$.*

Besides $\mathcal{S}_j \cup \emptyset$, direct interferers may also collide with node j if they cannot sense each other and yet happen to transmit simultaneously. The existence of three different interference sources for transmission collision (deep fades, hidden interference, direct interference) makes the problem much more complicated compared to the static channel setting. The minimum hitting set method cannot be applied in this scenario. One obvious reason is that channel states cannot be measured, so $\emptyset$ can be used to explain every collision, i.e., $\emptyset$ is the minimum hitting set for every AP j. Therefore, $\mathcal{S}_j$ cannot be attained by the minimum hitting set method.

In the static channel setting discussed in Sec. III-C, the estimated $\hat{G}_D$ is not required as an input for the hidden interference graph estimation. Since the channel state between each pair of APs is static, if the APs are neighbors in G_D they are guaranteed not to transmit simultaneously. This is not the situation in the random channel state setting since a pair of neighboring APs in G_D can transmit simultaneously. In order to isolate the effect of the direct interferers from that of the hidden interferers our algorithm focuses on the collision events of an AP when none of its direct interferers are active. Determining this subset of events requires knowledge of G_D. Thus our algorithm starts by estimating $\hat{G}_D$ using the algorithm of Sec. IV-B. That estimate is then used as an input to estimate G_H.

In the following we assume that the estimate $\hat{G}_D$ has already been made. The algorithm for inferring the hidden interference graph makes use of the following two indicator functions, both defined for any $\mathcal{S}$ and AP pair i, j where $\mathcal{S} \subseteq \mathcal{V} \backslash \{i, j\}$. The first is

$$I_{ij,\mathcal{S}}(t) = \mathbf{1}\{X_{\mathcal{S}}(t) = \mathbf{0}, X_{\hat{\mathcal{N}}_j}(t) = \mathbf{0}, X_i(t) = X_j(t) = 1\},$$

where $\hat{\mathcal{N}}_j$ is the estimate of $\mathcal{N}_j$ deduced from $\hat{G}_D$. This indicator selects the sessions in which nodes i and j both transmit, the (estimated) neighbors of node j do not transmit, and nor do any nodes in the set $\mathcal{S}$. The second indicator

$$I^0_{ij,\mathcal{S}}(t) = I_{ij,\mathcal{S}}(t) \cdot \mathbf{1}\{Y_j(t) = 0\}$$

further narrows the sessions selected by the first indicator to ones in which node j transmission fails.

We define the estimate

$$\hat{p}_{ij} := \min_{\substack{\mathcal{S} \subseteq \mathcal{V} \backslash \hat{\mathcal{N}}_j \cup \{i,j\} \\ |\mathcal{S}| \leq s}} \frac{\sum_{t=1}^{k} I^0_{ij,\mathcal{S}}(t)}{\sum_{t=1}^{k} I_{ij,\mathcal{S}}(t)}.$$

Searching for the optimal $\mathcal{S}$ to attain $\hat{p}_{ij}$ is to isolate the effect of hidden interferers other than i, together with that of $\hat{\mathcal{N}}_j$, from that of i. Thus, if $\hat{\mathcal{N}}_j = \mathcal{N}_j$ (which would be the case if $\hat{G}_D = G_D$), i and $\emptyset$ are the only possible sources for the counted collisions.

If $i \in \mathcal{S}_j$, then $\hat{p}_{ij}$ is an estimate of

$$\mathbb{P}(Y_j = 0 | X_j = 1, X_i = 1, X_{\mathcal{S}_j \backslash i} = \mathbf{0}, X_{\hat{\mathcal{N}}_j} = \mathbf{0}).$$

If $i \notin \mathcal{S}_j$, then $\hat{p}_{ij}$ is an estimate of

$$\mathbb{P}(Y_j = 0 | X_j = 1, X_i = 1, X_{\mathcal{S}_j} = \mathbf{0}, X_{\hat{\mathcal{N}}_j} = \mathbf{0}).$$

The importance of this estimate results from the following lemma proved in Appendix G.

Lemma 2 *For any $j \in \mathcal{V}$ there exists a unique set $\mathcal{S}_j$ such that, for any $i \in \mathcal{S}_j$,*

$$\mathbb{P}(Y_j = 0 | X_j = 1, X_i = 1, X_{\mathcal{S}_j \backslash i} = \mathbf{0}, X_{\mathcal{N}_j} = \mathbf{0}) \geq p_{ij}. \quad (4)$$

Further, for any $i \in \mathcal{V} \backslash \{\mathcal{S}_j \cup \mathcal{N}_j\}$,

$$\mathbb{P}(Y_j = 0 | X_j = 1, X_i = 1, X_{\mathcal{S}_j} = \mathbf{0}, X_{\mathcal{N}_j} = \mathbf{0}) = p_{\emptyset j}. \quad (5)$$

Thus, if $\hat{\mathcal{N}}_j = \mathcal{N}_j$ the lemma tells us that the value of $\hat{p}_{ij}$ indicates whether $i \in \mathcal{S}_j$.

Say that for some pair (i, j) of APs, $\hat{p}_{ij} \geq \delta$ where δ is equal to the average value of $p_{\min}$ and $p_\emptyset$. The algorithm is to add the edge (i, j) to the estimated hidden interference graph $\hat{G}_H$ for all such pairs.

In Appendix H we prove the following:

Theorem 6 *Under Assumptions 1, 2, and 3, if*

$$\epsilon = \frac{p^2 q_{\text{up}}(1-1/c_3)}{2d^2},$$
$$\delta = \frac{p_{\min} + p_{\emptyset}}{2},$$
$$\delta_w = \frac{(p_{\min} - p_{\emptyset})(1-p)^s(1-q_{\text{low}})^d p^2}{(1+\delta)d^2},$$

and

$$k \geq \max\left\{\frac{1}{\epsilon^2}\left(\log n + \frac{1}{2}\log\frac{2}{\gamma}\right), \frac{2}{\delta_w^2}\left((s+2)\log n + \log\frac{4}{\gamma}\right)\right\}, \tag{6}$$

then the estimated hidden interference graph $\hat{G}_H$ is equal to G_H with probability $1-\gamma$ after k observations.

The first term in (6) guarantees that k is sufficient large to infer G_D reliably, and the second term ensures that $\hat{p}_{ij}$ concentrates around its mean value. Asymptotically, the second term in the bound of k dominates the first term, and the upper bound on k for reliable estimation of the hidden inference graph is $O\left(\frac{d^4}{(1-p)^{2s}(1-q_{\text{low}})^{2d}}\log n\right)$.

V. Simulation Results

In this section we present simulation results for the following model. Access points and clients are deployed over a rectangular area that can be partitioned into square cells $50m$ on a side. In each cell an AP and several clients are placed uniformly at random. The locations of APs and clients are fixed throughout the period of observation. Each client is associated with the nearest AP. By generating the networks in this manner the traffic load across APs is relatively balanced.

For each client in its communication range, the AP generates an independent traffic flow for it according to a Poisson process with 0.1 packet per session. In order to predict received signal strength between APs and clients, we employ the log-distance path loss model. In this model, received power (in dBm) at unit distance (in meters) from the transmitter is given by:

$$\Gamma(l) = \Gamma(l_0) - 10\alpha\log(l/l_0) + X_\sigma. \tag{7}$$

In the above, $\Gamma(l_0)$ is the signal strength at reference distance l_0 from the transmitter, α is the path loss exponent, and X_σ represents a Gaussian random variable with zero mean and variance σ^2 in db. We choose l_0 to be 1km, σ^2 to be 3db, and α to be 4. We also assume that the "shadowing" (represented by X_σ) between any pair of APs and AP-client pair is fixed throughout the period of observation. Thus the underlying interference graph is constant within the observation period.

The system operates according to IEEE 802.11 CSMA/CA protocol. In comparison to the statistical assumptions made to derive the theoretical learning bounds, this setup mimics wireless networks more closely.

We first study the direct interference estimation algorithm of Section III-A. We fix the carrier sensing range for the APs to be 70m and simulate five networks, respectively consisting of 4×7, 4×8, 4×9 and 4×10 cells. For each network size,

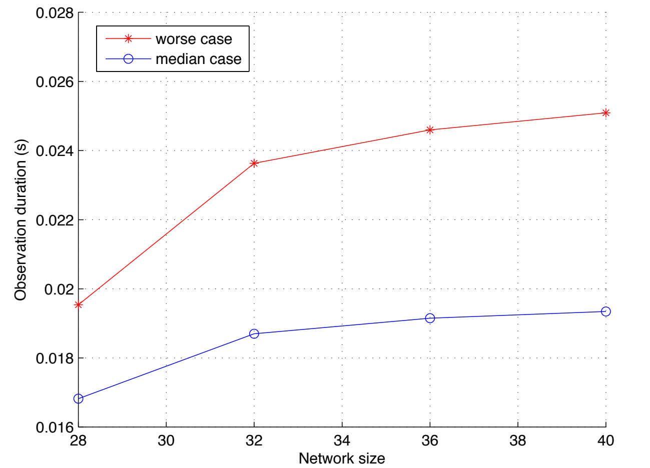

Fig. 2. The observation duration to recover direct interference graph for networks with $d = 5$ and $s = 2$, plotted as a function of the number of cells in the network.

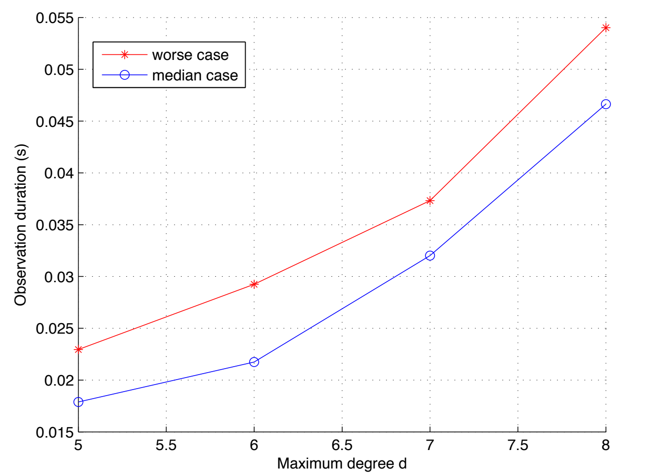

Fig. 3. The observation duration to recover direct interference graph for networks with 4×8 cells, plotted as a function of maximum degree d.

we randomly generate 10 topologies, i.e., randomly placed APs and clients. For each topology, we use the algorithm to recover G_D five times with different (randomly generated) traffic traces. We report the average observation duration (i.e., number of sessions) required to recover the direct interference graph for each topology. The median observation time over the 10 topologies is plotted in Fig. 2. The worst case is also plotted.

Next, we fix the network size to be 4×8 cells and randomly generate topologies. For each topology, we vary the carrier sensing range to be – 70m, 80m, 90m, 100m – and randomly generate topologies. We check the maximum degree d (the number of direct interference edges per node) and s (the number of hidden interferers per node) for each topology. We select ten topologies with $s = 2$ for each d varying from 5 to 8 and each carrier sensing range. We then simulate each case. We plot the median observation time required to recover G_D

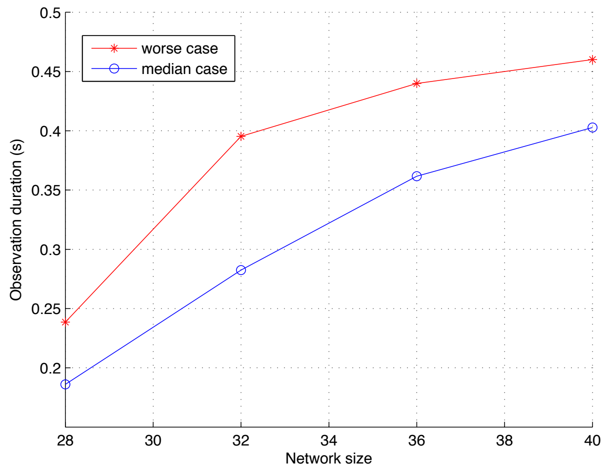

Fig. 4. The observation duration to recover hidden interference graph for networks with $d = 5, s = 2$, plotted as a function of the number of cells in the network.

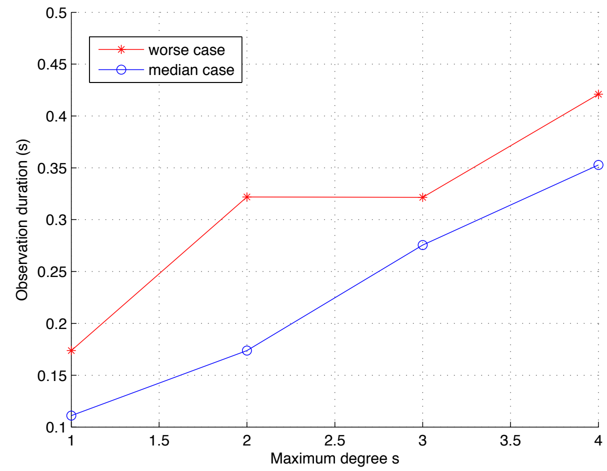

Fig. 5. The observation duration required to hidden interference graph for topologies with 4×8 cells, plotted as a function of maximum degree s.

as a function of the maximum degree of the network in Fig. 3. The worst case is also plotted in the same figure.

The plots show the scaling that was predicted by the theory. The necessary observation duration scales sub-linearly (logarithmically) with network size n, and super-linearly (quadratically) in d.

Next, we study estimation of the hidden interference graph. In Fig. 4 we plot the observation duration required to correctly identify the minimum hitting set for each node and recover the hidden interference graph as a function of network size. The same simulation conditions hold as were described in the discussion of Fig. 2. For this algorithm we again observe that the necessary observation duration scales sub-linearly (logarithmically) with network size n.

In Fig. 5 we examine the dependence of the necessary observation duration on s for hidden interference graph inference. In these simulations we fix the network size to be 4×8 cells and the carrier sensing range to be 80m. We randomly generate topologies with fixed $d = 5$, and s varying from 1 to 4. The necessary observation duration as a function of s is then plotted. We observe that the observation duration increases as s increases. However, it scales in a complicated form. This is because the hidden interferers cause transmission collisions in the networks, which actually increases the traffic intensity at each AP.

We also examine the robust direct interference estimation algorithm of Section IV-B. Unlike the static scenario where we assume that the shadowing (i.e., X_σ) between a pair of APs is a constant during the observation period, under the random channel state setting, we let X_σ in (7) be an i.i.d Gaussian random variable with zero mean and 3dB variance across the observations. We repeat the simulation procedure under this random channel setting, and evaluate the performance of the robust direct interference estimation algorithm with $\epsilon = 0.007$.

We plot the median average observation time over 10 topologies required to correctly recover G_D in Fig. 6. The worst case is also plotted. Next, we plot the median and worst

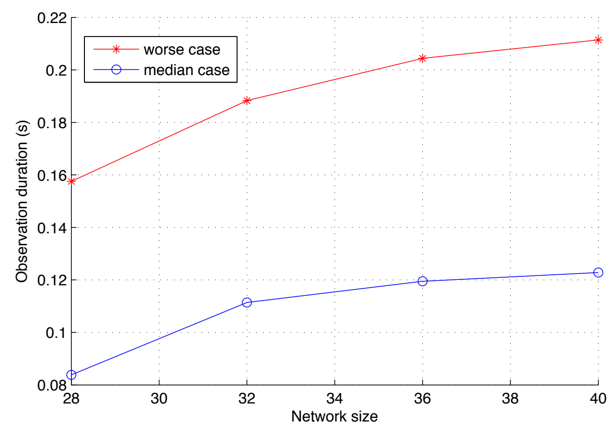

Fig. 6. The observation duration to recover direct interference graph for networks with $d = 5$ and $s = 2$ under random channel state setting, plotted as a function of the number of cells in the network.

observation times required to recover G_D as a function of the maximum degree of the network in Fig. 7. The plots show the scaling that was predicted by the theory under the random channel states setting. Again, we observe that he necessary observation duration scales sub-linearly with network size n, and super-linearly in d. Compared with the simulation results under the static setting, we note that it takes much longer observation time to successfully recover G_D in the random channel scenario. This is consistent with the theoretical results on the upper bounds.

VI. Conclusions

In this paper, we proposed passive interference learning algorithms and analyzed their learning bounds for both static and random settings. For the static setting, we first upper bounded the number of measurements required to estimate the direct interference graph. Then, we provided a minimax lower bound by constructing a sequence of networks and transforming it into an M-ary hypothesis test. The lower bound matches the

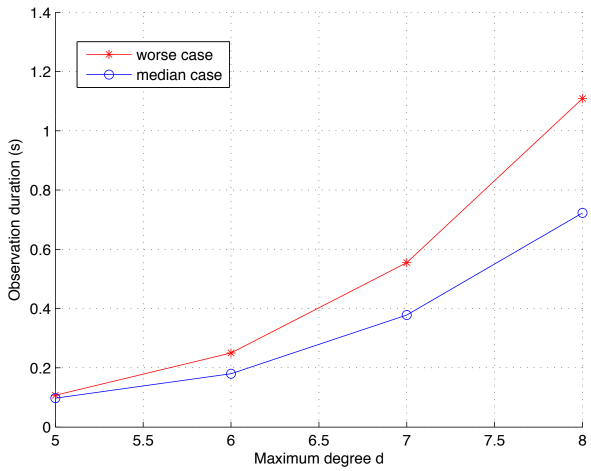

Fig. 7. The observation duration to recover direct interference graph for networks with 4×8 cells under random channel state setting, plotted as a function of maximum degree d.

upper bound (up to a constant), thus the bound is tight and the algorithm is asymptotically optimal. We then analyzed the hidden interference graph estimation, and provided matched lower and upper bounds following similar approaches. For the random setting, we proposed robust graph inference algorithms to estimate the direct and hidden interference, and presented upper bounds on the number of measurements required for reliable estimations. We also presented an experimental study that lends support to the theoretical analysis.

Appendix

A. Proof of Theorem 1

Consider any two nonadjacent nodes i, j in $G_D = (\mathcal{V}, \mathcal{E}_D)$. Let $\mathcal{N}_{ij} = \mathcal{N}_i \cup \mathcal{N}_j$, and $\mathcal{N}_{i\backslash j} = N_i \cap N_j^c$. nodes that are the neighbors of node i but are not neighbors of node j.

Under the CSMA/CA protocol, transmitter i only contends for the channel when $Q_i = 1$. For ease of analysis, in this proof, we assume that every transmitter competes for the channel at the beginning of each session no matter whether or not $Q_i = 1$. Then, if transmitter i gets the channel, it starts transmit if $Q_i = 1$. Otherwise, even thought node i has the opportunity to transmit, it keeps silent. The latter event leaves the opportunity open for one of its neighbors with a longer back-off time to transmit. Since this does not change the outcome of transmission pattern given that the back-off times of those transmitters with nonempty queues are fixed, the statistics of $\mathbf{X} := (X_1, X_2, \ldots, X_n)$ stay the same under this interpretation.

Define $T_{\mathcal{N}_i}$ as the minimum back-off time of the nodes in the set $\mathcal{N}_i$. Then,

$$\mathbb{P}\left(T_i < T_{\mathcal{N}_i}, T_j < T_{\mathcal{N}_{j\backslash i}}\right) = \frac{\binom{|\mathcal{N}_{ij}|+2}{|\mathcal{N}_i|+1} \cdot |\mathcal{N}_i|! \cdot |\mathcal{N}_{j\backslash i}|!}{(|\mathcal{N}_{ij}|+2)!} \quad (8)$$

$$\geq \frac{1}{(d_i+1)(d_j+1)} \geq \frac{1}{d^2}.$$

The probability in (8) is obtained in this way: Considering nodes i, j and their neighbors, there are $|\mathcal{N}_{ij}|+2$ nodes in total, and there are $(|\mathcal{N}_{ij}|+2)!$ orderings of their back-off times. Among these orderings, $\binom{|\mathcal{N}_{ij}|+2}{|\mathcal{N}_i|+1} \cdot |\mathcal{N}_i|! \cdot |\mathcal{N}_{j\backslash i}|!$ orderings correspond to $T_i < T_{\mathcal{N}_i}, T_j < T_{\mathcal{N}_{j\backslash i}}$. Such ordering can be obtained in this way: suppose these $|\mathcal{N}_{ij}|+2$ nodes are ordered according to their back-off times. Node i and its neighbors take $|\mathcal{N}_i|+1$ positions in the ordering. There are $\binom{|\mathcal{N}_{ij}|+2}{|\mathcal{N}_i|+1}$ different combinations of positions. Since $T_i < T_{\mathcal{N}_i}$, node i takes the first position out of the chosen $|\mathcal{N}_i|+1$ positions, the remaining $|\mathcal{N}_i|$ positions are for its neighbors. This results in $|\mathcal{N}_i|!$ orderings for each combinations of positions. Node j and the nodes in $\mathcal{N}_{j\backslash i}$ take the rest of the positions, where node j takes the first. This gives us the factor $|\mathcal{N}_{j\backslash i}|!$.

When $T_i < T_{\mathcal{N}_i}$, and $Q_i = 1$, based on the protocol, node i gets the channel and thus $X_i = 1$. At the same time, transmission from all nodes in $\mathcal{N}_i$ are suppressed. Therefore, for node j, if $T_j < T_{\mathcal{N}_{j\backslash i}}$ and $Q_j = 1$, node j also gets a channel. Thus, we have $X_i = X_j = 1$. Since other scenarios result in $X_i = X_j = 1$, we have the following

$$\begin{aligned}
&\mathbb{P}(X_i = 1, X_j = 1) \\
&\quad \geq \mathbb{P}\left(T_i < T_{\mathcal{N}_i}, T_j < T_{\mathcal{N}_{j\backslash i}}, Q_i > 0, Q_j > 0\right) \\
&\quad = \mathbb{P}\left(T_i < T_{\mathcal{N}_i}, T_j < T_{\mathcal{N}_{j\backslash i}}\right) \mathbb{P}(Q_i > 0, Q_j > 0) \geq \frac{p^2}{d^2} \\
&\mathbb{P}(\text{edge } (i,j) \text{ is not removed with one observation } \mathbf{X}) \\
&\quad = 1 - \mathbb{P}(X_i = 1, X_j = 1) \leq 1 - \frac{p^2}{d^2}.
\end{aligned}$$

Define $\mathcal{A}_{ij}$ as the event that edge (i, j) is not removed after k observations. Then, the probability that, after k observations, the graph cannot be identified successfully is

$$\mathbb{P}(\hat{G}_D \neq G_D) = \mathbb{P}(\cup_{(i,j)\notin\mathcal{E}_D} \mathcal{A}_{ij}) \leq \binom{n}{2}\left(1 - \frac{p^2}{d^2}\right)^k.$$

The inequality follows from the fact that the number of nonadjacent pairs in $G_D = (\mathcal{V}, \mathcal{E}_D)$ is upper bounded by $\binom{n}{2}$. Under the assumption that $d \ll n$, this is a good approximation for the nonadjacent pairs in G_D.

B. Proof of Theorem 2

First, we define $\mathcal{G}_d$ as the set of graphs consisting of n nodes that have maximum degree $d-1$. We are going to construct a collection of $M+1$ graphs $\{G_{D0}, G_{D1}, \ldots, G_{DM}\}$ where $G_{Di} \in \mathcal{G}_d$ for all i. We denote the distribution of transmission patterns $\mathbf{x} \in \{0,1\}^n$ for each of these graphs as $P_0(\mathbf{x}), P_1(\mathbf{x}), \ldots, P_M(\mathbf{x})$, respectively. Define $\mathbb{P}(\mathcal{A}; G_D)$ to be the probability of event $\mathcal{A}$ occurring where the underlying direct interference graph is $G_D \in \mathcal{G}_d$. A metric of interest is the edit or "Levenshtein" distance between a pair of graphs. This is the number of operations needed to transform one graph into the other. We denote the Levenshtein distance between G_{Di} and G_{Dj} as $D_L(G_{Di}, G_{Dj})$. Then, we use the following theorem to obtain the lower bound. The theorem is adapted from [15].

Theorem 7 *Let $k \in \mathbb{Z}^+$, $M \geq 2$, $\{G_{D0}, \ldots, G_{DM}\} \in \mathcal{G}_d$ be such that*

- $D_L(G_{Di}, G_{Dj}) \geq 2r$, *for* $0 \leq i < j \leq M$, *where* D_L *is a semi-distance,*
- $\frac{k}{M}\sum_{i=1}^{M} D_{KL}(P_i \| P_0) \leq \alpha \log M$, *with* $0 < \alpha < 1/8$.

Then

$$\begin{aligned}\inf_{\hat{G}_D \in \mathcal{G}_d} \sup_{G_D \in \mathcal{G}_d} & \mathbb{P}(D_L(\hat{G}_D, G_D) \geq r; G_D) \\ &\geq \inf_{\hat{G}_D \in \mathcal{G}_d} \max_i \mathbb{P}(D_L(\hat{G}_D, G_{Di}) \geq r; G_{Di}) \\ &\geq \frac{\sqrt{M}}{1+\sqrt{M}}\left(1 - 2\alpha - \sqrt{\frac{2\alpha}{\log M}}\right) > 0\end{aligned}$$

In the following, we construct the $M+1$ graphs $\{G_{D0}, G_{D1}, \ldots, G_{DM}\}$ and leverage Theorem 7 to obtain a lower bound on k. We first construct G_{D0} and characterize $P_0(\mathbf{x})$. Then, we construct the rest of the M graphs by perturbing G_{D0}. These graphs will be symmetric in the sense that $D_{KL}(P_i \| P_0)$ will be the same for $1 \leq i \leq M$. We calculate $D_{KL}(P_1 \| P_0)$ and then lower bound the number of required observations k.

1) G_{D0} and its transmission pattern distribution $P_0(\mathbf{x})$: We let graph G_{D0} consist of $\lceil n/(d-1) \rceil$ disconnected cliques. The first $\lfloor n/(d-1) \rfloor := m_0$ cliques are fully connected subgraphs with $d-1$ nodes, as shown in Fig. 8. The last clique is also full connected, and the number of nodes is less than $d-1$. In the following analysis, we focus on the first m_0 cliques.

Denote $\mathcal{C}_m$ as the set of nodes in the mth clique and let $\mathbf{X}_{\mathcal{C}_m}$ be the restriction of the transmission pattern $\mathbf{X}$ to the nodes in the clique ($\mathbf{X}_{\mathcal{C}_m}$ is a set of subvectors, $m = 1, 2, \ldots, m_0$ that partition $\mathbf{X}$). Define $\mathbf{e}_i$ as the unit vector of dimension $d-1$ whose ith element is one. Define $\mathbf{0}$ as the $d-1$ dimensional all-zeros vector. Due to the fully connected structure of the cliques, no more than one node in any $\mathcal{C}_m$ can transmit at any time. We denote the set of all feasible transmission patterns for clique m as $\mathcal{X}_m := \{\mathbf{0}, \mathbf{e}_1, \ldots, \mathbf{e}_{|\mathcal{C}_m|}\}$. If $\mathbf{X}_{\mathcal{C}_m} = \mathbf{0}$, then no node in $\mathcal{C}_m$ is transmitting in that session. For each individual clique, this event happens only when none of the nodes in that clique has traffic to send, i.e.,

$$\mathbb{P}(\mathbf{X}_{\mathcal{C}_m} = \mathbf{0}; G_{D0}) = (1-p)^{d-1}. \tag{9}$$

Otherwise, when at least one AP has a packet to send, the channel will not idle. Because the Q_is and T_is are i.i.d. across nodes, and the clique is fully connected, each node in that clique has the same probability of occupying the channel. Thus, for any j, $j \in \{1, 2 \ldots, |\mathcal{C}_m|\}$,

$$\mathbb{P}(\mathbf{X}_{\mathcal{C}_m} = \mathbf{e}_j; G_{D0}) = \frac{1-(1-p)^{d-1}}{d-1} \triangleq q, \tag{10}$$

Since the behavior of the cliques are independent, we have

$$P_0(\mathbf{x}) = \prod_{m=1}^{m_0+1} \mathbb{P}(\mathbf{X}_{\mathcal{C}_m} = \mathbf{x}_{\mathcal{C}_m}; G_{D0}). \tag{11}$$

2) Construct $M = n$ graphs: In this subsection, we construct a sequence of graphs G_{D1}, G_{D2}, $\ldots$, G_{DM}. We will wan to construct n graphs, i.e., $M = n$. We construct each graph by picking a pair of nodes from distinct cliques in first m_0 cliques in graph G_{D0}. We add an edge between the selected pair. We leave the last ((m_0+1)th) clique unmodified for $G_{D1}, \ldots, G_{DM}$. We can construct $\binom{m_0}{2}(d-1)^2 = (d-1)^2 m_0(m_0-1)/2$ different graphs in this manner. Under the assumption that

$$d-1 \leq n/2, \tag{12}$$

and the fact that $m_0 = \lfloor \frac{n}{d-1} \rfloor \geq \frac{n}{d-1} - 1$, we have

$$\begin{aligned}\frac{(d-1)^2}{2} m_0(m_0-1) &\geq \frac{(d-1)^2}{2}\left(\frac{n}{d-1} - 1\right)\left(\frac{n}{d-1} - 2\right) \\ &= (d-1)^2 - \frac{3n}{2}(d-1) + \frac{n^2}{2}\end{aligned} \tag{13}$$

This is a quadratic function of $d-1$ for any fixed n. We want to have $M = n$ and the value of (13) equals n if

$$d-1 = \frac{3n \pm \sqrt{n^2+16n}}{4}.$$

Since $d-1 \leq n$, and $\sqrt{n^2+16n} > n$, only the smaller solution is feasible. When $n \geq 5$, we have $\frac{3n-\sqrt{n^2+16n}}{4} > 1$, thus the assumption $d-1 \geq 1$ is satisfied. Meanwhile, since

$$\frac{3n-\sqrt{n^2+16n}}{4} \leq \frac{3n-n}{4} = \frac{n}{2}, \tag{14}$$

it is a tighter constraint on $d-1$ than (12). Therefore, under the condition that

$$n \geq 5, \quad d-1 \leq \frac{3n-\sqrt{n^2+16n}}{4},$$

we have $(d-1)^2 m_0(m_0-1)/2 \geq n$ and thus we can always pick n graphs that are perturbations of G_{D0} in the above sense. We note that for each of these graphs $D_L(G_{D0}, G_{Di}) = 1$, and $D_L(G_{Di}, G_{Dj}) = 2$ for any $0 < i, j \leq M$ where $i \neq j$.

3) G_{D1} and its transmission pattern distribution $P_1(\mathbf{x})$: We now calculate $P_1(\mathbf{x})$, which differs from $P_0(\mathbf{x})$ due to the added constraint resulting from the additional edge. Due to the symmetric construction of the M graphs, without loss of generality we concentrate on a single graph. Let G_{D1} be the graph formed from G_{D0} by connecting the ith node in $\mathcal{C}_1$ to the jth node in $\mathcal{C}_2$ with an edge. Since the remaining $m_0 - 1$ cliques are unchanged, the distribution of their transmission patterns $\mathbf{X}_{\mathcal{C}_m}$ is the same as under G_{D0}. Furthermore, they are independent of each other and of $(\mathbf{X}_{\mathcal{C}_1}, \mathbf{X}_{\mathcal{C}_2})$. Thus we express the transmission pattern distribution under G_{D1} as

$$\begin{aligned}P_1(\mathbf{x}) &= \mathbb{P}(\mathbf{X}_{\mathcal{C}_1} = \mathbf{x}_{\mathcal{C}_1}, \mathbf{X}_{\mathcal{C}_2} = \mathbf{x}_{\mathcal{C}_2}; G_{D1}) \\ &\cdot \prod_{m=3}^{m_0+1} \mathbb{P}(\mathbf{X}_{\mathcal{C}_m} = \mathbf{x}_{\mathcal{C}_m}; G_{D0}).\end{aligned} \tag{15}$$

We want to calculate the KL-divergence between $P_0(\mathbf{x})$ from (11) and $P_1(\mathbf{x})$. The final $m_0 - 2$ terms of both are identical. Thus, for the remainder of this subsection we focus on the distribution of the activation pattern in the first two cliques, i.e., $\mathbb{P}(\mathbf{X}_{\mathcal{C}_1} = \mathbf{x}_{\mathcal{C}_1}, \mathbf{X}_{\mathcal{C}_2} = \mathbf{x}_{\mathcal{C}_2}; G_{D1})$.

The differences between $\mathbb{P}(\mathbf{X}_{\mathcal{C}_1} = \mathbf{x}_{\mathcal{C}_1}, \mathbf{X}_{\mathcal{C}_2} = \mathbf{x}_{\mathcal{C}_2}; G_{D1})$ and $\mathbb{P}(\mathbf{X}_{\mathcal{C}_1} = \mathbf{x}_{\mathcal{C}_1}, \mathbf{X}_{\mathcal{C}_2} = \mathbf{x}_{\mathcal{C}_2}; G_{D0})$ are due to the added edge, which constraints the allowable patterns. In particular, the event $\mathbf{X}_{\mathcal{C}_1} = \mathbf{e}_i, \mathbf{X}_{\mathcal{C}_2} = \mathbf{e}_j$, which occurs with nonzero probability under G_{D0}, cannot occur under G_{D1}. If, under

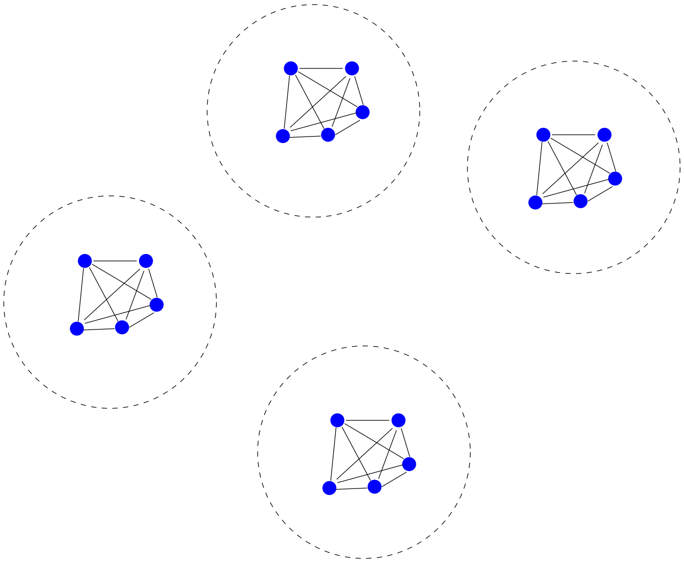

Fig. 8. A network consists of $n/(d-1)$ cliques, where nodes in each clique are within each other's carrier sensing range. Nodes belonging to different cliques cannot hear each other. $n = 20, d = 6$.

G_{D0} both nodes i and j would have transmitted, under G_{D1} only one will transmit. The carrier sensing mechanism will suppress the transmission of the other. Which node will transmit and which will be suppressed will depend on the respective back-off times. The result will be an increase in the probability of some other node transmitting, since the suppressed node will not compete for the channel.

To get the analysis rolling, we first consider the simpler situations of joint transmission patterns $(\mathbf{x}_{\mathcal{C}_1}, \mathbf{x}_{\mathcal{C}_2})$ where there is no possibility that transmission by node i could have suppressed transmission by node j (or vice-versa). For these situations the probability of the joint pattern under G_{D1} is the same as under G_{D0}. There are three such cases. The first case consists patterns such that (i) some node in $\mathcal{C}_i$ *other than* i transmits and (ii) no node in $\mathcal{C}_2$ transmits, i.e., $\mathbf{x}_{\mathcal{C}_1} \in \mathcal{X}_1 \backslash \mathbf{e}_i$ and $\mathbf{x}_{\mathcal{C}_2} = \mathbf{0}$. Condition (i) means that node i could not have suppressed the transmission of node j. Thus, but condition (ii) no node in $\mathcal{C}_2$ has any data to transmit. Therefore,

$$\mathbb{P}(\mathbf{X}_{\mathcal{C}_1} = \mathbf{x}_{\mathcal{C}_1}, \mathbf{X}_{\mathcal{C}_2} = \mathbf{0}; G_{D1}) = \mathbb{P}(\mathbf{X}_{\mathcal{C}_1} = \mathbf{x}_{\mathcal{C}_1}, \mathbf{X}_{\mathcal{C}_2} = \mathbf{0}; G_{D0}).$$

The second case is the reverse of the first, i.e., (i) $\mathbf{x}_{\mathcal{C}_2} \in \mathcal{X}_2 \backslash \mathbf{e}_j$ and (ii) $\mathbf{x}_{\mathcal{C}_1} = \mathbf{0}$. By the same logic, $\mathbb{P}(\mathbf{X}_{\mathcal{C}_1} = \mathbf{0}, \mathbf{X}_{\mathcal{C}_2} = \mathbf{x}_{\mathcal{C}_2}; G_{D1}) = \mathbb{P}(\mathbf{X}_{\mathcal{C}_1} = \mathbf{0}, \mathbf{X}_{\mathcal{C}_2} = \mathbf{x}_{\mathcal{C}_2}; G_{D0})$. The third case consists of situations where (i) transmissions occur in both cliques, but (ii) neither i nor j transmit. Condition (ii) implies that suppression of node j by node i (or i by j) could not have occurred, and thus the probability of the joint pattern under G_{D1} is the same as under G_{D0}. These conditions are summarized as $\mathbf{x}_{\mathcal{C}_1} \in \mathcal{X}_1 \backslash \{\mathbf{0}, \mathbf{e}_i\}$ and $\mathbf{x}_{\mathcal{C}_2} \in \mathcal{X}_2 \backslash \{\mathbf{0}, \mathbf{e}_j\}$. Thus,

$$\mathbb{P}(\mathbf{X}_{\mathcal{C}_1} = \mathbf{x}_{\mathcal{C}_1}, \mathbf{X}_{\mathcal{C}_2} = \mathbf{x}_{\mathcal{C}_2}; G_{D1}) = \mathbb{P}(\mathbf{X}_{\mathcal{C}_1} = \mathbf{x}_{\mathcal{C}_1}, \mathbf{X}_{\mathcal{C}_2} = \mathbf{x}_{\mathcal{C}_2}; G_{D0}).$$

Finally, by (10) and (11) we know that $\mathbb{P}(\mathbf{X}_{\mathcal{C}_1} = \mathbf{x}_{\mathcal{C}_1}, \mathbf{X}_{\mathcal{C}_2} = \mathbf{x}_{\mathcal{C}_2}; G_{D0}) = q^2$.

We now turn to transmission patterns where nodes i and j may have interacted. These are patterns $(\mathbf{x}_{\mathcal{C}_1}, \mathbf{x}_{\mathcal{C}_2})$ the probability of which may be higher under G_{D1} than under G_{D0} due the the infeasibility of the pattern $(\mathbf{e}_i, \mathbf{e}_j)$ under G_{D1}. To analyze these cases consider the paired queue/backoff-time vectors $(\mathbf{Q}, \mathbf{T})$ that, under G_{D0}, would have resulted in $(\mathbf{X}_{\mathcal{C}_1}, \mathbf{X}_{\mathcal{C}_2}) = (\mathbf{e}_i, \mathbf{e}_j)$. Depending on the particular realization of $(\mathbf{Q}, \mathbf{T})$ there are four possible outcomes under G_{D1}:

(a) $T_i < T_j$ and $\mathbf{Q}_{\mathcal{C}_2} = \mathbf{e}_j$: Node i will get the channel and transmit so $\mathbf{X}_{\mathcal{C}_1} = \mathbf{e}_i$. However, the transmission of node j will be suppressed. Since $\mathbf{Q}_{\mathcal{C}_2} = \mathbf{e}_j$, node j is the only node in $\mathcal{C}_2$ with data, so $\mathbf{X}_{\mathcal{C}_2} = \mathbf{0}$.

(b) $T_i > T_j$ and $\mathbf{Q}_{\mathcal{C}_2} = \mathbf{e}_j$: The analysis here is analogous to (a) with the roles of i and j reversed. Thus, the transmission pattern will be $\mathbf{X}_{\mathcal{C}_1} = \mathbf{0}, \mathbf{X}_{\mathcal{C}_2} = \mathbf{e}_j$.

(c) $T_i < T_j$ and $\mathbf{Q}_{\mathcal{C}_2} \neq \mathbf{e}_j$: In this situation at least one other node in $\mathcal{C}_2$ has data to transmit. Thus, even though $\mathbf{X}_{\mathcal{C}_1} = \mathbf{e}_i$ suppresses the transmission of node j, some other node in $\mathcal{C}_2$ will transmit. Thus, $\mathbf{X}_{\mathcal{C}_1} \in \mathcal{X}_1 \backslash \{\mathbf{e}_i, \mathbf{0}\}$. Furthermore, since the statistics of the queue status and backoff times are identical for all nodes, the transmitting node in $\mathcal{C}_2$ will be uniformly distributed across the other $d-2$ nodes in that clique.

(d) $T_i > T_j$ and $\mathbf{Q}_{\mathcal{C}_1} \neq \mathbf{e}_i$: The analysis here is analogous to (c) with the roles of i and j reversed. $\mathbf{X}_{\mathcal{C}_2} = \mathbf{e}_j$, and $\mathbf{X}_{\mathcal{C}_1}$ is uniformly distributed across $\mathbf{x}_{\mathcal{C}_1} \in \mathcal{X}_1 \backslash \{\mathbf{e}_i, \mathbf{0}\}$.

We note that the above four cases partition the event space where nodes i and j transmit concurrently under G_{D0}. The four terms in the following correspond, respectively, to (a)–(d), above:

$$\begin{aligned}
&\mathbb{P}(\mathbf{X}_{\mathcal{C}_1} = \mathbf{e}_i, \mathbf{X}_{\mathcal{C}_2} = \mathbf{e}_j; G_{D0}) \\
&= \mathbb{P}(\mathbf{X}_{\mathcal{C}_1} = \mathbf{e}_i, \mathbf{X}_{\mathcal{C}_2} = \mathbf{e}_j, T_i < T_j, \mathbf{Q}_{\mathcal{C}_2} = \mathbf{e}_j; G_{D0}) \\
&\quad + \mathbb{P}(\mathbf{X}_{\mathcal{C}_1} = \mathbf{e}_i, \mathbf{X}_{\mathcal{C}_2} = \mathbf{e}_j, T_i > T_j, \mathbf{Q}_{\mathcal{C}_1} = \mathbf{e}_i; G_{D0}) \\
&\quad + \mathbb{P}(\mathbf{X}_{\mathcal{C}_1} = \mathbf{e}_i, \mathbf{X}_{\mathcal{C}_2} = \mathbf{e}_j, T_i < T_j, \mathbf{Q}_{\mathcal{C}_2} \neq \mathbf{e}_j; G_{D0}) \\
&\quad + \mathbb{P}(\mathbf{X}_{\mathcal{C}_1} = \mathbf{e}_i, \mathbf{X}_{\mathcal{C}_2} = \mathbf{e}_j, T_i > T_j, \mathbf{Q}_{\mathcal{C}_1} \neq \mathbf{e}_i; G_{D0}) \\
&= 2\beta_1 + 2\beta_2. \qquad (16)
\end{aligned}$$

The first two terms are equal due to the symmetry of the conditions and graph structure. Similar logic implies that the third and fourth terms are also equal. We respectively define β_1 and β_2 to be the two probabilities.

We now consider the four cases of joint transmission patterns $(\mathbf{x}_{\mathcal{C}_1}, \mathbf{x}_{\mathcal{C}_2})$ under G_{D1} not yet considered. These will each connect to one of the cases (a)–(d), above. First considered the probability of pattern $(\mathbf{x}_{\mathcal{C}_1}, \mathbf{x}_{\mathcal{C}_2}) = (\mathbf{e}_i, \mathbf{0})$ under G_{D1}. The probability of this pattern under G_{D1} will be larger than under under G_{D0} since certain pairs $(\mathbf{Q}, \mathbf{T})$ that result in $(\mathbf{e}_i, \mathbf{e}_j)$ under G_{D0} result in to $(\mathbf{e}_i, \mathbf{0})$ under G_{D1}. The probability of observing $(\mathbf{e}_i, \mathbf{0})$ under G_{D1} equals the probability of observing that pattern under G_{D0} plus the probability of the event occurring that was considered in case (a), above. This latter event is the bump in probability due to the interaction of i and j. Therefore, we find that

$$\begin{aligned}&\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,\mathbf{X}_{\mathcal{C}_2}=\mathbf{0};G_{D1})\\&=\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,\mathbf{X}_{\mathcal{C}_2}=\mathbf{0};G_{D0})\\&\quad+\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j,T_i<T_j,\mathbf{Q}_{\mathcal{C}_2}=\mathbf{e}_j;G_{D0})\\&=q(1-p)^{d-1}+\beta_1,\end{aligned}\tag{17}$$

where the first term follows from the independence of the node-wise transmission patterns under G_{D0}, (10), and the fact that no nodes in $\mathcal{C}_2$ have data to transmit. We defer the calculation of $\beta_1=\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j,T_i<T_j,\mathbf{Q}_{\mathcal{C}_2}=\mathbf{e}_j;G_{D0})$ till **??**. Following a similar line of argument for $(\mathbf{x}_{\mathcal{C}_1},\mathbf{x}_{\mathcal{C}_2})=(\mathbf{0},\mathbf{e}_j)$, and considering case (b), we find that

$$\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{0},\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j;G_{D1})=q(1-p)^{d-1}+\beta_1\tag{18}$$

The third case concerns the patterns $(\mathbf{e}_i,\mathbf{x}_{\mathcal{C}_2})$ for all $\mathbf{x}_{\mathcal{C}_2}\in\mathcal{X}_2\backslash\{\mathbf{0},\mathbf{e}_j\}$. Similar to (a) and (b), the probability of the pattern will equal the probability of the pattern under G_{D0} plus a boost due to the infeasibility of the $(\mathbf{e}_i,\mathbf{e}_j)$ pattern under G_{D1}. The boost corresponds to the event discussed in (c), above. For any $\mathbf{x}_{\mathcal{C}_2}\in\mathcal{X}_2\backslash\{\mathbf{e}_j,\mathbf{0}\}$ we find that

$$\begin{aligned}&\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,\mathbf{X}_{\mathcal{C}_2}=\mathbf{x}_{\mathcal{C}_2};G_{D1})\\&=\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,\mathbf{X}_{\mathcal{C}_2}=\mathbf{x}_{\mathcal{C}_2};G_{D0})\\&\quad+\frac{1}{d-2}\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j,T_i<T_j,\mathbf{Q}_{\mathcal{C}_2}\neq\mathbf{e}_j;G_{D0})\\&=q^2+\frac{1}{d-2}\beta_2.\end{aligned}\tag{19}$$

The factor of $1/(d-2)$ in the second term results from the uniformity over the other $d-2$ transmission patterns in $\mathcal{C}_2$, mentioned in (c). The probability $\beta_2=\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j,T_i<T_j,\mathbf{Q}_{\mathcal{C}_2}\neq\mathbf{e}_j;G_{D0}$ will be calculated in (22). Finally, by symmetric logic, we find that for any $\mathbf{x}_{\mathcal{C}_1}\in\mathcal{X}_1\backslash\{\mathbf{e}_i,\mathbf{0}\}$

$$\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{x}_{\mathcal{C}_1},\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j;G_{D1})=q^2+\frac{1}{d-2}\beta_2.\tag{20}$$

We now calculate β_1, required in (17) and (18). We start by rewriting the first term (16) using Bayes' rule as

$$\begin{aligned}&\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j,T_i<T_j,\mathbf{Q}_{\mathcal{C}_2}=\mathbf{e}_j;G_{D0})\\&\quad=\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,T_i<T_j;G_{D0})\mathbb{P}(\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j,\mathbf{Q}_{\mathcal{C}_2}=\mathbf{e}_j;G_{D0}).\end{aligned}$$

In the application of Bayes' rule we have used the fact that $\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,T_i<T_j|\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j,\mathbf{Q}_{\mathcal{C}_2}=\mathbf{e}_;G_{D0})=\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,T_i<T_j;G_{D0})$, due to the independence of transmission patterns under G_{D0}. Now, note that $\mathbb{P}(\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j,\mathbf{Q}_{\mathcal{C}_2}=\mathbf{e}_j;G_{D0})=p(1-p)^{d-2}$ as node j is the only node in $\mathcal{C}_2$ with something to transmit. Next, rewrite $\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,T_i<T_j;G_{D0})$ as

$$\begin{aligned}&\sum_{l=0}^{d-2}\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,T_i<T_j,|\{k\in\mathcal{C}_1:T_k<T_i\}|=l;G_{D0})\\&=\sum_{l=0}^{d-2}\mathbb{P}\Big(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i\Big|T_i<T_j,|\{k\in\mathcal{C}_1:T_k<T_i\}|=l;G_{D0}\Big)\\&\quad\cdot\mathbb{P}(T_i<T_j,|\{k\in\mathcal{C}_1:T_k<T_i\}|=l;G_{D0}).\end{aligned}$$

The first factor $\mathbb{P}\Big(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i\Big|T_i<T_j,|\{k\in\mathcal{C}_1:T_k<T_i\}|=l;G_{D0}\Big)=p(1-p)^l$. The second factor is just the fraction of the $d!$ orderings such that there are l nodes in $\mathcal{C}_1$ with $T_k<T_i$ and such that node $T_i<T_j$. The number of such orderings is $\binom{d-2}{l}l!\binom{d-1-1}{1}(d-l-2)!=(d-2)!(d-l-1)$. Putting the pieces together we find that

$$\begin{aligned}&\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j,T_i<T_j,\mathbf{Q}_{\mathcal{C}_2}=\mathbf{e}_j;G_{D0})\\&\quad=\left[\sum_{l=0}^{d-2}(d-l-1)(1-p)^l\right]\frac{p^2(1-p)^{(d-2)}}{d(d-1)}\triangleq\beta_1.\end{aligned}\tag{21}$$

And, since transmitters i and j have the same statistics, when $T_i>T_j$, we also have

$$\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{0},\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j,T_i>T_j,\mathbf{Q}_{\mathcal{C}_1}=\mathbf{e}_i;G_{D0})=\beta_1,$$

which justifies (18).

Finally, to calculate β_2 we simply combine (16) with (21).

$$\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j;G_{D0})=2\beta_1+2(d-2)\beta_2=q^2,$$

where the final inequality follows from the independence of $\mathbf{X}_{\mathcal{C}_1}$ and $\mathbf{X}_{\mathcal{C}_2}$ under G_{D0}. Thus,

$$\beta_2=\frac{q^2-2\beta_1}{2}..\tag{22}$$

4) Bounding $D_{KL}(P_0\|P_1)$: The Kullback-Leibler divergence between the transmission pattern distributions under G_{D1} and G_{D0}, denoted as P_1,P_0, respectively, can be calculated as

$$\begin{aligned}&D_{KL}(P_1\|P_0)\\&=\sum_{\mathbf{x}\in\{0,1\}^n}P_1(\mathbf{x})\log\frac{P_1(\mathbf{x})}{P_0(\mathbf{x})}\\&=\sum_{\mathbf{x}_{\mathcal{C}_1}\in\{0,1\}^{d-1},\mathbf{x}_{\mathcal{C}_2}\in\{0,1\}^{d-1}}\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{x}_{\mathcal{C}_1},\mathbf{X}_{\mathcal{C}_2}=\mathbf{x}_{\mathcal{C}_2};G_{D1})\\&\qquad\cdot\log\frac{\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{x}_{\mathcal{C}_1},\mathbf{X}_{\mathcal{C}_2}=\mathbf{x}_{\mathcal{C}_2};G_{D1})}{\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{x}_{\mathcal{C}_1};G_{D0})\mathbb{P}(\mathbf{X}_{\mathcal{C}_2}=\mathbf{x}_{\mathcal{C}_2};G_{D0})}\end{aligned}\tag{23}$$

$$\begin{aligned}&=2\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,\mathbf{X}_{\mathcal{C}_2}=\mathbf{0};G_{D1})\\&\qquad\cdot\log\frac{\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,\mathbf{X}_{\mathcal{C}_2}=\mathbf{0};G_{D1})}{\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i;G_{D0})\mathbb{P}(\mathbf{X}_{\mathcal{C}_2}=\mathbf{0};G_{D0})}\\&\quad+2(d-2)\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_{j'};G_{D1})\\&\qquad\cdot\log\frac{\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_{j'};G_{D1})}{\mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i;G_{D0})\mathbb{P}(\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_{j'};G_{D0})}\end{aligned}\tag{24}$$

$$\begin{aligned}&=2[q(1-p)^{d-1}+\beta_1]\log\left[\frac{q(1-p)^{d-1}+\beta_1}{q(1-p)^{d-1}}\right]\\&\qquad+2(d-2)\left[q^2+\frac{\beta_2}{d-2}\right]\log\left[\frac{q^2+\frac{\beta_2}{d-2}}{q^2}\right]\end{aligned}\tag{25}$$

$$\leq\left(2+\frac{1}{1-p}\right)\frac{1}{(d-1)^2}.\tag{26}$$

In (23) we cancel (and marginalize over) the m_0-1 common factors of the form $\mathbb{P}(\mathbf{X}_{\mathcal{C}_m}=\mathbf{x}_{\mathcal{C}_m};G_{D0})$, $3\leq m\leq m_0+1$ cf. (11) and (15). In (24), we focus in on the terms that don't cancel out. The first two terms therein correspond to cases (a) and (b), cf. (17) and (18). The latter $2(d-2)$ terms correspond the cases (c) and (d), cf. (19) and (20) where j' is some node $j'\in\mathcal{C}_2$ but $j'\neq j$. In (25) we use (17) and (19) in the numerators and (9) and (11) in the denominators. Finally, substituting in for the definitions of β_1 and β_2 from (21) and (22), equation (26) follows from the fact that $\log(1+x) \leq x$.

5) Put pieces together: Since the G_{Di}s are constructed in the same manner, $D_{KL}(P_i \| P_0) = D_{KL}(P_1 \| P_0)$ for all i. Thus, we have

$$\frac{1}{M}\sum_{i=1}^{M} D_{KL}(P_i \| P_0) \leq \left(2+\frac{1}{1-p}\right)\frac{1}{(d-1)^2}.$$

In summary, we have $D_L(G_{Di}, G_{Dj}) \geq 1$ for $0 \leq i < j \leq M$ and we can always pick $M = n$. Thus, according to Thm. 7, when

$$k \leq \frac{\alpha \log n}{\left(2+\frac{1}{1-p}\right)\frac{1}{(d-1)^2}}$$

we have

$$\begin{aligned}
&\inf_{\hat{G}_D \in \mathcal{G}_d} \sup_{G_D \in \mathcal{G}_d} \mathbb{P}(\hat{G}_D \neq G_D; G_D) \\
&\quad = \inf_{\hat{G}_D \in \mathcal{G}_d} \sup_{G_D \in \mathcal{G}_d} \mathbb{P}(D_L(\hat{G}_D, G_D) \geq 1/2; G_D) \\
&\quad > \frac{\sqrt{n}}{1+\sqrt{n}}\left(1 - 2\alpha - \sqrt{\frac{2\alpha}{\log n}}\right) > 0.
\end{aligned}$$

C. Proof of Lemma 1

Since $Y_j = 0$ must be caused by some active interferer, $\mathcal{S}_j^t \cap \mathcal{S}_j \neq \varnothing$ for every $t \in \mathcal{K}_j(k)$. Therefore, $\mathcal{S}_j$ is a hitting set for $\{\mathcal{S}_j^t\}_{t\in\mathcal{K}_j(k)}$.

Next, we prove that $\mathcal{S}_j$ is the unique minimum hitting set. We prove this through contradiction. Assume that there exist a different hitting set $\hat{\mathcal{S}}_j(k)$ with $|\hat{\mathcal{S}}_j(k)| \leq |\mathcal{S}_j|$. Since $\hat{\mathcal{S}}_j(k)$ is different from $\mathcal{S}_j$, there must exist a node $i \in \mathcal{S}_j$, that is not in $\hat{\mathcal{S}}_j(k)$, i.e., $i \in \mathcal{S}_j \backslash \hat{\mathcal{S}}_j(k)$. Consider the following probability

$$\begin{aligned}
&\mathbb{P}(X_i = 1, X_j = 1, Y_j = 0, \mathbf{X}_{\hat{\mathcal{S}}_j(k)} = \mathbf{0}) \\
&= \mathbb{P}(X_i = 1, X_j = 1, \mathbf{X}_{\hat{\mathcal{S}}_j(k)} = \mathbf{0}) \\
&\quad \cdot \mathbb{P}(Y_j = 0 | X_i = 1, X_j = 1, \mathbf{X}_{\hat{\mathcal{S}}_j(k)} = \mathbf{0}) \\
&\geq \mathbb{P}(X_i = 1, X_j = 1, \mathbf{Q}_{\hat{\mathcal{S}}_j(k)} = \mathbf{0}) \\
&\quad \cdot \mathbb{P}(Y_j = 0 | X_i = 1, X_j = 1, \mathbf{X}_{\hat{\mathcal{S}}_j(k)} = \mathbf{0}) \\
&\geq \frac{p^2}{d^2}(1-p)^{|\hat{\mathcal{S}}_j(k)|} p_{ij} \geq \frac{p^2}{d^2}(1-p)^s p_{\min},
\end{aligned} \tag{27}$$

where (27) follows from the assumptions that $i \in \mathcal{S}_j$, $|\hat{\mathcal{S}}_j(k)| \leq |\mathcal{S}_j| \leq s$, and (2).

For any observation with $X_i(t) = 1, X_j(t) = 1, Y_j(t) = 0, \mathbf{X}_{\hat{\mathcal{S}}_j(k)}(t) = \mathbf{0}$, its time index t belongs to $\mathcal{K}_j(k)$, however, $\mathcal{S}_j^t \cap \hat{\mathcal{S}}_j(k) = \varnothing$, which contradicts with the assumption that $\hat{\mathcal{S}}_j(k)$ is a hitting set. Since this event has a lower bounded probability for any fixed d, s, as $k \to \infty$, this event happens with probability one. Therefore, $\mathcal{S}_j$ is the unique minimum hitting set as $k \to \infty$. This proves the lemma.

D. Proof of Theorem 3

Define the error event $\mathcal{C}_j$ as the event that the estimated minimum hitting set $\hat{\mathcal{S}}_j$ is not equal to $\mathcal{S}_j$ after k observations. This only happens when $|\hat{\mathcal{S}}_j| \leq |\mathcal{S}_j|$, because otherwise $\mathcal{S}_j$ is a feasible hitting set with smaller size. So when $|\hat{\mathcal{S}}_j| \leq |\mathcal{S}_j|$, there must exist at least transmitter $i \in \mathcal{S}_j$ that is not included in $\hat{\mathcal{S}}_j$. Then, following the steps in the proof of Lemma 1, we have

$$\begin{aligned}
\mathbb{P}(\mathcal{C}_j) &= \mathbb{P}(\cup_{i\in\mathcal{S}_j} i \notin \hat{\mathcal{S}}_j) \\
&\leq \sum_{i\in\mathcal{S}_j}(1 - \mathbb{P}(X_i = 1, X_j = 1, Y_j = 0, \mathbf{X}_{\hat{\mathcal{S}}_j} = \mathbf{0}))^k \\
&\leq s\left(1 - \frac{p^2}{d^2}(1-p)^s p_{\min}\right)^k.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{P}(\hat{G}_H \neq G_H) &= \mathbb{P}(\cup_j \mathcal{C}_j) \\
&\leq \sum_j s\left(1 - \frac{p^2}{d^2}(1-p)^{s-1} p_{\min}\right)^k \\
&= ns\left(1 - \frac{p^2(1-p)^s p_{\min}}{d^2}\right)^k.
\end{aligned}$$

E. Proof of Theorem 4

To prove Theorem 4 we follow a similar approach to that taken in the proof of Theorem 2. For any given direct interference graph $G_D = (\mathcal{V}, \mathcal{E}_D)$, we define $\mathcal{G}_s$ to be the set of hidden interference graphs satisfying the assumption that $s_j \leq s$ for every j. We construct a collections of graphs, $G_{H0}, G_{H1}, \ldots, G_{HM}$, all in $\mathcal{G}_s$, and reduce the interference estimation problem to an M-ary hypothesis test. These graphs share the same node set and direct interference edges, however, the hidden interference edges differ. With a little abuse of the notation, we use $P_i(\mathbf{x}, \mathbf{y})$ to denote the joint distribution of transmission pattern $\mathbf{x}$ and feedback information vector $\mathbf{y}$ under G_D and G_{Hi}, $0 \leq i \leq M$. We use $\mathbb{P}(\mathcal{A}; G_{Hi})$ to denote the probability of event $\mathcal{A}$ under distribution $P_i(\mathbf{x}, \mathbf{y})$. Note that $\mathbb{P}(\mathcal{A}; G_{Hi})$ implicitly depends on the underlying direct interference graph G_D.

1) Construct G_D and G_{H0}: We now construct an underlying direct interference graph G_D, and add directed edges to form G_{H0}. An illustrative example of a possible G_D is provided in Fig. 9. We partition the node set into $\lceil n/(2d+s-1)\rceil$ groups. The first $\lfloor n/(2d+s-1)\rfloor$ groups consist of $2d+s-1$ nodes. The last group consists of the remaining nodes. In each group, except the last, we cluster $2d$ nodes into a pair of cliques, each clique consisting of d nodes. The remaining $s-1$ nodes are "independent" nodes or "atoms", disconnected from all other nodes in the network. Thus, their activation status depends only on their own queue status; it is independent of everything else.

We construct G_{H0} by adding directed edges to G_D. These edges will be added only between nodes in the same group. Hidden interference will thus exist only among nodes within the same group. It will not exist between groups. To get the hidden interference G_{H0} consider each node in each clique in each group. Let all $s-1$ independent nodes in that same group be hidden interferers as well as one (any one) node in the other clique in the same group. Thus every node in each clique has exactly s hidden interferers. We note that a node in a clique

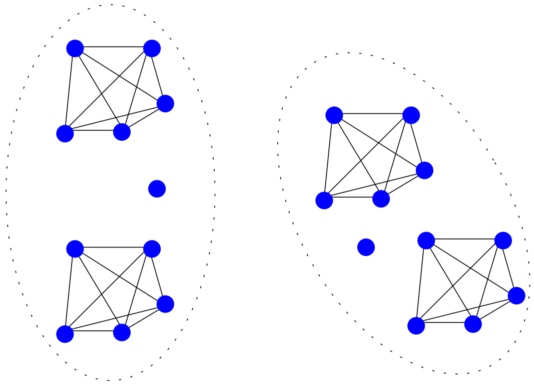

Fig. 9. A direct interference graph G_D of $n/(2d+s-1)=2$ groups, where each group consists of two fully connected cliques of size d and $s-1$ detached APs; $n=22, d=5, s=2$.

is allowed to interfere with more than one node in the other clique. The last group can have an arbitrary edge structure as long as it satisfies the maximum degree constraints.

Part of the hidden graph structure are the probabilities p_{ij}, defined in (1), of whether hidden interferer $i \in \mathcal{S}_j$ interferes with the transmission of node j. We now specify these probabilities for G_{H0}. For all $i \in \mathcal{S}_j$ and $\mathcal{S} \subseteq \mathcal{V}\backslash\{i,j\}$, the hidden interference satisfies

$$\begin{aligned}&\mathbb{P}(Y_j=0|X_i=1,X_j=1,\mathbf{X}_{\mathcal{V}\backslash\{i,j\}}=\mathbf{0};G_{H0})\\&=\mathbb{P}(Y_j=0|X_i=1,X_j=1,\mathbf{X}_{\mathcal{S}}=\mathbf{x}_{\mathcal{S}};G_{H0})=p_{\min},\end{aligned}\tag{28}$$

where $\mathbf{x}_{\mathcal{S}}$ is any transmission pattern feasible under the direct interference graph G_D. Thus, in contrast to the inequality (2) in the general setting, for this network the bound holds with *equality* for all $i \in \mathcal{S}_j$. The implication is that the transmission collision probability for an AP j doesn't increase if there are more than one hidden interferer transmitting. This assumption holds for every hidden interference graph G_{Hi} discussed in this section.

2) Construct M hidden interference graphs: We now construct a set of hidden interference graphs $G_{H1}, G_{H2}, \ldots, G_{HM}$ as perturbations of G_{H0}. We get each by removing a single directed edge in G_{H0}. The edge removed must have connected a pair of nodes in distinct cliques in a group in G_{H0}. That is, we do not remove an edge between an independent node and a node in a clique.

In each group, there are $2d$ such edges. There are thus $2d\lfloor n/(2d+s-1)\rfloor$ distinct edges in G_{H0} that can be removed. If

$$d \le c_1 n, \quad s-1 \le c_2 n, \quad \text{and} \quad 2c_1+c_2<1, \tag{29}$$

where c_1, c_2 are positive constants, there are more than $2c_1(\frac{1}{2c_1+c_2}-1)n := M$ such edges. For each of these graphs, $D_L(D_{H0}, D_{Hi}) = 1$ and $D_L(D_{Hi}, D_{Hj}) = 2$ where $1 \le i, j \le M$.

3) Characterize $D_{KL}(P_1\|P_0)$: Without loss of generality, we assume that the directed edge (i,j) is removed from G_{H0} to form G_{H1}, where $i \in \mathcal{C}_1$, and $j \in \mathcal{C}_2$. Since we are *removing* an edge from G_{H0} to get G_{H1}, we assume property (28) is inherited. Thus, the only difference between the distributions under G_{H1} and G_{H0} occurs when $\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i, \mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j$, and $\mathbf{X}_{\mathcal{S}_j\backslash i}=\mathbf{0}$. Specifically,

$$\mathbb{P}(Y_j=0|\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j,\mathbf{X}_{\mathcal{S}_j\backslash i}=\mathbf{0};G_{H0})=p_{\min},$$

while

$$\mathbb{P}(Y_j=0|\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i,\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j,\mathbf{X}_{\mathcal{S}_j\backslash i}=\mathbf{0};G_{H1})=0.$$

Therefore,

$$\begin{aligned}D_{KL}(P_1\|P_0)&=\sum_{\mathbf{x},\mathbf{y}}P_1(\mathbf{x},\mathbf{y})\log\frac{P_1(\mathbf{x},\mathbf{y})}{P_0(\mathbf{x},\mathbf{y})}\\&=\sum_{\mathbf{x},\mathbf{y}}P_1(\mathbf{x})P_1(\mathbf{y}|\mathbf{x})\log\frac{P_1(\mathbf{y}|\mathbf{x})P_1(\mathbf{x})}{P_0(\mathbf{y}|\mathbf{x})P_0(\mathbf{x})}\\&=\mathbb{P}(\mathcal{A};G_{H1})\sum_{y_j\in\{0,1\}}P_1(y_j|\mathcal{A})\log\frac{P_1(y_j|\mathcal{A})}{P_0(y_j|\mathcal{A})}\\&=-\left(\frac{1-(1-p)^d}{d}\right)^2(1-p)^{s-1}\log(1-p_{\min})\end{aligned}$$

where $P_1(\mathbf{x}) = P_0(\mathbf{x})$ since G_D is held fixed, $\mathcal{A} := \{\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i, \mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j, \mathbf{X}_{\mathcal{S}_j\backslash i}=\mathbf{0}\}$. $\mathbb{P}(\mathcal{A};G_{H1}) = \mathbb{P}(\mathbf{X}_{\mathcal{C}_1}=\mathbf{e}_i;G_{H_1})\mathbb{P}(\mathbf{X}_{\mathcal{C}_2}=\mathbf{e}_j;G_{H1})\mathbb{P}(\mathbf{X}_{\mathcal{S}_j\backslash i}=\mathbf{0};G_{H1})$ because $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{S}_j\backslash i$ are disconnected in G_D. $\mathbb{P}(\mathcal{A};G_{H1})$ is then calculated following similar steps to obtain (11).

4) Put the pieces together: In summary, we have $M = 2c_1(\frac{1}{2c_1+c_2}-1)n$ and $D_L(G_{Hi}, G_{Hj}) \ge 1$ for $0 \le i < j \le M$. Theorem 4 is thus proved by application of Theorem 7.

F. Proof of Theorem 5

The probability that, after k observations, the graph cannot be identified successfully is

$$\mathbb{P}(\hat{G}_D \ne G_D) = \mathbb{P}\big(\cup_{(i,j)\in\mathcal{V}\times\mathcal{V}}\mathcal{A}_{ij}\big)$$

where $\mathcal{A}_{ij}$ is the error event associated with any AP pair (i,j). There are two types of errors, namely, type I errors $\mathcal{A}_{ij}^{I}$ ("false positive" or "false alarm" errors) and type II error $\mathcal{A}_{ij}^{II}$ ("false negative" or "missed detection" errors). In the following, we provide on the bounds on these two types of errors.

Under the noisy carrier sensing channel assumption, for edge $(i,j) \in \mathcal{E}_D$,

$$\begin{aligned}I_{ij}^1&:=\mathbb{P}(X_i=1,X_j=1|(i,j)\in\mathcal{E}_D)\\&\le\mathbb{P}\left(Q_i>0,Q_j>0\right)q_{ij}\\&\le p^2q_{\text{low}}\end{aligned}$$

where the upper bound corresponds to the situation that (i,j) is the only edge in $\mathcal{E}_D$ associated with AP i or AP j, and the factor q_{ij} comes from the fact that APs i and j should not sense each other in order to transmit simultaneously. Since $(i,j) \in \mathcal{E}_D$, $q_{ij} < q_{\text{low}}$, and the last inequality follows.

For the remaining AP pairs (i.e., non-edges),

$$\begin{aligned}I_{ij}^0&:=\mathbb{P}(X_i=1,X_j=1|(i,j)\notin\mathcal{E}_D)\\&\ge\mathbb{P}\left(T_i<T_{\mathcal{N}_i},T_j<T_{\mathcal{N}_j},Q_i>0,Q_j>0\right)q_{ij}\\&\ge\frac{p^2q_{\text{up}}}{d^2}.\end{aligned}$$

Then, by Assumption 2, we have

$$I_{ij}^1 < \epsilon < I_{ij}^0. \tag{30}$$

We note that Assumption 2 serves as a sufficient, but not necessary, condition for (30) to hold. This makes it possible to distinguish between edges and non-edges based on the probability of simultaneous transmission.

Applying Hoeffding's inequality, we have

$$\begin{aligned}\mathbb{P}(\mathcal{A}_{ij}^I) &= \mathbb{P}\left(\hat{I}_{ij} < \epsilon \middle| (i,j) \notin \mathcal{E}_D\right)\\ &= \mathbb{P}\left(\hat{I}_{ij} - I_{ij}^1 \geq I_{ij}^0 - \epsilon \middle| (i,j) \in \mathcal{E}_D\right)\\ &\leq \exp\left\{-2\Big(\frac{p^2 q_{\text{up}}}{d^2} - \epsilon\Big)^2 k\right\}.\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}(\mathcal{A}_{ij}^{II}) &= \mathbb{P}\left(\hat{I}_{ij} \geq \epsilon \middle| (i,j) \in \mathcal{E}_D\right)\\ &= \mathbb{P}\left(\hat{I}_{ij} - I_{ij}^1 \geq \epsilon - I_{ij}^1 \middle| (i,j) \in \mathcal{E}_D\right)\\ &\leq \exp\left\{-2(\epsilon - p^2 q_{\text{low}})^2 k\right\}\end{aligned}$$

Thus,

$$\mathbb{P}(\mathcal{A}_{ij}) \leq \max\left\{\mathbb{P}(\mathcal{A}_{ij}^I), \mathbb{P}(\mathcal{A}_{ij}^{II})\right\} = \exp\{-2\epsilon^2 k\}$$

where we recall from the theorem statement that $\epsilon = \frac{p^2 q_{\text{up}}(1-1/c_3)}{2d^2}$. Thus, the probability that, after k observations, the graph cannot be successfully identified is

$$\begin{aligned}\mathbb{P}(\hat{G}_D \neq G_D) &= \mathbb{P}\big(\cup_{(i,j)\in\mathcal{V}\times\mathcal{V}} \mathcal{A}_{ij}\big) &(31)\\ &\leq n^2 \exp\{-2\epsilon^2 k\}. &(32)\end{aligned}$$

G. Proof of Lemma 2

First, based on the definition of hidden interference in (1) and the assumption made in (2), for any $i \in \mathcal{S}_j$, we have

$$\mathbb{P}(Y_j = 0 | X_i = 1, X_j = 1, X_{\mathcal{S}_j \backslash i} = \mathbf{0}, X_{\mathcal{N}_j} = \mathbf{0}) \geq p_{ij}.$$

On the other hand, for any $i \in \mathcal{V}\backslash\{\mathcal{S}_j \cup \mathcal{N}_j\}$, since $\emptyset$ is the only source of interference for AP j when $X_{\mathcal{S}_j} = \mathbf{0}$ and $X_{\mathcal{N}_j} = \mathbf{0}$, and we assume $p_{\emptyset j}$ is independent of the status of APs in the network, we have

$$\begin{aligned}&\mathbb{P}(Y_j = 0 | X_i = 1, X_j = 1, X_{\mathcal{S}_j} = \mathbf{0}, X_{\mathcal{N}_j} = \mathbf{0})\\ &= \mathbb{P}(Y_j = 0 | X_j = 1, X_{\mathcal{V}\backslash\{j\}} = \mathbf{0}) = p_{\emptyset j}.\end{aligned}$$

Thus, the true hidden interferer set $\mathcal{S}_j$ satisfies the two properties described in (4) and (5). We need to prove that this set is unique, i.e., there does not exist another set $\mathcal{S}'_j \neq \mathcal{S}_j$ satisfying those two properties.

We first show that $\mathcal{S}_j \subseteq \mathcal{S}'_j$. If this were not the case, there must exist a node $i \in \mathcal{S}_j \backslash \mathcal{S}'_j$. Since $X_{\mathcal{S}'_j}$ is a interferer set, according to (5), we have

$$\mathbb{P}(Y_j = 0 | X_i = X_j = 1, X_{\mathcal{S}'_j} = \mathbf{0}, X_{\mathcal{N}_j} = \mathbf{0}) = p_{\emptyset_j} < p_{ij},$$

which contradicts the assumption that $i \in \mathcal{S}_j$, and

$$\mathbb{P}(Y_j = 0 | X_i = X_j = 1, X_{\mathcal{S}'_j} = \mathbf{0}, X_{\mathcal{N}_j} = \mathbf{0}) \geq p_{ij}.$$

We now show that $\mathcal{S}_j = \mathcal{S}'_j$. If this were not the case, there must exist a node $i \in \mathcal{S}'_j \backslash \mathcal{S}_j$. According to (4), we have

$$\mathbb{P}(Y_j = 0 | X_i = X_j = 1, X_{\mathcal{S}'_j \backslash i} = \mathbf{0}, X_{\mathcal{N}_j} = \mathbf{0}) \geq p_{ij}. \tag{33}$$

However, since $i \notin \mathcal{S}_j$ then, according to (5), we have

$$\begin{aligned}&\mathbb{P}(Y_j = 0 | X_i = X_j = 1, X_{\mathcal{S}'_j \backslash i} = \mathbf{0}, X_{\mathcal{N}_j} = \mathbf{0})\\ &= \mathbb{P}(Y_j = 0 | X_i = X_j = 1, X_{\mathcal{S}_j} = \mathbf{0}, X_{\mathcal{N}_j} = \mathbf{0})\\ &= p_{\emptyset j}.\end{aligned}$$

which contradicts (33).

H. Proof of Theorem 6

Before we proceed to prove the theorem, we introduce the following lemma which will be used to prove the theorem.

Lemma 3 *Let (U_i, W_i) for $i = 1, 2, \ldots, k$ be a length-k i.i.d. sequence of a pair of random variables, each having support $[0,1]$ and respective mean $E[U_i] = u_0$ and $E[W_i] = w_0$. For any $\delta > 0$ the sample averages $\bar{U} = (1/k)\sum_{i=1}^k U_i$ and $\bar{W} = (1/k)\sum_{i=1}^k W_i$ satisfy*

$$\begin{aligned}&\mathbb{P}\Big(\frac{\bar{U}}{\bar{W}} - \frac{u_0}{w_0} > \delta\Big)\\ &\leq \exp\Big\{-\frac{k(\delta(w_0 - \delta_w) - \frac{u_0}{w_0}\delta_w)^2}{2}\Big\} + \exp\Big\{-\frac{k\delta_w^2}{2}\Big\}\end{aligned}$$

where δ_w is any parameter satisfying $0 < \delta_w < w_0$ and $\delta(w_0 - \delta_w) - \frac{u_0}{w_0}\delta_w > 0$.

Similarly,

$$\begin{aligned}&\mathbb{P}\Big(\frac{\bar{U}}{\bar{W}} - \frac{u_0}{w_0} < -\delta\Big)\\ &\leq \exp\Big\{-\frac{k(\delta(w_0 + \delta_w) - \frac{u_0}{w_0}\delta_w)^2}{2}\Big\} + \exp\Big\{-\frac{k\delta_w^2}{2}\Big\}\end{aligned}$$

where δ_w is any parameter satisfying $\delta_w > 0$, $\delta(w_0 + \delta_w) - \frac{u_0}{w_0}\delta_w > 0$.

Proof:

$$\begin{aligned}&\mathbb{P}\Big(\frac{\bar{U}}{\bar{W}} - \frac{u_0}{w_0} > \delta\Big)\\ &= \mathbb{P}\Big(\frac{\bar{U}}{\bar{W}} - \frac{u_0}{w_0} > \delta, \bar{W} \geq w_0 - \delta_w\Big)\\ &\quad + \mathbb{P}\Big(\frac{\bar{U}}{\bar{W}} - \frac{u_0}{w_0} > \delta, \bar{W} < w_0 - \delta_w\Big)\\ &\leq \mathbb{P}\Big(\bar{U} - u_0 > \delta(w_0 - \delta_w) - \frac{u_0}{w_0}\delta_w\Big) + \mathbb{P}\Big(\bar{W} - w_0 < -\delta_w\Big)\\ &\leq \exp\Big\{-\frac{k(\delta(w_0 - \delta_w) - \frac{u_0}{w_0}\delta_w)^2}{2}\Big\} + \exp\Big\{-\frac{k\delta_w^2}{2}\Big\}.\end{aligned}$$

The last inequality follows from Hoeffding's inequality. Following similar steps, we have

$$\begin{aligned}
&\mathbb{P}\Big(\frac{\bar{U}}{\bar{W}} - \frac{u_0}{w_0} < -\delta\Big) \\
&= \mathbb{P}\Big(\frac{\bar{U}}{\bar{W}} - \frac{u_0}{w_0} < -\delta, \bar{W} \geq w_0 + \delta_w\Big) \\
&\quad + \mathbb{P}\Big(\frac{\bar{U}}{\bar{W}} < -\delta, \bar{W} < w_0 + \delta_w\Big) \\
&\leq \mathbb{P}\Big(\bar{U} - u_0 < -\delta(w_0 + \delta_w) + \frac{u_0}{w_0}\delta_w\Big) + \mathbb{P}\Big(\bar{W} > w_0 + \delta_w\Big) \\
&\leq \exp\Big\{ - \frac{k(\delta(w_0 + \delta_w) - \frac{u_0}{w_0}\delta_w)^2}{2}\Big\} + \exp\Big\{ - \frac{k\delta_w^2}{2}\Big\}.
\end{aligned}$$

■

For the hidden interference graph learning problem in the random setting there are three sources of error. The first is incorrect estimation of G_D, the direct interference graph. We denote this event $\mathcal{F}_0$. The other two are missed detection of an edge and false detection of an edge (a "false alarm"). Define $\mathcal{F}_{ij}$ to be the error event associated with (i, j) given that the input $\hat{G}_D$ is correct, i.e., $\mathcal{F}_0^c$. We denote the type I error by $\mathcal{F}_{ij}^I$ and the type II error by $\mathcal{F}_{ij}^{II}$.

Define

$$\begin{aligned}
U_{ij,\mathcal{S}} &:= \mathbf{1}\{Y_j = 0, X_{\mathcal{S}} = \mathbf{0}, X_{\mathcal{N}_j} = \mathbf{0}, X_i = X_j = 1\} \\
W_{ij,\mathcal{S}} &:= \mathbf{1}\{X_{\mathcal{S}} = \mathbf{0}, X_{\mathcal{N}_j} = \mathbf{0}, X_i = X_j = 1\}
\end{aligned}$$

where $\mathcal{S} \subseteq \mathcal{V}\backslash\{\mathcal{N}_j \cup \{i, j\}\}, |\mathcal{S}| \leq d$. Then, $U_{ij,\mathcal{S}}$ and $W_{ij,\mathcal{S}}$ are two random variables. Each has independent and identical distribution across k observations. Denote their mean values as $u_{ij,\mathcal{S}}^0$ and $w_{ij,\mathcal{S}}^0$, respectively. Note that

$$\begin{aligned}
w_{ij,\mathcal{S}}^0 &:= \mathbb{P}(X_{\mathcal{S}} = \mathbf{0}, X_{\mathcal{N}_j} = \mathbf{0}, X_i = X_j = 1) \qquad (34)\\
&\geq (1-p)^s(1-q_{\text{low}})^d p^2/d^2 := w^0. \qquad (35)
\end{aligned}$$

According to the definition of hidden interferers and Lemma 2, if $i \notin \mathcal{S}_j$,

$$\begin{aligned}
&\min_{\substack{\mathcal{S}\subseteq\mathcal{V}\backslash\{\mathcal{N}_j\cup\{i,j\}\}\\|\mathcal{S}|\leq d}} \frac{u_{ij,\mathcal{S}}^0}{w_{ij,\mathcal{S}}^0} \\
&:= \min_{\substack{\mathcal{S}\subseteq\mathcal{V}\backslash\{\mathcal{N}_j\cup\{i,j\}\}\\|\mathcal{S}|\leq d}} \mathbb{P}(Y_j = 0 | X_{\mathcal{S}} = \mathbf{0}, X_{\mathcal{N}_j} = \mathbf{0}, X_i = X_j = 1) \\
&= \mathbb{P}(Y_j = 0 | X_{\mathcal{S}_j} = \mathbf{0}, X_{\mathcal{N}_j} = \mathbf{0}, X_i = X_j = 1) \\
&= \frac{u_{ij,\mathcal{S}_j}^0}{w_{ij,\mathcal{S}_j}^0} = p_{\emptyset_j} \leq p_{\emptyset}.
\end{aligned}$$

Otherwise, if $i \in \mathcal{S}_j$,

$$\begin{aligned}
&\min_{\substack{\mathcal{S}\subseteq\mathcal{V}\backslash\{\mathcal{N}_j\cup\{i,j\}\}\\|\mathcal{S}|\leq d}} \frac{u_{ij,\mathcal{S}}^0}{w_{ij,\mathcal{S}}^0} \\
&= \mathbb{P}(Y_j = 0 | X_{\mathcal{S}_j\backslash i} = \mathbf{0}, X_{\mathcal{N}_j} = \mathbf{0}, X_i = X_j = 1) \\
&= \frac{u_{ij,\mathcal{S}_j\backslash i}^0}{w_{ij,\mathcal{S}_j\backslash i}^0} \geq p_{ij} \geq p_{\min}. \qquad (36)
\end{aligned}$$

Set δ to be in the range $p_\emptyset < \delta < p_{\min}$. Based on analysis above, we now examine the type I and type II errors.

$$\begin{aligned}
\mathbb{P}(\mathcal{F}_{ij}^I) &= \mathbb{P}(\hat{p}_{ij} > \delta | i \notin \mathcal{S}_j) \qquad (37)\\
&= \mathbb{P}\left(\min_{\substack{\mathcal{S}\subseteq\mathcal{V}\backslash\{\mathcal{N}_j\cup\{i,j\}\}\\|\mathcal{S}|\leq d}} \frac{\bar{U}_{ij,\mathcal{S}}}{\bar{W}_{ij,\mathcal{S}}} > \delta\right) \\
&\leq \mathbb{P}\Big(\frac{\bar{U}_{ij,\mathcal{S}_j}}{\bar{W}_{ij,\mathcal{S}_j}} > \delta\Big) \\
&= \mathbb{P}\Big(\frac{\bar{U}_{ij,\mathcal{S}_j}}{\bar{W}_{ij,\mathcal{S}_j}} - p_{\emptyset_j} > \delta - p_{\emptyset_j}\Big) \\
&\leq \exp\Big\{ - \frac{k((\delta - p_{\emptyset_j})(w_{ij,\mathcal{S}_j}^0 - \delta_w) - p_{\emptyset_j}\delta_w)^2}{2}\Big\} \\
&\quad + \exp\Big\{ - \frac{k\delta_w^2}{2}\Big\} \qquad (38)\\
&\leq \exp\Big\{-\frac{k((\delta - p_\emptyset)w^0 - \delta\delta_w)^2}{2}\Big\} + \exp\Big\{-\frac{k\delta_w^2}{2}\Big\}, \qquad (39)
\end{aligned}$$

where δ_w is parameter satisfying $(\delta - p_\emptyset)w^0 - \delta\delta_w > 0$, (38) follows from Lemma 3, and (39) follows from Assumption A7 and (34).

On the other hand,

$$\begin{aligned}
&\mathbb{P}(\mathcal{F}_{ij}^{II}) = \mathbb{P}(\hat{p}_{ij} < \delta | i \in \mathcal{S}_j) \\
&= \mathbb{P}\Big(\min_{\substack{\mathcal{S}\subseteq\mathcal{V}\backslash\{\mathcal{N}_j\cup\{i,j\}\}\\|\mathcal{S}|\leq d}} \frac{\bar{U}_{ij,\mathcal{S}}}{\bar{W}_{ij,\mathcal{S}}} < \delta\Big) \\
&= \mathbb{P}\left(\bigcup_{\substack{\mathcal{S}:\mathcal{S}\subseteq\mathcal{V}\backslash\{\mathcal{N}_j\cup\{i,j\}\}\\|\mathcal{S}|\leq d}} \frac{\bar{U}_{ij,\mathcal{S}}}{\bar{W}_{ij,\mathcal{S}}} < \delta\right) \\
&\leq \sum_{\substack{\mathcal{S}:\mathcal{S}\subseteq\mathcal{V}\backslash\{\mathcal{N}_j\cup\{i,j\}\}\\|\mathcal{S}|\leq d}} \mathbb{P}\left(\frac{\bar{U}_{ij,\mathcal{S}}}{\bar{W}_{ij,\mathcal{S}}} < \delta\right) \\
&= \sum_{\substack{\mathcal{S}:\mathcal{S}\subseteq\mathcal{V}\backslash\{\mathcal{N}_j\cup\{i,j\}\}\\|\mathcal{S}|\leq d}} \mathbb{P}\Big(\frac{\bar{U}_{ij,\mathcal{S}}}{\bar{W}_{ij,\mathcal{S}}} - \frac{u_{ij,\mathcal{S}}^0}{w_{ij,\mathcal{S}}^0} < \delta - \frac{u_{ij,\mathcal{S}}^0}{w_{ij,\mathcal{S}}^0}\Big).
\end{aligned} \qquad (40)$$

According to (36), we have $\delta - \frac{u_{ij,\mathcal{S}}^0}{w_{ij,\mathcal{S}}^0} \leq \delta - p_{\min} < 0$. Thus,

$$\begin{aligned}
&\mathbb{P}\Big(\frac{\bar{U}_{ij,\mathcal{S}}}{\bar{W}_{ij,\mathcal{S}}} - \frac{u_{ij,\mathcal{S}}^0}{w_{ij,\mathcal{S}}^0} < \delta - \frac{u_{ij,\mathcal{S}}^0}{w_{ij,\mathcal{S}}^0}\Big) \\
&\leq \exp\left\{-\frac{k((\frac{u_{ij,\mathcal{S}}^0}{w_{ij,\mathcal{S}}^0} - \delta)(w_{ij,\mathcal{S}}^0 + \delta_w') - \frac{u_{ij,\mathcal{S}}^0}{w_{ij,\mathcal{S}}^0}\delta_w')^2}{2}\right\} \\
&\quad + \exp\Big\{-\frac{k(\delta_w')^2}{2}\Big\} \qquad (41)\\
&\leq \exp\Big\{-\frac{k((p_{\min} - \delta)w^0 - \delta\delta_w')^2}{2}\Big\} + \exp\Big\{-\frac{k(\delta_w')^2}{2}\Big\},
\end{aligned}$$

where δ'_w is parameter satisfying $(p_{\min} - \delta)w^0 - \delta\delta'_w > 0$, and (41) follows from Lemma 3. Plugging in (40), we have

$$\mathbb{P}(\mathcal{F}_{ij}^{II}) \leq n^s \Bigg[\exp\left\{ -\frac{k((p_{\min} - \delta)w^0 - \delta\delta'_w)^2}{2} \right\} + \exp\left\{ -\frac{k(\delta'_w)^2}{2} \right\} \Bigg]. \tag{42}$$

Let $\delta = \frac{p_{\min}+p_\emptyset}{2}$, $\Delta = \frac{p_{\min}-p_\emptyset}{2}$, and $\delta_w = \delta'_w = \frac{\Delta w^0}{1+\delta}$. Combining (31), (39) and (42), we have

$$\begin{aligned} &\mathbb{P}\Big(\mathcal{F}^0 \cup \big(\cup_{(i,j)\notin\mathcal{E}_H} \mathcal{F}_{ij}^I \big) \cup \big(\cup_{(i,j)\in\mathcal{E}_H} \mathcal{F}_{ij}^{II} \big)\Big) \\ &\leq \mathbb{P}(\mathcal{F}^0) + \sum_{(i,j)\notin\mathcal{E}_H} \mathbb{P}(\mathcal{F}_{ij}^I) + \sum_{(i,j)\in\mathcal{E}_H} \mathbb{P}(\mathcal{F}_{ij}^{II}) \\ &\leq n^2 \exp\{-2\epsilon^2 k\} + n^{s+2} \cdot 2\exp\Big\{ -\frac{k\delta_w^2}{2} \Big\} \\ &\leq \gamma/2 + \gamma/2 = \gamma. \end{aligned}$$

The last inequality holds when

$$k \geq \max\left\{ \frac{1}{\epsilon^2}\left(\log n + \frac{1}{2}\log\frac{2}{\gamma}\right), \frac{2}{\delta_w^2}\left((s+2)\log n + \log\frac{4}{\gamma}\right) \right\}.$$

References

[1] J. Yang, S. Draper, and R. Nowak, "Passive learning of the interference graph of a wireless network," *IEEE International Symposium on Information Theory*, July 2012.

[2] D. Niculescu, "Interference map for 802.11 networks," *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, 2007.

[3] N. Ahmed and S. K. Smarta, "A self-managing architecture for thin access points," *Proceedings of the ACM CoNEXT conference*, 2006.

[4] Y.-C. Cheng, J. Bellardo, P. Benko, A. C. Snoeren, G. M. Voelker, and S. Savage, "Jigsaw: solving the puzzle of enterprise 802.11 analysis," *ACM SIGCOMM*, 2006.

[5] Y.-C. Cheng, M. Afanasyev, P. Verkaik, P. Benko, J. Chiang, A. C. Snoeren, S. Savage, and G. M. Voelker, "Automating cross-layer diagnosis of enterprise wireless networks," *ACM SIGCOMM*, 2007.

[6] R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan, "Analyzing the MAC-level behavior of wireless networks in the wild," *ACM SIGCOMM*, 2006.

[7] V. V. Shrivastava, "Optimizing enterprise wireless networks through centralization," Ph.D. dissertation, University of Wisconsin–Madison, 2010.

[8] V. Shrivastava, S. Rayanchu, S. Banerjee, and K. Pagaginnaki, "PIE in the sky: Online passive interference estimation for enterprise wlans," *8th USENIX Symposium on Networked Systems Design and Implementation*, 2011.

[9] *IEEE 802.15 WPAN Task Group 6 (TG6) Body Area Networks*, IEEE Standards Association. 09-Jun-2011. Retrieved 09-dec-2011.

[10] Http://www.zigbee.org/About/AboutAlliance/TheAlliance.aspx.

[11] *IEEE 802.11 standard*, http://www.ieee802.org/11/.

[12] R. M. Karp, *Reducibility Among Combinatorial Problems*. New York: Plenum, 1972.

[13] M. Kalisch and P. Büuhlmann, "Estimating high-dimensional directed acyclic graphs with the PC-algorithm," *Journal of Machine Learning Research*, March 2007.

[14] T. Cour, B. Sapp, and B. Taskar, "Learning from partial labels," *Journal of Machine Learning Research*, April 2011.

[15] A. Tsybakov, *Introduction à l'estimation non-paramétrique*. *Mathématiques et Application*. Springer, 2004.