



Università degli Studi di Salerno

.DIEM

**Dipartimento di Ingegneria dell'Informazione
ed Elettrica e Matematica Applicata**

Dottorato di Ricerca in Ingegneria dell'Informazione
Ciclo 35

TESI DI DOTTORATO / PH.D. THESIS

Stima di Grafi a Connessione Preferenziale nell'Ipotesi di Osservabilità Parziale

MICHELE CIRILLO

SUPERVISOR: **PROF. VINCENZO MATTA**

PHD PROGRAM DIRECTOR: **PROF. PASQUALE CHIACCHIO**

Anno Accademico 2022/2023

Abstract

Questa tesi analizza il problema della stima della topologia di una rete a partire dai segnali prodotti dai suoi nodi durante l'esecuzione di un algoritmo distribuito. In particolare, i suddetti segnali sono generati nel tempo secondo un modello autoregressivo vettoriale, ossia un modello diffusivo lineare in cui ogni nodo scambia messaggi con i suoi vicini (dunque, in base al grafo che descrive la rete) e li combina seguendo una certa regola. Si considera il complesso scenario in cui la stima della topologia debba avvenire sotto l'ipotesi di osservabilità parziale, secondo la quale solo parte dei nodi può essere osservata, e vengono studiate le condizioni sotto le quali il grafo relativo ai nodi osservati può essere ricostruito; si tratta di un problema complesso, in quanto i segnali osservati sono influenzati dalla presenza di nodi latenti, i cui segnali fungono da rumore e possono in principio inficiare la ricostruzione del grafo. Si considerano due aspetti fondamentali del problema. Il primo aspetto riguarda la fattibilità: sotto quali condizioni il problema di stima della topologia è risolvibile? Ossia, esiste uno stimatore in grado di ricostruire il grafo in esame a partire dai segnali emessi dai nodi osservati? In generale, una risposta affermativa a queste domande permette soltanto di sapere che uno stimatore esiste, e che funzionerà con un numero sufficientemente grande di campioni. Di conseguenza, il secondo aspetto analizzato riguarda l'onerosità della soluzione: dato uno stimatore individuato dalla suddetta analisi di fattibilità, quanti campioni richiede per funzionare correttamente in pratica? Studi recenti hanno analizzato il problema quando il grafo in esame è generato secondo il modello aleatorio di Erdős-Rényi. Tuttavia, il modello di Erdős-Rényi utilizza un meccanismo di costruzione semplicistico che produce (non sempre realistiche) reti omogenee con collegamenti indipendenti gli uni dagli altri. In questo lavoro si superano tali limitazioni considerando i cosiddetti grafi a connessione preferenziale (*preferential attachment*) che sono caratterizzati da una significativa eterogeneità, aventi nodi con molte connessioni (che corrispondono agli "hub" nelle reti reali) e nodi periferici con poche connessioni. Inoltre, il modello a connessione preferenziale impone un'elevata dipendenza probabilistica tra i collegamenti del grafo. Queste sono proprietà importanti, che ritroviamo spesso nelle reti reali. Il principale contributo del lavoro esamina il caso in cui un processo autoregressivo vettoriale del primo ordine, governato da una matrice di combinazione laplaciana stabile, è eseguito su un grafo generato secondo il popolare modello preferential attachment di Bollobás-Riordan. Si intro-

duce innanzitutto un framework unificante per descrivere il problema della stima del grafo sotto l'ipotesi di osservabilità parziale, che comprende sia i precedenti risultati su grafi di Erdős-Rényi sia i nuovi risultati riguardanti il modello di Bollobás-Riordan. Il risultato principale dimostrato in questa tesi afferma che esiste uno stimatore della matrice di combinazione, noto come stimatore di Granger, che rende possibile la stima del grafo nell'ipotesi di osservabilità parziale. Si caratterizza poi la complessità di questo stimatore su grafi di Bollobás-Riordan, dimostrando che esso funziona correttamente con un numero di campioni che cresce con la dimensione della rete seguendo una legge essenzialmente lineare. Confrontando questo risultato con quelli già noti per i grafi di Erdős-Rényi si osserva che: i) la stima di grafi di Erdős-Rényi densi è la più onerosa, e il numero di campioni necessari cresce quadraticamente con la dimensione della rete; ii) la stima di grafi di Bollobás-Riordan ha una complessità inferiore, che si quantifica in una crescita essenzialmente lineare del numero di campioni; iii) infine, la stima di grafi di Erdős-Rényi sparsi richiede una crescita sublineare.



Università degli Studi di Salerno

.DIEM

**Dipartimento di Ingegneria dell'Informazione
ed Elettrica e Matematica Applicata**

Dottorato di Ricerca in Ingegneria dell'Informazione
Ciclo 35

TESI DI DOTTORATO / PH.D. THESIS

Learning Preferential Attachment Graphs Under Partial Observability

MICHELE CIRILLO

SUPERVISOR: **PROF. VINCENZO MATTA**

PHD PROGRAM DIRECTOR: **PROF. PASQUALE CHIACCHIO**

Anno Accademico 2022/2023

Abstract

This thesis work deals with the problem of learning the topology of a network starting from the signals emitted by the network nodes while executing some distributed processing task. In particular, these signals are generated over time through a vector autoregressive process, which is a linear diffusion process where each node exchanges messages with its neighbors (therefore, according to the underlying network graph) and aggregate them according to a certain combination policy. We consider the demanding setting of graph learning under partial observability, where only part of the nodes can be probed, and we study under which conditions the subgraph relative to the probed nodes can be correctly estimated. This is a challenging problem since the observed signals are also influenced by the presence of latent nodes, whose signals act as noise and can in principle prevent faithful graph reconstruction. In particular, we consider two fundamental questions. The first question is about achievability: Under which conditions the graph learning problem can be solved? Namely, for meaningful classes of graphs, is there a graph estimator which, starting from signals of the probed nodes, is able to recover the related subgraph? Usually, a positive answer for achievability only says that one or more estimators exist, and that it works for a sufficiently large number of samples. Therefore, the second question arises in the context of sample complexity: Given a graph estimator found by the achievability analysis, how many samples it needs to work properly in practice? Recent results in the literature examined this problem when the underlying graph is generated according to an Erdős-Rényi random model. The main limitation of this assumption is that Erdős-Rényi graphs use a simple construction mechanism that produces (sometimes unrealistic) homogeneous networks with independent edges. We overcome this issue by solving the problem of graph learning over preferential attachment graphs, which are characterized by a large heterogeneity, featuring both very connected nodes (which model real-world network “hubs”) and peripheral nodes having few connections. Moreover, preferential-attachment graphs enforce a strong dependence between the edges of the graph. These are important properties that can be observed in real-world networks. In particular, our main contribution examines the case where a first-order vector autoregressive process, equipped with a stable Laplacian combination matrix, is run over the a graph drawn according to the popular Bollobás-Riordan preferential attachment model. In this thesis we first introduce a unifying framework for graph

learning under partial observability. This framework covers in particular the previous results on Erdős-Rényi graphs and our novel results on Bollobás-Riordan graphs. The main achievability result established in the present thesis is that a combination matrix estimator known as Granger estimator achieves graph learning under partial observability. We also characterize the sample complexity over Bollobás-Riordan graphs, establishing that it is essentially linear in the network size. Comparing this result with what was observed before for Erdős-Rényi graphs, we obtain the following interesting classification: i) dense Erdős-Rényi graphs require the highest sample complexity rate, where the number of samples scales quadratically as the network size grows; ii) the intermediate sample complexity rate is given by Bollobás-Riordan graphs, which require an almost-linear sample scaling law; finally, iii) sparse Erdős-Rényi graphs have a lighter, sublinear sample scaling law.