ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Blind wireless network topology inference

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Testi, E., Giorgetti, A. (2021). Blind wireless network topology inference. IEEE TRANSACTIONS ON COMMUNICATIONS, 69(2), 1109-1120 [10.1109/TCOMM.2020.3036058].

Availability:
This version is available at: https://hdl.handle.net/11585/859426 since: 2022-02-16

Published:

DOI: http://doi.org/10.1109/TCOMM.2020.3036058

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

27 November 2024

Blind Wireless Network Topology Inference

Enrico Testi, *Student Member, IEEE*, and Andrea Giorgetti, *Senior Member, IEEE*

***Abstract*— This work proposes a framework to discover the topology of a non-collaborative packet-based wireless network using radio-frequency (RF) sensors. The methodology developed is blind, allowing topology sensing of a network whose key features (i.e., number of nodes, physical layer signals, and medium access control (MAC) and routing protocols) are unknown. Because of the wireless medium, over-the-air signals captured by the sensors are mixed; therefore, blind source separation (BSS) and measurement association are used to separate traffic patterns. Then, to infer the topology, we detect directed data flows among nodes by identifying causal relationships between the separated transmitted patterns. We propose causal inference methods such as Granger causality (GC), transfer entropy (TE), and conditional transfer entropy (CTE) that use the times series of traffic profiles, and a solution based on a neural network (NN) that exploits distilled time-based features. The framework is validated on an ad-hoc wireless network accounting for MAC protocol, packet collisions, nodes mobility, the spatial density of sensors, and channel impairments, such as path-loss, shadowing, and noise. Numerical results reveal that the proposed approach reaches a high probability of link detection and a moderate false alarm rate in mild shadowing regimes and low to moderate network nodes mobility.**

***Index Terms*—Topology inference, blind source separation, independent component analysis, link detection, Granger causality, transfer entropy, neural network.**

I. Introduction

The importance of networks is massively growing in modern-day society thanks to unprecedented communication capabilities offered by technology. A group of users in a social network, a sensor network deployed to collect environmental data, or a tactical network aiming to exchange information between soldiers, are only a few examples of networks widely diffused today.

In this scenario of ultra-densely connected objects, topology discovery is an essential aspect that can help predict traffic flow, infer communications between nodes, estimate the network connectivity, detect communities, help network maintenance, and performs optimization and orchestration. Moreover, in defense applications, understanding the structure of an adversary's network may considerably help avoid dangerous situations, make predictions, and design decision-making strategies. Focusing on wireless networks, the rapidly growing demand for radio services by billions of devices makes the radio spectrum an increasingly valuable resource. From this perspective, while the current cognitive radio (CR) paradigm is grounded on spectrum sensing, more in-depth knowledge of how a network uses the wireless medium and, thus, the structure of the network itself, may contribute to the development of much effective spectrum sharing strategies [1]–[5].

For this reason, there has been a rising interest in the possibility of reconstructing a network's structure from few observed quantities at some nodes (or at the edges) with little, if not zero, prior knowledge [6]–[9]. If the problem appears rather complicated for a wired network, it can be even more challenging in a wireless scenario because of interference, path-loss, shadowing, fading, and the so-called hidden terminal problem. While connectivity between the nodes could be inferred based upon the distance between them, since many nodes can be within the range of many others, inferring which ones are communicating might better rely on their activity patterns.

A. Existing works

There are different approaches and methodologies for network topology inference proposed in the literature. Some of them rely on access to the packet's content, which is not always feasible and may increment the network overhead [6], [7]. In [10], a path inference approach that exploits routing information within packets is proposed. Others fall into the network tomography category, which requires access to information at endpoints [8], [9]. For example, in [11], a low-complexity inference algorithm based on the Kullback-Leibler (KL) divergence that requires a link rate estimation is developed. Without accessing the packet content, the solution proposed in [12] exploits spectral coherence based on the Lomb-Scargle periodogram as a measure of causality between two signals. Such an approach relies upon the notion of correlation, which, in principle, does not necessarily imply causation. In [13], a Bayesian nonparametric model to learn the topology of an unknown ad-hoc network is proposed; the solution is based on Hidden Semi-Markov model (HSMM) for segmenting nodes transmission activity.

A different research field that contributes to topology inference is represented by graph signal processing (GSP) applied to networks. Graph learning as an edge subset selection problem or a neighborhood-based sparse linear regression is proposed in [14]. In [15], non-linear structural equation models for detecting the topology of a graph from the observations of a process propagating through it are investigated. In [16], a novel method based on an elastic net solver [17], that performs well even in scenarios where the data are highly correlated, is presented. However, topology inference in GSP assumes that the network is subject to a diffusion process, i.e., there is a piece of information propagating among all the nodes.

This work was supported by MIUR under the program "Departments of Excellence (2018-2022) - Precise-CPS," and the CoACh project funded by the POR FESR 2014-2020 program. Part of this work has been presented at the IEEE Int. Conf. on Signal Proc. and Comm. Systems (ICSPCS), Dec. 2019.
The authors are with the Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi" (DEI), CNIT, University of Bologna, Italy (e-mail: {enrico.testi4, andrea.giorgetti}@unibo.it).

The task of network topology inference can be seen as learning temporal causal structures among multiple time series. This has roots in the causal inference problem described by Pearl [18] and Granger [19]–[21]. In particular, the Granger test based on a auto-regressive (AR) model introduced in [19] has become the basis for the development of further causal analysis and topology inference methods. An approach for causal inference on networks involving a specific formulation of Granger causality (GC), named asymmetric Granger causality (AGC), is exploited in [22], where the parametric tests are carried out over groups of time series. Hawkes point processes are a statistical tool to model causal relationships, and recently their connection with GC has been investigated [23]. The use of multivariate Hawkes processes for topology inference through causal analysis between time series is exploited in [24], [25]. Another well-known tool for causal inference is based on the information-theoretic measure, named transfer entropy (TE), proposed in [26]. In [27], the authors propose a TE-based topology inference approach and evaluate its robustness with respect to GC. To overcome some limits of TE, [28] proposes causation entropy (CE), an approach for causal inference on networks that is optimum under certain *Markovian* assumptions. A non-parametric learning method related to GC and TE that measure the impact of one node activity over another is developed in [29]. An interesting feature of this approach is that prior knowledge of the underlying network protocol is not needed.

B. Our contribution

Previous contributions on wireless network topology inference assume that the temporal profiles of packets sent by each node are perfectly known, which presumes the absence of noise, interference, and channel impairments. In a realistic scenario, this is possible only if sensors are in close proximity to the nodes, which in turn require that every node is physically accessible and that the number of sensors is the same as the number of nodes. To overcome such limitations, in this work, we propose a new framework for topology inference of a non-collaborative wireless network where the observations of over-the-air signals are carried out by sensors randomly distributed in the network landscape, as depicted in Fig. 1. The contributions of this work are the following.

- Assuming that most of the network's features are unknown (the number of nodes, their exact position, the traffic type, the communication protocol), we aim to learn as much information as possible about its structure observing it from the outside. We estimate the number of sources (network nodes) and separate the traffic patterns at the sensors by combining techniques for blind source separation (BSS) and a novel solution for the measurement association.
- We propose a new approach for topology inference based on machine learning (ML) and the extraction of time-related features from the observed data, turning the problem of identifying causal relations to binary classification. Such an approach applies to statistical time-division multiplexing (STDM) based networks, which encompasses a variety of multiple access algorithms (e.g., CSMA/CA), and appear to be lightweight than known methods, especially when considering networks with many nodes.
- Our analysis accounts for noise, propagation impairments (i.e., path-loss and shadowing), and mobility of network nodes. In particular, we investigate how shadowing affects the quality of the reconstructed traffic patterns, hence the accuracy of the topology discovered.

C. System architecture and applications

The proposed framework aims to infer the topology of a non-collaborative network, either because it is competing for the spectrum usage, or it is private, encrypted, and hence cannot be accessed. Thus, the BSS and the topology inference are performed from the outside. In particular, for the sake of clarity, two specific application scenarios are the following:

- Defense: in this scenario, a network of radio-frequency (RF) sensors can be deployed in an unknown environment to detect and extract information about an adversarial network's structure. The sensors collect over-the-air received power profiles and send them to a fusion center (FC), which can be either one of the sensors or a specific device. The FC performs the BSS and then infers the topology of the adversarial network. Since the sensors can continuously monitor the received power, if the time spent to send the data to the FC is negligible, the topology inference can be performed online.
- Spectrum awareness in CR: in this scenario, a primary wireless network wishes to know if another network uses the same spectrum (legitimately or not). In this case, the wireless nodes schedule a periodic sensing phase to sense the RF medium [3], [30], [31]. While in conventional periodic sensing nodes try to detect a transmission, in our setting, nodes collect power samples for the FC. The FC then detects the adversarial network and performs BSS and topology inference. Once the topology is inferred, the wireless network can make decisions about the spectrum usage, perform communication optimization based on the adversarial network's behavior, or notify spectrum regulators about violations by nonlegitimate communications. In this scenario, the time spent to sense the spectrum reduces the primary network's throughput while gaining insight into the spectrum usage.

Throughout the paper, capital boldface letters denote matrices, lowercase bold letters denote vectors, $(\cdot)^{\mathrm{T}}$ stands for transposition, $||\cdot||_p$ is the l_p-norm, $|\cdot|$ is the module operator, $\odot$ stands for the element-wise product. With $v_{i,j}$, $\mathbf{v}_{i,:}$, and $\mathbf{v}_{:,j}$, we represent, respectively, the element, the ith row, and the jth column of the matrix $\mathbf{V}$ (when unambiguous, the ith row of $\mathbf{V}$ is $\mathbf{v}_i$), and with $\mathbf{v}_{i,j:k}$ we select the elements between the jth and the kth entry of the ith row of $\mathbf{V}$, extremes included. We use $\mathcal{N}(\mu, \sigma^2)$ to denote a real Gaussian distribution with mean μ and variance σ^2, $\mathcal{U}(a, b)$ to denote a uniform distribution between a and b, $\mathbb{E}\{\cdot\}$ to denote the expectation operator, and $\langle\cdot\rangle$ to indicate the sample mean operator.

The remainder of this paper is organized as follows. In Section II, we introduce the scenario and system model.

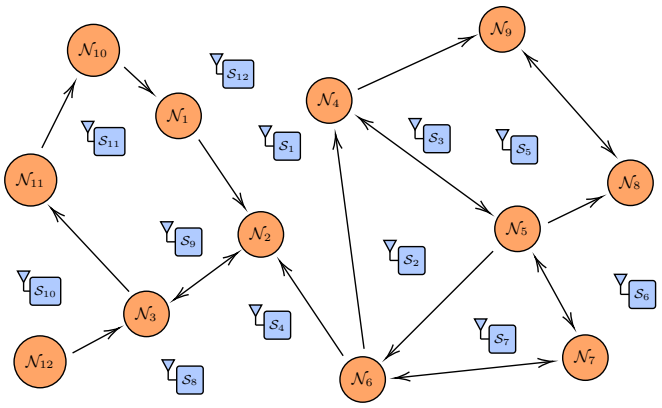

Fig. 1. A cloud of randomly distributed RF sensors (in blue) across the wireless network landscape (in orange).

Section III provides an overview of the blind source separation problem and describes the proposed solutions. In Section IV, the state-of-the-art methods for topology inference are presented, and our novel solution is introduced. Numerical results are given in Section V. Conclusions are drawn in Section VI.

II. System Overview and Problem Setup

Let us consider a scenario with a non-collaborative wireless network of N nodes (the network in the following) and a network of M RF sensors (the sensors in the following) randomly deployed on a two-dimensional landscape at known positions.[1] We assume that the technical specifications of the network (i.e., number of nodes, physical layer signals, and medium access control (MAC) and routing protocols) are unknown. The topology of the network is represented by a directed graph and its associated adjacency matrix $\mathbf{A} \in \{0,1\}^{N\times N}$ where

$$a_{i,j} = \begin{cases} 1 & \text{if information flows from node } i \text{ to } j \\ 0 & \text{otherwise.} \end{cases}$$

For example, with reference to Fig. 1, $a_{12,3} = 1$, $a_{3,12} = 0$, and $a_{3,2} = a_{2,3} = 1$.

The goal is to find an estimate of the adjacency matrix, $\widehat{\mathbf{A}}$, of the wireless network from raw RF measurements carried out by sensors within an observation window of duration T_{ob}. Since no interaction is expected between the observed network and the sensors, all the subsequent tasks, summarized in Fig. 2, are performed without demodulating the received signals, so that a simple energy detector (ED) receiver suffices [31], [34], [35].

A. Data acquisition and channel model

Let us consider the equivalent low-pass representation of the signal received by the mth sensor

$$r_m(t) = \sum_{n=1}^{N} q_n(t) g_{m,n} + \nu_m(t) \tag{1}$$

where $q_n(t)$ is the signal transmitted by node n, $g_{m,n}$ is the channel gain between node n and sensor m, and $\nu_m(t)$ is the additive white Gaussian noise (AWGN) with independent, identically distributed (i.i.d.) real and imaginary parts, each with two-sided power spectral density N_0.

For the sake of topology inference we seek to collect the transmitted power profiles for each node. Let us define such profile at node n as a vector of K samples $\mathbf{p}_n = (p_{n,1}, p_{n,2}, \ldots, p_{n,K})^{\mathrm{T}}$, whose elements

$$p_{n,k} = \frac{1}{T_{\text{b}}} \int_{(k-1)T_{\text{b}}}^{kT_{\text{b}}} |q_n(t)|^2 \mathrm{d}t \qquad k = 1, \ldots, K \tag{2}$$

correspond to the transmitted power calculated over short intervals of duration T_{b} such that $T_{\text{ob}} = KT_{\text{b}}$. Then, let us collect the transmitted power profiles in the matrix

$$\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N)^{\mathrm{T}} \in \mathbb{R}^{N\times K}. \tag{3}$$

Similarly, the output of the RF sensor m is a vector, $\mathbf{x}_m = (x_{m,1}, x_{m,2}, \ldots, x_{m,K})^{\mathrm{T}}$, whose samples correspond to the received power within T_{b}, i.e.,

$$x_{m,k} = \frac{1}{T_{\text{b}}} \int_{(k-1)T_{\text{b}}}^{kT_{\text{b}}} |r_m(t)|^2 \mathrm{d}t \qquad k = 1, \ldots, K. \tag{4}$$

The samples (4) are obtained by an ED composed of a bandpass zonal filter with bandwidth W (the same of the transmitted signals), followed by a square-law device and an integrator with finite integration time T_{b} [31], [34]. Collecting the measured powers of all sensors in the matrix

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M)^{\mathrm{T}} \in \mathbb{R}^{M\times K}$$

it is possible to relate $\mathbf{X}$ with $\mathbf{P}$ by

$$\mathbf{X} = \mathbf{H}\mathbf{P} + \mathbf{N} \tag{5}$$

where $\mathbf{H} \in \mathbb{R}^{M\times N}$ is the matrix of power gains $h_{m,n} = |g_{m,n}|^2$, and $\mathbf{N} \in \mathbb{R}^{M\times K}$ is the matrix of noise power samples at the output of the ED, $n_{m,k} = \frac{1}{T_{\text{b}}} \int_{(k-1)T_{\text{b}}}^{kT_{\text{b}}} |\nu_m(t)|^2 \mathrm{d}t$. The elements of $\mathbf{N}$ are i.i.d. central chi-squared random variables (r.v.s) with a number of degrees of freedom $N_{\text{DOF}} = 2WT_{\text{b}}$ [34].[2] If N_{DOF} is large enough, by the central limit theorem, we can approximate the distribution of the elements of $\mathbf{N}$ as $n_{m,k} \sim \mathcal{N}(\sigma^2_{\text{N}}, 2\sigma^4_{\text{N}}/N_{\text{DOF}})$, where $\sigma^2_{\text{N}} = 2N_0W$ [31], [34]. The channel gain consists of two components $h_{m,n} = h'_{m,n} e^{\sigma_{\text{S}} s_{m,n}}$, where $h'_{m,n}$ is the path gain, and $s_{m,n} \sim \mathcal{N}(0,1)$ are i.i.d. Gaussian r.v.s to model log-normal shadowing with intensity σ_{S} [36].[3] In deriving (5) we consider that the signals emitted by the network nodes are mutually uncorrelated and uncorrelated with the noise.

B. Blind source separation

Topology inference requires the temporal characterization of the transmitted packets for each node of the wireless network. Therefore, a reconstruction of the temporal traffic profiles, $\mathbf{P}$, as if they were measured at each node, is needed. However,

[1] Network nodes localization can be performed by the sensor network in a phase preceding topology inference [32], [33]. As will be clarified in Section V, only a coarse estimate of nodes position is required. Without loss of generality, the numerical results are derived considering the position (x_m, y_m) of the mth sensor uniformly distributed, i.e., $x_m, y_m \sim \mathcal{U}(-L/2, L/2)$, where L is the side length of a squared landscape.

[2] The intervals duration, T_b, is chosen such that $N_{\text{DOF}} = 2WT_{\text{b}}$ is integer.

[3] The shadowing parameter is usually expressed as the standard deviation of the channel loss in deciBel by $\sigma_{\text{S}}(\text{dB}) = \frac{10}{\ln 10}\sigma_{\text{S}}$.

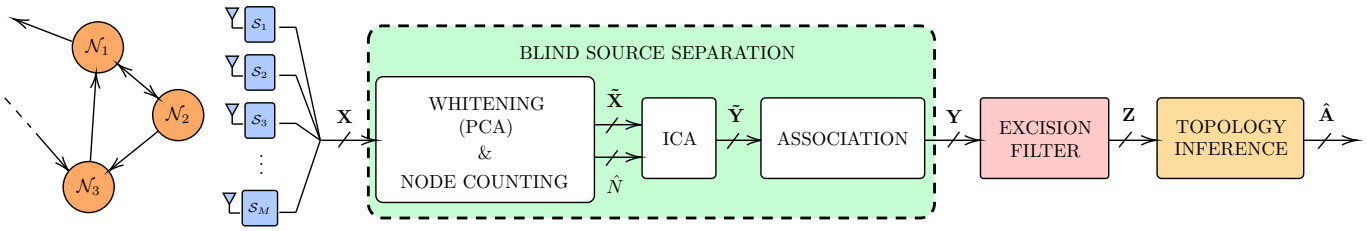

Fig. 2. Block scheme of the wireless network topology inference: sensor network, data pre-processing and inference.

because of the wireless medium, sensors observe a mixture of the signals emitted by the nodes, (5), and an unmixing operation is required to extract $\mathbf{P}$ [37], [38].

In literature, various methods for separating mixed signals have been proposed, e.g., matrix factorization [39] and tensor decomposition [40], [41], to name a few. In this work, we adopt an approach based on the combination of principal component analysis (PCA) and Fast-independent component analysis (ICA) techniques [42], and we compare it with a simpler approach named spatial filtering (SF). We then propose a novel solution to the measurement association problem. Since the unmixing operation is not perfect because of the presence of noise and shadowing, the output $\mathbf{Y}$ of BSS contains residual transmitted power profiles from other nodes (crosstalk) which is removed by an excision filter.

C. Excision filter

The topology inference algorithms presented in Section IV are based on the temporal characterization of the packet flows exchanged by the nodes. To extract temporal features, it is necessary to process the time series in $\mathbf{Y}$ to obtain sequences of 0s and 1s; this is carried out by an excision filter that forces to zero power samples due to crosstalk. The output of the filter is the matrix $\mathbf{Z}$ with elements

$$z_{n,k} = \begin{cases} 1 & \text{if } y_{n,k} \geq \lambda_n \\ 0 & \text{otherwise} \end{cases}$$

where the threshold λ_n is set as a fraction $q \in [0,1]$ of the maximum of $\mathbf{y}_{n,:}$, i.e.,

$$\lambda_n = q \cdot \max_k \{y_{n,k}\}, \quad n = 1, \ldots, N. \tag{6}$$

D. Topology inference

Topology inference aims to estimate the adjacency matrix, $\widehat{\mathbf{A}}$, exploiting relations (causality) between the traffic streams contained in the matrix $\mathbf{Z}$.

A first issue in testing the presence of information flow (packet exchange) between two nodes occurs when they are not communicating with each other, but their distance does not guarantee the absence of collisions. In such a case, the causality between the transmitted power profiles might not be zero, leading to false link detection. An opposite phenomenon is link misdetection, because of an unfavorable placement of the sensors in the landscape that degrade BSS performance. In this case, imperfect unmixing may lead to crosstalk, and detection of a causal relationship difficult. Another challenge is the time-varying nature of network topology. In these scenarios, the complexity is an important aspect that has to be taken into account when designing the inference algorithm.[4] In Section IV we present two state-of-the-art methods and our novel strategy for topology inference, based on GC, TE, and neural network (NN), respectively.

III. Blind Source Separation

BSS recovers the source signals $\mathbf{P}$ in (5), from a set of observed quantities $\mathbf{X}$, when the mixing matrix $\mathbf{H}$ is unknown [37]. Here we assume $M \geq N$, i.e., more sensors than sources; in this setting, the problem is named overdetermined.

A. PCA whitening and estimation of the number of sources

Signal separation can be effective if a pre-processing stage manipulates the data so that there are N mixtures centered and whitened at its output [37]. Since the number of network nodes N is unknown, this stage has to estimate the number of sources.

Firstly, we center the mixtures subtracting the row-wise mean from $\mathbf{X}$. Then, we apply PCA, a linear transformation to the observations so that their components become uncorrelated (whitening) with unit variance. Starting from the sample covariance matrix of the observations $\mathbf{\Sigma} = \frac{1}{K}\mathbf{X}\mathbf{X}^{\mathrm{T}}$, we perform the eigenvalue decomposition $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}}$, where $\mathbf{U}$ is the orthogonal matrix containing the eigenvectors, and $\mathbf{\Lambda}$ is a diagonal matrix of the eigenvalues, Λ_i, with $i = 1, \ldots, M$, sorted in descending order. Thus, the whitening matrix is

$$\mathbf{Q} = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{T}}. \tag{7}$$

To estimate the number of sources $\hat{N}$ generating the mixture, we use model order selection based on minimum description length (MDL) criteria [43], [44]

$$\hat{N} = \underset{n \in \{1,..,M\}}{\arg\min} \{\mathsf{MDL}(n)\} \tag{8}$$

with

$$\mathsf{MDL}(n) = -\log\left(\frac{\prod_{i=n+1}^{M}\Lambda_i^{1/(M-n)}}{\frac{1}{M-n}\sum_{i=n+1}^{M}\Lambda_i}\right)^{(M-n)K} + \frac{1}{2}n(2M-n)\log K$$

[4] In this regard, although it would be desirable to infer the entire network's topology simultaneously (i.e., through a multivariate method) to account for all the possible causal relationships among nodes, we test all the possible pairs of nodes separately. This does not imply that the pair under test is independent of the rest of the network, because the effects of the whole network, e.g., due to the collision avoidance mechanism, affect the features extracted. The advantage of this approach lies in a much computationally lighter approach [20].

Algorithm 1: Fast-ICA for BSS

Input : Whitened signals $\tilde{\mathbf{X}}$, $\epsilon_t > 0$, $\hat{N}$
Output: Unmixing matrix $\mathbf{W}$

$\epsilon \leftarrow \infty$
$\mathbf{W} \leftarrow$ random initialization such that $||\mathbf{W}||_1 = 1$
$\tilde{\mathbf{W}} \leftarrow \mathbf{W}$
for n *from* 1 *to* $\hat{N}$ **do**
  while $\epsilon \geq \epsilon_t$ **do**
    $\mathbf{w}_{i,n} \leftarrow 4\langle \tilde{\mathbf{x}}_{i,:} \odot (\tilde{\mathbf{w}}_{:,n}^{\mathrm{T}} \tilde{\mathbf{X}})^3 \rangle - 3\tilde{\mathbf{w}}_{i,n}, \ \forall i = 1, .., \hat{N}$
    $\mathbf{w}_{:,n} \leftarrow \mathbf{w}_{:,n} - \sum_{i=1}^{n-1} (\mathbf{w}_{:,n}^{\mathrm{T}} \mathbf{w}_{:,i}) \mathbf{w}_{:,i}$
    $\mathbf{w}_{:,n} \leftarrow \frac{\mathbf{w}_{:,n}}{||\mathbf{w}_{:,n}||_2}$
    $\epsilon = \sum_{m=1}^{\hat{N}} |w_{m,n} - \tilde{w}_{m,n}|$
    $\tilde{\mathbf{w}}_{:,n} \leftarrow \mathbf{w}_{:,n}$
  end
end

where n is the unknown model order. Once the number of sources is estimated, we project the mixture onto the subspace spanned by the eigenvectors corresponding to the $\hat{N}$ largest eigenvalues, reducing the dimensionality from M to $\hat{N}$. This is accomplished by a projection matrix $\tilde{\mathbf{Q}}$ obtained from the first $\hat{N}$ rows of $\mathbf{Q}$, so that the whitened mixture is

$$\tilde{\mathbf{X}} = \tilde{\mathbf{Q}}\mathbf{X}. \tag{9}$$

B. Independent Component Analysis

ICA is a data processing method that finds statistically independent and non-Gaussian components from data. In our setting, ICA is applied to $\tilde{\mathbf{X}}$ to reconstruct the transmitted power profiles $\mathbf{P}$ in (5). Its output is an unmixing matrix $\mathbf{W} \in \mathbb{R}^{\hat{N} \times \hat{N}}$ such that

$$\tilde{\mathbf{Y}} = \mathbf{W}^{\mathrm{T}} \tilde{\mathbf{X}} \tag{10}$$

where $\tilde{\mathbf{Y}} \in \mathbb{R}^{\hat{N} \times K}$ is the matrix of the separated components. We propose Fast-ICA, an iterative algorithm with kurtosis as a measure of non-Gaussianity, and decorrelation based on the Gram-Schmidt orthonormalization [42], [45]. In particular, the update rule to calculate the unmixing matrix is [42, eq. 20]

$$\mathbf{w}_{:,n} \leftarrow 4 \begin{pmatrix} \langle \tilde{\mathbf{x}}_{1,:} \odot (\tilde{\mathbf{w}}_{:,n}^{\mathrm{T}} \tilde{\mathbf{X}})^3 \rangle \\ \cdots \\ \langle \tilde{\mathbf{x}}_{\hat{N},:} \odot (\tilde{\mathbf{w}}_{:,n}^{\mathrm{T}} \tilde{\mathbf{X}})^3 \rangle \end{pmatrix} - 3\tilde{\mathbf{w}}_{:,n}. \tag{11}$$

The complete iterative method is reported in Algorithm 1, where ϵ_t is the termination parameter.

However, the order of recovered signals is not preserved; thus $\mathbf{P}$ could be obtained from $\tilde{\mathbf{Y}}$ through a permutation of the rows. In the next section, a novel solution to this issue, tailored for our scenario is proposed.

C. Unmixed signals association

We propose an iterative method to associate the reconstructed sequences to the nodes of the network and measure the correctness of this matching. On this purpose, let us define the matrix $\mathbf{D} \in \mathbb{R}^{M \times \hat{N}}$, whose elements $d_{m,n}$ are the distances between sensor m and node n. First, we select a node n from a list of all the nodes of the network and find its nearest RF sensor m. Then, we correlate the sequence measured at sensor m with all the unmixed sequences (rows of $\tilde{\mathbf{Y}}$) separately. The row $\tilde{\mathbf{y}}_{j,:}$ that shows the highest positive correlation is associated with n and, thus, is copied into the nth row of $\mathbf{Y}$. Then, we remove node n from the list, delete the jth sequence from $\tilde{\mathbf{Y}}$ and iterate the algorithm. The method is detailed in Algorithm 2. Its complexity is $\mathcal{O}(\hat{N} \log \hat{N})$, acceptable also when dealing with large networks.

Algorithm 2: Unmixed signals association

Input : Unmixed signal $\tilde{\mathbf{Y}}$, $\hat{N}$, $\mathbf{D}$
Output: Aligned unmixed signal $\mathbf{Y}$

for n *from* 1 *to* $\hat{N}$ **do**
  $i \leftarrow \arg\min_m \{d_{m,n}\}$
  for j *from* 1 *to length of* $\tilde{\mathbf{y}}_{:,1}$ **do**
    $peaks_j \leftarrow \max\{\mathrm{corr}(\tilde{\mathbf{y}}_{j,:}; \mathbf{x}_{i,:})\}^{\dagger}$
  end
  $k \leftarrow \arg\max_j \{\mathbf{peaks}\}$
  $\mathbf{Y}_{n,:} \leftarrow \tilde{\mathbf{y}}_{k,:}$
  $\tilde{\mathbf{Y}} \leftarrow \tilde{\mathbf{Y}} / \tilde{\mathbf{y}}_{k,:}^{\ddagger}$
end

† $\mathrm{corr}(\mathbf{a}; \mathbf{b})$ operator returns the cross-correlation between vectors $\mathbf{a}$ and $\mathbf{b}$.
‡ $\mathbf{A}/\mathbf{a}_{\mathbf{k},:}$ operator removes the kth row from the matrix $\mathbf{A}$.

D. Spatial filtering

As a benchmark, we consider the alternative to BSS proposed in [46], where a path-loss law is used to weight the sensors measurements and reconstruct the source signals via linear filtering. The estimation of the number of sources, $\hat{N}$, is obtained via the MDL criteria (9). Then, we build the matrix $\boldsymbol{\alpha} \in \mathbb{R}^{M \times \hat{N}}$ of weights inversely proportional to the path-loss between sensor m and node n as $\alpha_{m,n} = (d_{m,n})^{-\eta}$, where η is the filtering parameter. The matrix $\mathbf{Y} \in \mathbb{R}^{\hat{N} \times K}$ containing the temporal power profiles reconstructed for all the $\hat{N}$ nodes is given by

$$\mathbf{Y} = \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{X}.$$

As a result, the traffic profiles due to distant nodes are filtered out, to some extent, by the weights $\alpha_{m,n}$. This approach is simpler than the BSS algorithm based on ICA, but it requires the choice of a filtering parameter η that has to be tuned depending on the specific propagation scenario and sensor and nodes deployment. Furthermore, its performance is strongly influenced by the presence of shadowing, as shown in Section V.

IV. Topology Inference Models and Tests

In this section, GC and TE based approaches are presented as performance benchmark, and a novel solution based on NN is introduced.

A. Granger causality

GC test methods are based on linear L-order AR models. In particular, considering a pair of time series $\mathbf{z}_i$ and $\mathbf{z}_j$ (i.e., two rows of $\mathbf{Z}$ that correspond to the transmitted power profiles of nodes i and j, respectively) two models (hypotheses) can be formulated

$$\mathcal{H}_1 : z_{j,k} = \sum_{l=1}^{L} \beta_l z_{j,k-l} + \sum_{l=1}^{L} \gamma_l z_{i,k-l} + \varepsilon_k \tag{12}$$

$$\mathcal{H}_0 : z_{j,k} = \sum_{l=1}^{L} \delta_l z_{j,k-l} + \omega_k \tag{13}$$

where $\{\beta_l\}_{l=1}^{L}$, $\{\gamma_l\}_{l=1}^{L}$, and $\{\delta_l\}_{l=1}^{L}$ are the regression coefficients, and ε_k, ω_k, are samples of independent AWGN. The model (12) corresponds to hypothesis $\mathcal{H}_1$ and considers the possibility of a causal relationship between the two time series, while (13) is the null hypothesis $\mathcal{H}_0$ and excludes the contribution of the past values of $\mathbf{z}_i$ in the prediction of $\mathbf{z}_j$. Note that (13) can be seen as a particular case of (12) where $\gamma_l = 0$, $l = 1, \ldots, L$. This means that if $\mathbf{z}_i$ *Granger causes* $\mathbf{z}_j$ the prediction error of model (12) is less than the one of (13). On the other hand, if $\mathbf{z}_i$ has no causal influence on $\mathbf{z}_j$ the errors are approximately equal. In [20], [21] a GC test based on squared sum of residuals is proposed

$$\mathsf{GC}_{i\to j} = \frac{\sum_{t=1}^{T} |\omega_t|^2 - \sum_{t=1}^{T} |\varepsilon_t|^2}{\sum_{t=1}^{T} |\varepsilon_t|^2} \cdot \frac{T-2K-1}{K} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \theta \tag{14}$$

where $T = K - L$ and K is the length of the time series. Since both ε_k and ω_k are Gaussian distributed, the sum of squared residuals follows a central chi-squared distribution and the test (14) is then the ratio of chi-squared r.v.'s which results in a $\mathcal{F}$-distribution [20]

$$\mathsf{GC}_{i\to j} \sim \mathcal{F}(L, T-2L-1).$$

The threshold θ for the test can be set fixing the false alarm probability. Alternatively, in [47] a useful tool for quantifying the degree of connectivity between two nodes i and j, named causal magnitude, is defined as

$$\mathsf{F}_{i\to j} = \log\left(\frac{\mathsf{V}(\boldsymbol{\omega})}{\mathsf{V}(\boldsymbol{\epsilon})}\right) \tag{15}$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_T)$, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_T)$ and $\mathsf{V}(\cdot)$ is the unbiased estimator of the variance.

B. Transfer entropy

In [26], a model-independent method to measure the information flow between two random processes by a specific type of conditional mutual information named TE is proposed. Considering two time series $\mathbf{z}_i$ and $\mathbf{z}_j$, modeled as random processes, the TE from node i to node j can be expressed as

$$\begin{aligned}\mathsf{TE}_{i\to j}(R,Q) &= \mathcal{I}(z_{j,k}; \mathbf{z}_{i,k-1:k-R}, \mathbf{z}_{j,k-1:k-Q}) \\ &= \mathbb{E}\left\{ \log_2 \frac{p(z_{j,k}|\mathbf{z}_{i,k-1:k-R}, \mathbf{z}_{j,k-1:k-Q})}{p(z_{j,k}|\mathbf{z}_{j,k-1:k-Q})} \right\}\end{aligned} \tag{16}$$

where $\mathbf{z}_i^-$ and $\mathbf{z}_j^-$ denote the past samples of $\mathbf{z}_i$ and $\mathbf{z}_j$ up to time instant k, respectively. In general, the evaluation of conditional probability densities requires the knowledge of infinite past samples of $\mathbf{z}_i$ and $\mathbf{z}_j$. However, in this particular application, TE is calculated considering only R and Q past samples of $\mathbf{z}_i$ and $\mathbf{z}_j$, respectively. The decision threshold θ is obtained by the null distribution of the TE, estimated from an appropriate manipulation of the time series [27], setting a predefined false-alarm probability. Then, the test becomes

$$\mathsf{TE}_{i\to j} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \theta. \tag{17}$$

The flow of information from node i to node j might take some time to generate a response, i.e., sending an acknowledgment (ACK). To predict the causal interaction, the lags R and Q should be very large and, thus, the algorithm's complexity gets overwhelming. Hence, an additional interaction delay parameter, n_0, to select the past values of the time series, is proposed in [48]. The definition of TE is then modified as

$$\mathsf{TE}_{i\to j}(R,Q,n_0) = \mathcal{I}(z_{j,k}; \mathbf{z}_{i,k-n_0-1:k-n_0-R}, \mathbf{z}_{j,k-1:k-Q}). \tag{18}$$

Note that the interaction delay has been considered only on the time series $\mathbf{z}_i$. This because the useful information on $z_{j,k}$ has been extracted from its past $\mathbf{z}_{j,n-1:n-Q}$, and only the influence of $\mathbf{z}_i$ need to be investigated.

C. Conditional transfer entropy

TE is a simplified version of CE and, as shown in [46], it tends to overestimate the number of inferred links. If for a given node i, we evaluate $\mathrm{TE}_{i\to j}(R,Q,n_0)$, $j = 1, \ldots, N$ with $j \neq i$, and then apply the binary hypothesis test for each pair $\{i,j\}$, we identify the set of possible neighbours of node i. Then, to avoid the presence of spurious links, the set of possible neighbors should be tested again with a variant of TE called conditional transfer entropy (CTE), where the effects of all the possible neighbors on the causal inference are taken into account. CTE from node i to j is defined as

$$\begin{aligned}&\mathsf{CTE}_{i\to j}(R,Q,n_0,g) \\ &= \mathcal{I}(z_{j,k}; \mathbf{z}_{i,k-n_0-1:k-n_0-R} | \mathbf{z}_{j,k-1:k-Q}, \mathbf{z}_{g,k-1:k-Q})\end{aligned} \tag{19}$$

where $g = 1, \ldots, N$ with $g \neq i, j$. For a complete description of the CTE algorithm and all its details, please refer to [27].

D. Neural network based method

The previously described methods compute a test based on the entire time series. A different approach is to use time-based features to infer the presence of causality via binary classification [49]. Indeed, in case of traffic flow from node i to node j, a packet-ACK time relation is expected to be found, as depicted in Fig. 3.[5] We indicate with $\tau^{i\to j}$ the time elapsed between the end of a packet sent by node i and the beginning of a packet from node j. If $a_{i,j} = 1$ the packet transmitted by

[5]The method might also apply to protocols that do not support ACKs. In that case, temporal features are extracted from the inter-transmission time between the end of a packet sent by node i and the following transmission from node j.

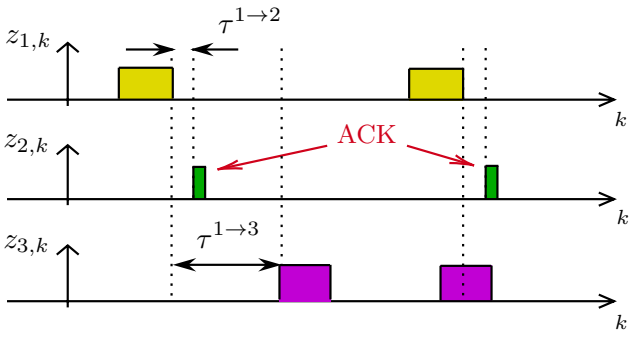

Fig. 3. An example of successful transmission of data packets between nodes $\mathcal{N}_1$ and $\mathcal{N}_2$ is shown. In this case node $\mathcal{N}_1$ is connected to node $\mathcal{N}_2$ but not to node $\mathcal{N}_3$. Notice how the time-to-ack $\tau^{1\to2}$ and $\tau^{1\to3}$ differ each other.

node j is likely to be an ACK. Denoting with N_{ta} the number of time-to-acks detected within the observation window, T_{ob}, there are three main relevant features which characterize the statistic of the time-to-ack $\tau^{i\to j}$:

- *Sample mean*

$$\mathsf{M}_{\tau^{i\to j}} = \frac{1}{N_{\text{ta}}} \sum_{p=1}^{N_{\text{ta}}} \tau_p^{i\to j} \tag{20}$$

- *Sample variance*

$$\mathsf{V}_{\tau^{i\to j}} = \frac{1}{N_{\text{ta}}-1} \sum_{p=1}^{N_{\text{ta}}} (\tau_p^{i\to j} - \mathsf{M}_{\tau^{i\to j}})^2 \tag{21}$$

- *Kurtosis*

$$\mathsf{K}_{\tau^{i\to j}} = \frac{\mathsf{m}_4^{i\to j}}{(\mathsf{m}_2^{i\to j})^2} \tag{22}$$

where $\mathsf{m}_4^{i\to j}$ and $\mathsf{m}_2^{i\to j}$ are respectively the 4th and the 2nd order moments, estimated from samples as

$$\mathsf{m}_q^{i\to j} = \frac{1}{N_{\text{ta}}} \sum_{p=1}^{N_{\text{ta}}} (\tau_p^{i\to j} - \mathsf{M}_{\tau^{i\to j}})^q. \tag{23}$$

The causal magnitude $\mathsf{F}_{i\to j}$ in (15) can be considered as additional feature to incorporate the benefits of GC for the classification. In the previous example, if the packet transmitted by node j is an ACK, the statistical features estimated from $\tau^{i\to j}$, e.g., the sample mean, will significantly change with respect to the case in which the packet is not an ACK.

If we consider, e.g., only two features, they can be represented on a plane such as the one reported in Fig. 4. Each red and blue point in the figure represents a couple $\{\mathsf{M}_{\tau^{i\to j}}, \mathsf{V}_{\tau^{i\to j}}\}$ extracted from time series measured at nodes i and j. In particular, a red point corresponds to the presence of a directed link from i to j, i.e., $a_{i,j} = 1$, while a blue point represents the absence of the link, that is $a_{i,j} = 0$. The color gradient shows the decision boundary identified by the classification algorithm. After a proper training phase, a classification algorithm (i.e., NN) can identify a boundary that separates the two classes. Accordingly, this approach needs a preliminary step where the features calculated from time series obtained by real or simulated networks are collected and used

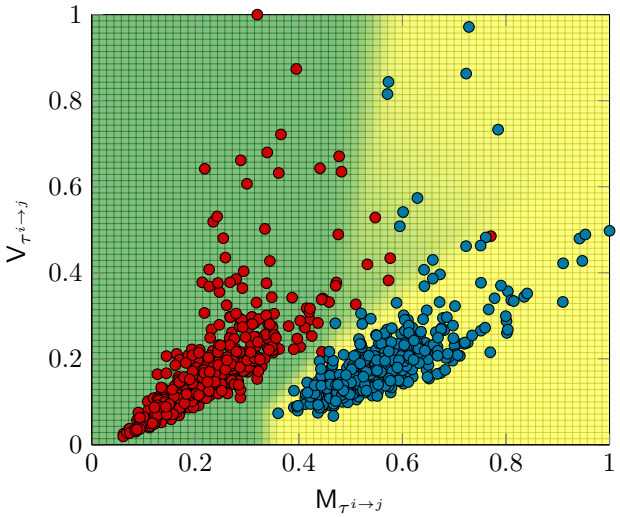

Fig. 4. Example of a two-dimensional normalized features space. In red: points corresponding to the presence of a link ($a_{i,j} = 1$). In blue: points representing couples of nodes that are not connected ($a_{i,j} = 0$). The color gradient shows the decision boundary identified by the classification algorithm.

for training. As it is shown in Fig. 4 the two groups of points are not linearly separable, therefore a NN has been selected as a proper classification algorithm in this work [50], [51]. Once the boundary has been found, it is possible to classify new points on-the-fly according to their position on the features space. In this way, we first perform the classification of every possible link and then merge all the outcomes to obtain the network topology. Such an approach applies to STDM based networks, which encompasses a variety of multiple access algorithms (e.g., CSMA/CA), and appear to be lightweight than known methods under certain conditions.

E. Computational complexity

In this section, we discuss the computational complexity of the topology inference algorithms as a function of the number of nodes of the network.

- **Granger causality.** A linear regression like (12) with K data points and $2L$ parameters has complexity $\mathcal{O}(K4L^2 + 8L^3)$. Similarly, including the linear regression in (13), and considering that $L^3 \ll K$ and that there are $N^2 - N$ couples of nodes in the network, the overall complexity is $\mathcal{O}(N^2 K 4L^2)$.
- **NN based method.** Since the training phase can be executed offline, we account for only the forward propagation in the complexity of the NN. The number of operations strictly depends on the number of neurons and layers and can be treated as a constant B. Thus, the complexity of the NN is $\mathcal{O}(N^2 B)$. The forward propagation is preceded by the feature extraction, whose complexity is dominated by the most computational expensive feature to extract, the causal magnitude — for this reason, considering that $L^2 \ll B$, the complete ML-based method has overall complexity $\mathcal{O}(N^2 K B)$.
- **Conditional transfer entropy.** The complexity is $\mathcal{O}(N^2 K C_1)$ and $\mathcal{O}(N^2 K C_2)$ for the two steps of the

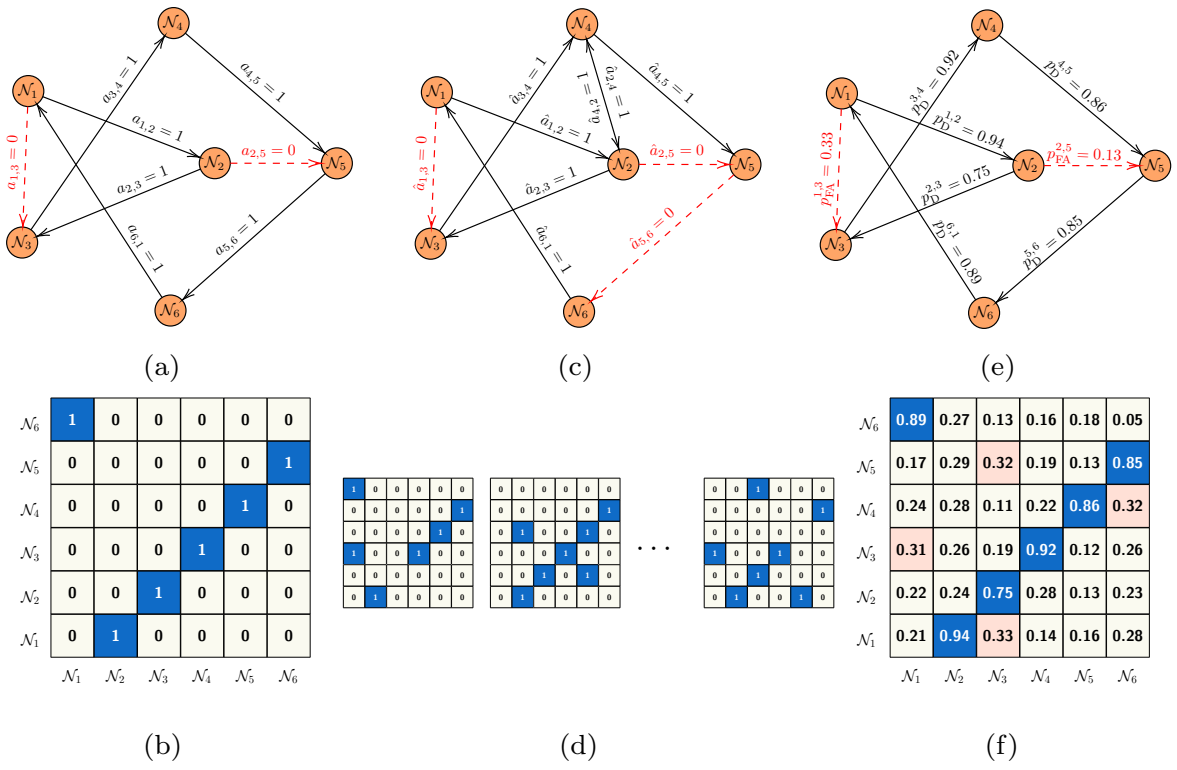

Fig. 5. (a) The directed graph representing the real topology of the network, whose adjacency matrix $\mathbf{A}$ is shown in (b); (c) and (d) examples of inferred topology graph and adjacency matrices $\hat{\mathbf{A}}$ estimated over the Monte Carlo trials of the same topology depicted in (a) varying the position of the nodes and of the the sensors in the landscape; (e) visualization of the performance where $p_{\mathrm{D}}^{i,j}$ and $p_{\mathrm{FA}}^{i,j}$ indicates the probability of detection and false alarm of the directed link from node i to node j, calculated over $M_{\mathrm{mc}} = 100$ Monte Carlo trials; the corresponding matrix of $p_{\mathrm{D}}^{i,j}$ and $p_{\mathrm{FA}}^{i,j}$ is shown in (f).

algorithm, respectively [27], [52]. C_1 and C_2 are two constants that take into account the operations for the choice of the interaction delay and the number of bootstraps iterations in both steps. Combining the two steps, the overall complexity is $\mathcal{O}(N^2K(C_1 + C_2))$.

Although the three algorithms have the same complexity trends, $\mathcal{O}(N^2K)$, considering a typical range of values for N in a practical scenario, the time complexity of the NN results considerably lower than the CTE. In fact, since in general $B \ll C_1 + C_2$, although they are constant factors, their values can differ by several orders of magnitude, so they are relevant for comparing the algorithms. This means that, in cases similar to those analyzed in the next section, the impact of such constant factors is not negligible.

V. Numerical Results

In this section, we present several tests to evaluate the performance of the whole processing chain, the impact of channel impairments on BSS, and compare the state-of-the-art solutions in topology inference with the NN-based method.

As a case study, we recreated an IEEE 802.11s ad-hoc network, operating at $f_0 = 2.412$ GHz, using a simulator developed through the *ns3* platform. The landscape containing the wireless network is a square area of side 10 m. The propagation scenario is characterized by omnidirectional antennas at the nodes and the sensors, path-loss, log-normal shadowing, and thermal noise. The path-loss model is of power-law type

TABLE I
Set of parameters used in the tests described in Section V

Parameter Set	A_0	A_1	A_2	B_0	B_1	B_2	C_0	C_1	C_2
ρ_{S} (nodes/m^2)	0.3	0.3	0.3	0.2	0.2	0.2	0.1	0.1	0.1
σ_{S} (dB)	0	3	6	0	3	6	0	3	6

with channel gain $h'_{m,n} = h_0(\frac{d_0}{d_{m,n}})^{\nu}$ where the path-loss exponent is $\nu = 3$, the reference distance is $d_0 = 1$ m, and $h_0 = -60.1$ dB [53]. The transmit power of the nodes is $P_{\mathrm{T}} = 10$ dBm, while the thermal noise power for both nodes and sensors is $\sigma_{\mathrm{N}}^2 = -93$ dBm. The bandwidth of the RF sensor is $W = 20$ MHz, according to the bandwidth of the IEEE 802.11s signal, while the integration time is $T_{\mathrm{b}} = 10\,\mu$s, hence $N_{\mathrm{DOF}} = 2WT_{\mathrm{b}} = 400$. The offered traffic of each node is 1 Mb/s, the size of data packets is set to 1024 Byte, while the ACK packets have a size of 112 Byte.

Regarding the signal processing chain, the loss parameter of the spatial filter, used as a benchmark, is set to $\eta = 4$, the excision filter threshold, λ_n, is set as in (6) with $q = 0.7$, and the termination parameter of the Fast-ICA is set to $\epsilon_t = 10^{-5}$. The results presented in this section are obtained from the data extracted by the simulations of $M_{\mathrm{top}} = 100$ different mesh topologies, such as the one depicted in Fig. 5. Then, for each topology, $M_{\mathrm{mc}} = 100$ Monte Carlo trials are performed to change the position of nodes and sensors randomly. In a real wireless network, the adjacency matrix $\mathbf{A}$ is sparse, hence the number of links to be detected is

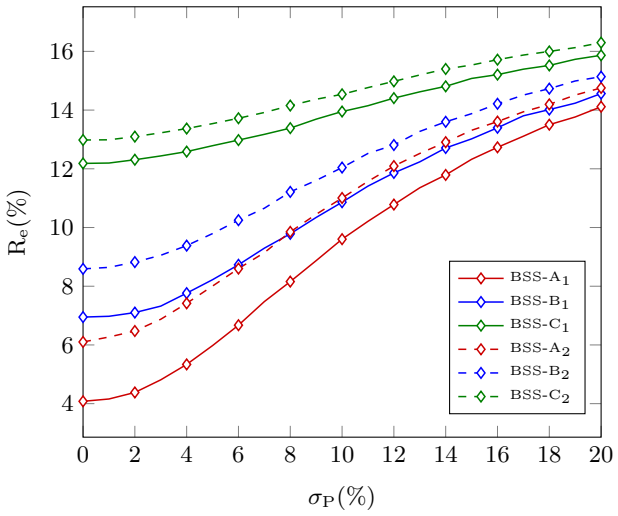

Fig. 6. BSS algorithm performance varying the standard deviation σ_{P} of the location uncertainty, the shadowing parameter σ_{S} and the density ρ_{S}.

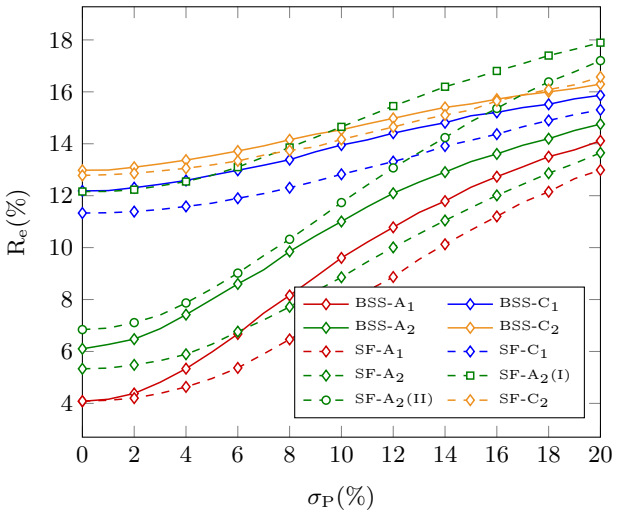

Fig. 7. Performance of BSS compared to the SF parametric method varying ρ_{S} and σ_{S} according to Table I.

much lower than the number of possible connections. Thus, the standard non-weighted metrics (i.e., accuracy) are not suitable for evaluating topology inference performance. Therefore, we adopt the detection probability (or recall), $p_{\mathrm{D}}^{i,j}$, and the false alarm probability (or false positive rate), $p_{\mathrm{FA}}^{i,j}$, of the directed link from node i to node j, defined as

$$p_{\mathrm{D}}^{i,j} = \mathbb{P}\{(\hat{a}_{i,j} = 1 | a_{i,j} = 1\}$$
$$p_{\mathrm{FA}}^{i,j} = \mathbb{P}\{\hat{a}_{i,j} = 1 | a_{i,j} = 0\}.$$

In Fig. 5, we show the detection and false alarm probabilities for some of the links of the network, using our NN-based approach, estimated from the results of Monte Carlo trials in which the position of the nodes vary inside the landscape, but the network maintain the same logical topology. Moreover, the true adjacency matrix $\mathbf{A}$, some of the adjacency matrices $\hat{\mathbf{A}}$ estimated during the Monte Carlo trials, and the matrix summarizing $p_{\mathrm{D}}^{i,j}$ and $p_{\mathrm{FA}}^{i,j}$ for each possible link of the network, are shown. In the following, p_{D} and p_{FA} are, respectively, the detection and false alarm probabilities averaged over all the network links to summarize the topology inference performance.

The algorithm used for classification is a 2-hidden-layer feed-forward NN with 40 neurons in the first hidden layer and 10 in the second one. All the layers are fully connected, and the activation functions are ReLU for the hidden layers and softmax for the output layer. The considered features are the mean, variance, kurtosis, and GCs causal magnitude. Thus, the input layer has a size of 4. The network is trained with the features extracted by the links of 70 different simulated topologies for 5000 epochs (iterations of the stochastic gradient descent algorithm) with an initial learning rate of 0.1. The learning rate decreases by a factor of 10 after 3000 epochs [50]. A k-fold cross-validation is performed to avoid overfitting with a validation set composed by the features extracted by the links of 30 different topologies. The simulated configurations used in training and test differ in the number of connections, the number of nodes, their position (i.e., network topology), and the position of the sensors. For the $\mathcal{F}-$test of GC a time lag $L = 4$ is set according to the Akaike information criterion (AIC) [44], [54], while for the CTE the parameters $R = 2$, $Q = 1$, and $n_0 = 3$, are chosen. For the decision threshold, the false alarm probability is set to 10^{-2} for both the algorithms.

The inferred topologies have to be considered instantaneous, i.e., the topology of the network in a time horizon confined by the observation time, $T_{\mathrm{ob}} = 1$ s in this case, which corresponds to $K = 100 \cdot 10^3$ samples.[6] This way, the proposed framework can capture the network's dynamical behavior, including nodes that join and leave the system.

A. BSS reconstruction error

To evaluate the performance of BSS we define the reconstruction error as

$$\mathrm{R_e} = \frac{\#\text{ of wrong samples}}{\#\text{ of total samples}} = \frac{||\mathbf{Z} - \bar{\mathbf{P}}||_1}{N \cdot K}$$

where the matrix $\bar{\mathbf{P}}$ has elements $\bar{p}_{n,k} = 1$ if node n is transmitting in the kth bin, i.e., $p_{n,k} > 0$, and 0 otherwise.[7] The accuracy of the time series reconstructed is degraded by noise and shadowing, while the node-source association could be affected by uncertainties on node positions. We model position uncertainty as a Gaussian distributed r.v. with standard deviation σ_{P} added to both the coordinates of each node. To characterize the impact of the number of sensors on the performance of the source separation methods, we define the density of sensors ρ_{S}, as the number of sensors per square meter. The nine configurations of parameters used in this test are summarized in Table I [55].

[6]The topology of the network can be obtained by collecting several instantaneous topologies and mixing all the estimations to have a topology representation on a broader time horizon, as suggested in [20].

[7]Note that topology inference exploits temporal statistics of the transmitted signals, so the quality of the reconstructed traffic profiles needs not to account for the recovered transmit power error. Hence, $\bar{\mathbf{P}}$ can be interpreted as normalization of $\mathbf{P}$ to discard irrelevant amplitude-related information.

In Fig. 6, the BSS algorithm has been tested varying the standard deviation $\sigma_{\mathrm{P}}(\%)$, defined as percentage of the side of the landscape, the shadowing parameter σ_{S}, and the density ρ_{S}. The figure depicts how $\mathrm{R_e}$ increases when σ_{P} gets higher, even at relatively high density, i.e., $\rho_{\mathrm{S}} = 0.3$. Moreover, the curves translate upward when the shadowing intensity increases, reaching an error $\mathrm{R_e} = 16\%$ with $\sigma_{\mathrm{S}} = 6\,\mathrm{dB}$ and $\sigma_{\mathrm{P}} = 20\%$. In Fig. 7 the performance of BSS is compared to the SF benchmark method described in Section IV. The figure shows how SF performance is strongly influenced by choice of the filtering parameter η. In many cases, it outperforms the BSS algorithm but requires an experimental tuning that might not always be possible. Moreover, in the presence of strong shadowing, the performance of this method rapidly degrades.

B. Topology inference and number of nodes

After BSS, the time series are processed to extract the topology information. In this section, the state-of-the-art methods for topology inference described in Section IV are compared varying the density of wireless nodes per square meter, ρ_{N}. In this test, the BSS was performed with $\rho_{\mathrm{S}} = 0.3\,\mathrm{sensors}/m^2$, $\sigma_{\mathrm{S}} = 3\,\mathrm{dB}$ and $\sigma_{\mathrm{P}} = 0$. Note that increasing the number of nodes in the landscape leads to an increase in collision probability, which results in network congestion. The NN has been trained only once on the data captured with $\rho_{\mathrm{N}} = 0.06\,\mathrm{nodes/m^2}$. A variation on the density of nodes ρ_{N} affects the topology inference when it significantly deviates from the density considered for the training. As depicted in Fig. 8, p_{D} for NN are comparable with that of GC when the density of nodes is close to the one used for training. However, when the nodes' density doubles compared to that considered for the training, GC outperforms the NN. On the contrary, when considering p_{FA} the NN is better than GC regardless of the density of nodes. As far as CTE is concerning, it presents the lower p_{FA}, but the p_{D} is lower than the other methods. Therefore, the error on the reconstruction impacts more CTE than the other approaches. Finally, considering the complexity of the algorithms, GC and NN resulted in less computationally demanding than CTE. To provide a qualitative example, for $N = 6$, inferring the complete topology requires an average execution time $t_{\mathrm{GC}} = 8.09\,\mathrm{s}$ for GC, and $t_{\mathrm{NN}} = 8.13\,\mathrm{s}$ for a NN that includes the GCs causal magnitude as feature. In the same setting, the execution of CTE requires $t_{\mathrm{CTE}} = 117.26\,\mathrm{s}$.

C. Impact of shadowing

Fig. 9 shows how increasing σ_{S} degrades the accuracy of the algorithms, as expected. In this case, we set the density of sensors $\rho_{\mathrm{S}} = 0.3\,\mathrm{sensors/m^2}$, the density of nodes $\rho_{\mathrm{N}} = 0.06\,\mathrm{nodes/m^2}$, and $\sigma_{\mathrm{P}} = 0$. Even in this scenario, CTE presents a false alarm rate lower than the other methods, but p_{D} is still the lowest. Furthermore, even if the NN outperforms the other methods for low σ_{S}, increasing the shadowing intensity results in a substantial increment of p_{FA}.

D. Impact of nodes mobility

In this test, the effect of the nodes' mobility on the performance of the topology inference is investigated. The mobility model chosen is the Random Walk [56]; within an observation window T_{ob}, each node moves along a random direction with speed v. In Fig. 10, the performance of the topology inference varying the speed of the nodes is shown. In particular, we set $v = 2, 10, 20\,\mathrm{m/s}$ to simulate human walking, a slow vehicle (i.e., low-altitude unmanned aerial vehicle (UAV)) and a fast vehicle, respectively. As highlighted by the figure, the topology inference is strongly affected by the network nodes' mobility. More specifically, in case of $v = 2\,\mathrm{m/s}$ the performance is preserved, with $v = 10\,\mathrm{m/s}$ the inference is degraded, with a detection probability reduced to 70%, while in case of $v = 20\,\mathrm{m/s}$ topology inference is compromised. This is due to the inability of BSS to reconstruct the power profile transmitted by the nodes correctly.

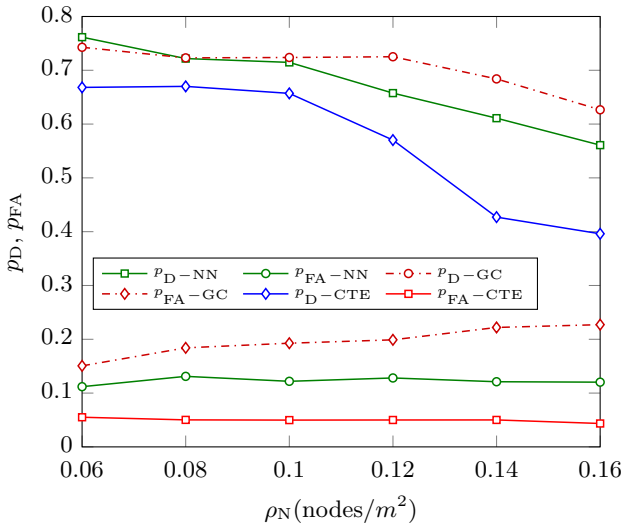

Fig. 8. p_{D} and p_{FA} of the topology inference algorithms as a function of the density of nodes ρ_{N} for $\rho_{\mathrm{S}} = 0.3$ sensors/m^2 and $\sigma_{\mathrm{S}} = 3\,\mathrm{dB}$.

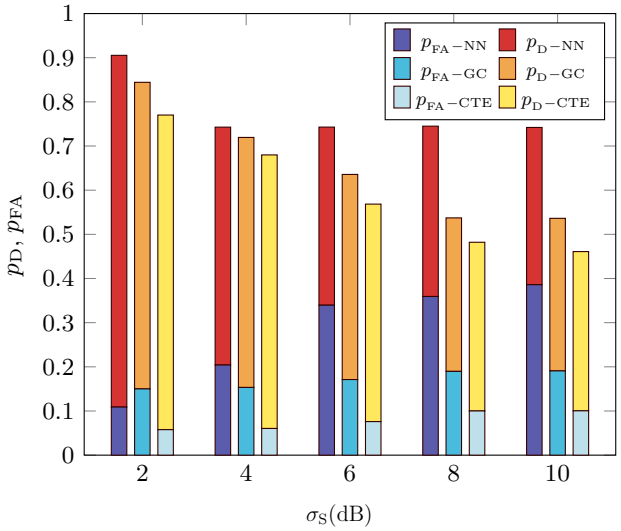

Fig. 9. p_{D} and p_{FA} of the topology inference algorithms as a function of the shadowing parameter $\sigma_{\mathrm{S}}(\mathrm{dB})$ for $\rho_{\mathrm{S}} = 0.3$ sensors/m^2.

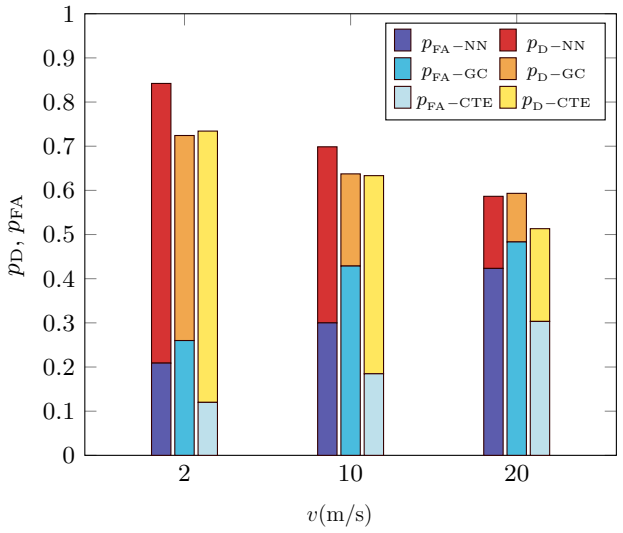

Fig. 10. $\mathrm{p_D}$ and $\mathrm{p_{FA}}$ of the topology inference algorithms in the parameter setting A_1 varying the velocity v of the nodes of the wireless network.

VI. Conclusion

We proposed a novel framework for blind topology sensing of a non-collaborative wireless network whose key features are unknown. The framework consists of combining BSS, measurement association, excision filtering, and topology inference. The last step was performed adopting state-of-the-art causal inference methods such as Granger causality (GC) and conditional transfer entropy (CTE), that exploit the times series of traffic profiles, and a novel solution based on a properly designed and trained NN that makes use of distilled time-based features. The numerical results accounting for packet collisions, nodes mobility, and realistic channel impairments, such as noise, and shadowing, revealed that, in this framework, topology inference of a wireless network is possible. Moreover, we found that in mild shadowing regime and low mobility, the performance in terms of probability of detection and probability of false alarm is remarkably good, especially for the proposed NN-based solution.

References

[1] K. Sithamparanathan and A. Giorgetti, *Cognitive Radio Techniques: Spectrum Sensing, Interference Mitigation and Localization*. Boston, USA: Artech House Publishers, Nov. 2012.

[2] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surv. Tutor.*, vol. 17, pp. 2347–2376, Jun. 2015.

[3] A. Mariani, S. Kandeepan, and A. Giorgetti, "Periodic spectrum sensing with non-continuous primary user transmissions," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1636–1649, Mar. 2015.

[4] E. Testi, E. Favarelli, and A. Giorgetti, "Machine learning for user traffic classification in wireless systems," in *Europ. Sig. Proc. Conf. (EUSIPCO)*, Rome, Italy, Sep. 2018, pp. 2040–2044.

[5] E. Favarelli, E. Testi, L. Pucci, and A. Giorgetti, "Anomaly detection using WiFi signals of opportunity," in *IEEE Int. Conf. on Signal Proc. and Comm. Syst. (ICSPCS)*, Surfers Paradise, Gold Coast, Australia, Dec. 2019.

[6] J. Wenli, G. Teng, and J. Meiyin, "Researching topology inference based on end-to-end date in wireless sensor networks," in *Int. Conf. on Int. Comp. Tech. and Autom. (ICICTA)*, vol. 2, Shenzhen, China, Apr. 2011, pp. 683–686.

[7] T. Kontos, G. S. Alyfantis, Y. Angelopoulos, and S. Hadjiefthymiades, "A topology inference algorithm for wireless sensor networks," in *IEEE Symp. on Comp. and Comm. (ISCC)*, Cappadocia, Turkey, Jul. 2012, pp. 479–484.

[8] C. Yu, K. Chen, and S. Cheng, "Cognitive radio network tomography," *IEEE Trans. on Veh. Technol.*, vol. 59, no. 4, pp. 1980–1997, May 2010.

[9] Y.Vardi, "Network tomography: estimating source-destination traffic intensities from link data," *J. of the American Stat. Ass.*, vol. 91, no. 433, pp. 365–377, Mar. 1996.

[10] Y. Gao, W. Dong, C. Chen, J. Bu, W. Wu, and X. Liu, "ipath: Path inference in wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 24, no. 1, pp. 517–528, Feb. 2016.

[11] Y. E. Sagduyu, Y. Shi, A. Fanous, and J. H. Li, "Wireless network inference and optimization: algorithm design and implementation," *IEEE Trans. Mobile Comput.*, vol. 16, no. 1, pp. 257–267, Jan. 2017.

[12] C. Partridge, D. Cousins, A. W. Jackson, R. Krishnan, T. Saxena, and W. T. Strayer, "Using signal processing to analyze wireless data traffic," in *Proc. ACM Work. on Wireless Security (WiSE)*, New York, NY, USA, Sep. 2002, pp. 67–76.

[13] S. Kokalj-Filipovic, C. Bertoncini Acosta, and M. Pepe, "Learning structural properties of wireless ad-hoc networks non-parametrically from spectral activity samples," in *IEEE Global Conf. on Signal and Info. Proc. (GlobalSIP)*, Washington, DC, USA, Dec 2016, pp. 1092–1097.

[14] G. Mateos, S. Segarra, A. Marques, and A. Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Process. Mag.*, vol. 36, pp. 16–43, May 2019.

[15] Y. Shen, B. Baingana, and G. B. Giannakis, "Nonlinear structural equation models for network topology inference," in *Annual Conference on Information Science and Systems (CISS)*, Princeton, NJ, USA, Mar. 2016, pp. 163–168.

[16] P. A. Traganitis, Y. Shen, and G. B. Giannakis, "Network topology inference via elastic net structural equation models," in *European Signal Processing Conf. (EUSIPCO)*, Kos, Greece, Aug. 2017, pp. 146–150.

[17] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[18] J. Pearl, *Causality: Models, Reasoning and Inference*, 2nd ed. New York, NY, USA: Cambridge University Press, 2009.

[19] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, Aug. 1969.

[20] P. Tilghman and D. Rosenbluth, "Inferring wireless communications links and network topology from externals using Granger causality," in *IEEE Mil. Comm. Conf. (MILCOM)*, San Diego, CA, USA, Nov. 2013, pp. 1284–1289.

[21] X. Qin and W. Lee, "Statistical causality analysis of Infosec alert data," in *Recent Advances in Intrusion Detection*, G. Vigna, C. Kruegel, and E. Jonsson, Eds. Springer Berlin Heidelberg, 2003, pp. 73–93.

[22] M. Laghate and D. Cabric, "Learning wireless networks' topologies using asymmetric Granger causality," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 233–247, Feb. 2018.

[23] H. Xu, M. Farajtabar, and H. Zha, "Learning Granger causality for Hawkes processes," in *International Conference on Machine Learning ICML*, vol. 48, New York, NY, USA, Feb. 2016.

[24] J. Etesami, N. Kiyavash, K. Zhang, and K. Singhal, "Learning network of multivariate Hawkes processes: A time series approach," in *Conf. on Uncertainty in Artif. Int. (UAI)*, Jersey City, New Jersey, USA, Mar. 2016, pp. 162–171.

[25] M. G. Moore and M. A. Davenport, "Analysis of wireless networks using Hawkes processes," in *IEEE Int. Work. on Signal Proc. Adv. in Wireless Comm. (SPAWC)*, Edinburgh, UK, July 2016, pp. 1–5.

[26] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, pp. 461–464, July 2000.

[27] P. Sharma, D. J. Bucci, S. K. Brahma, and P. K. Varshney, "Communication network topology inference via transfer entropy," *IEEE Trans. Netw. Sci. Eng.*, pp. 1–7, Jan. 2019.

[28] J. Sun, D. Taylor, and E. Bollt, "Causal network inference by optimal causation entropy," *SIAM Journal on Applied Dynamical Systems*, vol. 14, p. 73–106, Jan. 2014.

[29] S. Kokalj-Filipovic, P. Spasojevic, and A. Poylisher, "Non-parametric learning to infer wireless relays, routes and traffic patterns from time series of spectrum activity," in *Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Oct. 2017, pp. 920–927.

[30] IEEE 802.22, "Standard for Wireless Regional Area Networks—Part 22: Cognitive Wireless RAN Medium Access Control (MAC) and Physical

Layer (PHY) specifications: Policies and procedures for operation in the TV Bands," Jul. 2011.
[31] A. Mariani, A. Giorgetti, and M. Chiani, "Effects of noise power estimation on energy detection for cognitive radio applications," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3410–3420, Dec. 2011.
[32] H. Nurminen, M. Dashti, and R. Piché, "A survey on wireless transmitter localization using signal strength measurements," *Wireless Communications and Mobile Computing*, vol. 2017, pp. 1–12, Feb. 2017.
[33] K. Pahlavan, X. Li, M. Ylianttila, R. Chana, and M. Latva-aho, "An overview of wireless indoor geolocation techniques and systems," in *IFIP International Conference on Mobile and Wireless Communication Networks*, Paris, France, May 2000, pp. 1–13.
[34] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proc. IEEE*, vol. 55, no. 4, pp. 523–531, April 1967.
[35] A. Mariani, A. Giorgetti, and M. Chiani, "Wideband spectrum sensing by model order selection," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6710–6721, Dec. 2015.
[36] G. L. Stuber, *Principles of Mobile Communication*, 2nd ed. Norwell, MA, USA: Kluwer Academic Publishers, 2001.
[37] X. Yu, D. Hu, and J. Xu, *Blind Source Separation: Theory and Applications*, 1st ed. Wiley Publishing, 2014.
[38] M. Joho, H. Mathis, and R. H. Lambert, "Overdetermined blind source separation: using more sensors than source signals in a noisy mixture," in *Int. Conf. on Indep. Comp. Analysis and Blind Signal Sep. (ICA)*, Helsinki, Finland, Jun 2000, pp. 81–86.
[39] A. Cichocki, R. Zdunek, and S. Amari, "New algorithms for non-negative matrix factorization in applications to blind source separation," in *IEEE Int. Conf. on Acoustics Speech and Sig. Processing Proceedings*, vol. 5, Toulouse, France, June 2006.
[40] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, 2015.
[41] G. Ivkovic, P. Spasojevic, and I. Seskar, "Localization of packet based radio transmitters in space, time, and frequency," in *Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov. 2008.
[42] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.
[43] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, Apr. 1985.
[44] A. Mariani, A. Giorgetti, and M. Chiani, "Model order selection based on information theoretic criteria: Design of the penalty," *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2779–2789, Jun. 2015.
[45] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, Jul. 1997.
[46] E. Testi, E. Favarelli, L. Pucci, and A. Giorgetti, "Machine learning for wireless network topology inference," in *IEEE Int. Conf. on Signal Proc. and Comm. Syst. (ICSPCS)*, Surfers Paradise, Gold Coast, Australia, Dec. 2019.
[47] A. K. Seth, "A Matlab toolbox for Granger causal connectivity analysis," *J. of Neuroscience Methods*, vol. 186, no. 2, pp. 262 – 273, Feb. 2010.
[48] M. Wibral, N. Pampu, V. Priesemann, F. Siebenhühner, H. Seiwert, M. Lindner, J. T. Lizier, and R. Vicente, "Measuring information-transfer delays," *PLOS ONE*, vol. 8, pp. 1–19, Feb. 2013.
[49] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer Verlag, Aug. 2006.
[50] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2016.
[51] J. Watt, R. Borhani, and A. K. Katsaggelos, *Machine Learning Refined*. Cambridge University Press, 2016.
[52] J. T. Lizier, "Jidt: An information-theoretic toolkit for studying the dynamics of complex systems," *Frontiers in Robotics and AI*, vol. 1, pp. 1–11, Dec. 2014.
[53] M. Lacage and T. R. Henderson, "Yet another network simulator," in *Proc. Work. on Ns-2: the IP netw. sim. (WNS2)*, Pisa, Italy, Oct. 2006, p. 12–22.
[54] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Proc. Mag.*, vol. 21, no. 4, pp. 36–47, July 2004.
[55] P. Stuedi, O. Chinellato, and G. Alonso, "Connectivity in the presence of shadowing in 802.11 ad hoc networks," in *Proc. IEEE Wireless Comm. and Netw. Conf.*, vol. 4, New Orleans, LA, USA, Mar. 2005, pp. 2225–2230.
[56] C. Bettstetter, "Mobility modeling in wireless networks: categorization, smooth movement, and border effects," *ACM Mobile Computing and Communications Review*, vol. 5, no. 3, p. 55–66, Jul. 2001.