

Università degli studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche



TESI DI LAUREA

**CONTRASTARE LE RETI DI TERRORISTI.
UN MODELLO DI CLASSIFICAZIONE PER RETI**

Relatore Prof. Bruno Scarpa
Dipartimento di Scienze Statistiche

Laureando Emanuele Aliverti
Matricola N 1104081

Anno Accademico 2015/2016

Indice

Introduzione	5
1 Introduzione al problema	7
1.1 Reti sociali e terrorismo	7
1.2 Cenni di teoria dei grafi	9
1.3 La rete di Noordin Mohammad Top	10
1.3.1 Analisi descrittive	12
2 Modelli per reti	15
2.1 Motivazioni	15
2.2 Contesto	16
2.3 Approccio	17
2.4 Modello ad effetti fissi	19
2.4.1 Distribuzioni a priori	22
2.4.2 Distribuzione a posteriori	23
2.4.3 Gibbs sampling per il modello ad effetti fissi	24
2.5 Modello ad effetti casuali	26
2.5.1 Gibbs sampling per il modello ad effetti casuali	28
3 Simulazione	31
3.1 Scenari di simulazione	31
3.2 Modelli competitori	33
3.3 Applicazione ai dati simulati	34
3.4 Risultati	35

4	Applicazione ai dati reali	39
4.1	Simulazione dalla distribuzione a posteriori	39
4.1.1	Modello ad effetti fissi: risultati	40
4.1.2	Modello ad effetti casuali: risultati	43
4.2	Confronti	47
4.3	Riflessioni finali	50
A	Materiale aggiuntivo	53
A.1	Costruzione della matrice dei coefficienti casuali	53
A.2	Rappresentazione grafica delle reti coinvolte	53
	Bibliografia	55

Introduzione

Dopo l'attentato dell'11 settembre 2001, il mondo occidentale ha realizzato come le minacce terroristiche, fino a quel momento percepite come isolate e poco pertinenti, non fossero relegate solo a Paesi lontani; la risposta statunitense e, per antonomasia, occidentale a questo pericolo è giunta attraverso la dichiarazione di guerra a nazioni nelle quali si supponeva che i gruppi di terroristi avessero base, come ad esempio Afghanistan e Iraq.

Negli anni seguenti, tuttavia, i fenomeni terroristici non sono scomparsi; l'Europa stessa ha dovuto, molto recentemente, scontrarsi con questi incubi da tempo assopiti. Se in un primo momento l'obiettivo dei gruppi estremisti era la dimostrazione della propria potenza, raggiunta colpendo i centri del potere occidentale, oggi invece gli attentati vengono compiuti in modo tale da garantire un'intensa visibilità mediatica, per ricordare al mondo moderno la propria capacità di minare i valori democratici sui quali è stato costruito (Bauman, 2016). La situazione odierna conferma sia la scarsa efficacia che la poca lungimiranza delle strategie belliche adottate; bilanci positivi, purtroppo, si registrano solo nel numero di civili coinvolti durante le operazioni militari e nella contabilità dei venditori e trafficanti di armi.

Per questo motivo, sono stati proposti approcci differenti per contrastare le azioni terroristiche tramite operazioni preventive; sfruttando le capacità dei servizi di *intelligence* di ottenere informazioni dettagliate, la disgregazione dei gruppi di terroristi diventa un obiettivo decisamente verosimile (Everton, 2005). Ad esempio, bloccando le strutture di supporto logistico e finanziario in modo opportuno si potrebbero prevenire le operazioni più violente, senza dover ricorrere a misure offensive su larga scala come quelle utilizzate fino ad oggi.

L'utilizzo di strumenti statistici a favore di questo approccio risulta naturale:

è indispensabile utilizzare tecniche in grado di convertire i dati grezzi, raccolti dai servizi governativi, in informazione leggibile ed interpretabile; i risultati ottenuti potranno aiutare gli esperti nella programmazione delle operazioni in modo coerente, permettendo di scegliere la strategia migliore.

Nel primo capitolo di questa tesi, il lavoro verrà contestualizzato entro il panorama letterario sull'argomento, definendo quindi i dati che è intenzione analizzare, vero punto di partenza di ogni analisi.

Successivamente, verranno presentati gli obiettivi di ricerca e proposto un modello statistico potenzialmente in grado di rispondere agli interrogativi proposti; la formulazione sarà descritta in dettaglio nel capitolo 2.

Nel capitolo 3, la correttezza delle proposte sarà valutata tramite un'applicazione esaustiva su dati artificiali, confrontandone la *performance* con strumenti classici dell'analisi statistica.

Ultima, ma certamente non per importanza, l'applicazione ai dati disponibili e l'illustrazione critica dei risultati ottenuti, effettuata nel capitolo 4.

Capitolo 1

Introduzione al problema

1.1 Reti sociali e terrorismo

L'approccio delle reti sociali per l'analisi dei fenomeni militari vede le proprie origini durante la seconda guerra mondiale, attraverso una branca della criptologia denominata "*analisi dei traffici*" (Meter, 2002). Nello specifico, le analisi condotte da parte dei servizi segreti erano orientate principalmente verso la raccolta di informazioni riguardanti la comunicazione tra obiettivi bellici sensibili, in modo tale da individuare, ad esempio, quali soggetti maggiormente comunicassero tra loro, con quale frequenza e modalità; tramite queste tecniche i militari tentavano di individuare quali fossero i soggetti più importanti e potenzialmente pericolosi, concentrandosi sulla *forma* della comunicazione (presente/assente, diretta/indiretta) e non sul contenuto della stessa. Degni di nota anche metodi di indagine sul campo, nati durante i primi anni '60, che consistevano in etnografie antropologiche volte a descrivere in dettaglio le relazioni tra soggetti potenzialmente pericolosi sotto forma di reti sociali (McGehee, 1983); questo tipo di tecniche per la raccolta dati, sviluppata in ambito militare, fu cruciale durante le prime fasi della più generale analisi di reti sociali moderna (SNA): si pensi ad esempio alla rete di monaci di Sampson (1968), uno degli esempi di rete sociale più noti in letteratura, costruita unicamente tramite interviste ai soggetti interessati.

Questa particolare tipologia di dato viene rappresentata come delle *reti sociali*, ossia degli insiemi di nodi connessi tra di loro; ogni soggetto rappresenta quindi un

nodo, e una connessione tra due nodi indica la presenza di una particolare relazione. In alcuni casi sono disponibili informazioni riguardo *diversi* tipi di relazioni tra i medesimi soggetti, ed è quindi possibile definire una rete *multi-layer*; per una definizione di grafo più rigorosa, si rimanda alla sezione 1.2.

A partire dall'attentato dell'11 settembre, tecniche afferenti alla SNA iniziarono ad essere utilizzate specificatamente nei contesti di analisi di reti terroristico-criminali, denominate "*dark network*"; analisi specifiche di questo filone sono, ad esempio, Krebs (2002) e Rodríguez e Rodríguez (2005). Negli anni successivi, il potenziale esplicativo di questo tipo di analisi ha implicato che l'attenzione degli esperti si concentrasse principalmente nello sviluppo di nuove tecniche di spionaggio per raccogliere dati affidabili, i quali molto spesso venivano presentati sotto forma di reti sociali (Chen, 2012).

Un interessante contributo in questo filone letterario è presente in Roberts e Everton (2011), i quali propongono di utilizzare tecniche di analisi di reti sociali per coordinare in modo più efficiente le operazioni anti-terrorismo, orientando le strategie militari in modo più preciso.

La lotta alle *dark networks* è stata principalmente condotta tramite tecniche "cinetiche", ossia azioni militari dirette contro obiettivi sensibili, con obiettivo l'eliminazione o la cattura degli stessi. Tuttavia, durante gli ultimi anni, si è notato come queste associazioni criminali spesso siano molto resistenti anche verso attacchi mirati ai soggetti più influenti; i membri rimasti sono sovente in grado di rimpiazzare in tempo brevissimo i leader eliminati (Pedahzur e Perliger, 2006). Paradossalmente, questo tipo di operazioni registrano talvolta veri e propri effetti collaterali: l'eliminazione di soggetti centrali rischia infatti di consentire a fazioni marginali di assumere posizioni di monopolio, oppure di rendere la rete più decentralizzata e quindi più complicata da tenere sotto controllo (Flynn et al., 2010). Non trascurabili, inoltre, il rischio di un coinvolgimento di civili durante queste operazioni, e le conseguenze sociali di un aumento sconsiderato di violenza in contesti gravemente colpiti da minacce terroristiche.

Diviene centrale dunque lo sviluppo di tecniche "non-cinetiche", i.e. meno violente e più programmate rispetto all'approccio classico, in grado di sfruttare a pieno le informazioni provenienti dai servizi di *intelligence*. Tra le strategie proposte ve ne sono alcune, chiamate *Computer Network Operations*, che si focalizzano su metodi per eliminare indirettamente e preventivamente le minacce terroristiche,

bloccando ad esempio la struttura di supporto finanziario e logistico degli attori coinvolti, oppure individuando con precisione il ruolo particolare dei diversi soggetti all'interno delle organizzazioni (Roberts e Everton, 2011). Le indagini condotte in questo ambito consistono sostanzialmente in analisi descrittive, più o meno articolate, delle reti sociali in oggetto; gli studiosi in questo settore mirano infatti a sintesi delle proprietà topologiche della rete oppure a descrizioni delle proprietà degli individui in quanto membri di una rete sociale per individuare opportunamente più influenti (Everton, 2005).

1.2 Cenni di teoria dei grafi

Introduciamo ora alcune definizioni di base di teoria dei grafi, necessarie per poter descrivere opportunamente le variabili disponibili.

Definizione 1 (Grafo non orientato) *Un grafo non orientato $G = (V, E)$ è una struttura matematica composta da un insieme V di nodi e un insieme E di archi, i cui elementi sono coppie non ordinate $\{u, v\}$ di nodi $u, v \in V$.*

Un grafo semplice è un grafo non orientato, in cui agli archi non è associato alcun peso, e che non contiene né archi multipli (archi che compaiono ripetutamente) né “self-loops” (archi che collegano un nodo a sé stesso)

Definizione 2 (Grafo multilivello) *Un grafo indiretto multilivello è una collezione di M grafi indiretti $G^{(i)} = \{V, E^{(i)}\}$, $i = 1 \dots M$. Ogni livello consiste di un insieme di nodi condiviso e un insieme di archi specifico.*

La connettività di un grafo può essere agevolmente rappresentata da una matrice di adiacenza.

Definizione 3 (Matrice di adiacenza) *Una matrice di adiacenza \mathbf{A} relativa ad un grafo semplice è una matrice quadrata di dimensione $v \times v$, con $v = \text{card}(V)$, i cui elementi*

$$a_{ij} = \begin{cases} 1, & \text{se } \{i, j\} \in E \\ 0 & \text{altrimenti} \end{cases}$$

Se il grafo è semplice, inoltre, non vi è alcuna direzionalità nelle connessioni, e quindi un arco tra il nodo i e il nodo j può essere letto anche come un arco

tra j e i ; di conseguenza, $a_{ji} = a_{ij} \forall i, j$, il che implica simmetria per la matrice di adiacenza. Notiamo inoltre che, vista l'assenza di *self-loops*, gli elementi sulla diagonale della matrice di adiacenza sono nulli per definizione.

All'interno di un grafo, per misurare la distanza tra due nodi è sensato utilizzare la *distanza geodetica*, ossia il numero di archi necessari per connetterli; due nodi possono essere connessi attraverso diversi percorsi, quindi si utilizza come riferimento la lunghezza del cammino più breve (detto *shortest-path*).

In un grafo, infine, è possibile definire diverse proprietà, le quali saranno *locali* se riferite ai singoli nodi oppure *globali* se riferite al grafo nel suo complesso. Le proprietà necessarie alla comprensione della trattazione successiva sono le seguenti.

Definizione 4 (Grado, Densità, Transitività) *Si definiscono*

- *Grado di un nodo il numero totale di archi incidenti al nodo stesso.*
- *Densità di un grafo la frequenza relativa di archi realizzati rispetto al numero totale di archi possibili.*
- *Transitività di un grafo il rapporto tra il numero di triangoli (gruppi di tre nodi tra i quali sono presenti tre connessioni) e il numero totale di triple connesse (gruppi di tre nodi tra i quali sono presenti solo due connessioni).*

1.3 La rete di Noordin Mohammad Top

I dati che si intende esaminare contengono informazioni riguardo la composizione e struttura della rete terroristica avente come leader Noordin Mohammad Top (11 Agosto 1968 – 17 Settembre 2009). Fino al momento della morte, il terrorista era il criminale più ricercato da parte del governo indonesiano; ritenuto la mente, o comunque attivo partecipante, degli attentati più sanguinosi che hanno colpito l'Asia dal 2000 in poi, riuscì a sfuggire ai tentativi di cattura da parte della polizia per diversi anni. Tra le principali accuse a suo carico, si citano l'attacco al *JW Marriott hotel*, all'Ambasciata australiana di Jakarta (2003 e 2004), l'attentato di Bali del 2005 e il sanguinoso attacco del 2009 agli hotel *JW Marriott* e *Ritz-Carlton* di Jakarta. Si suppone inoltre che il criminale abbia avuto un ruolo attivo anche negli attentati di Bali del 2002. Nel 2006, quando l'ipotesi di un'internazionalizzazione

delle operazioni terroristiche divenne verosimile, Noordin venne inserito nelle liste dei ricercati internazionali dell’FBI; su di lui fu fissata una taglia pari ad un miliardo di *rupiah* indonesiane (Wikipedia, 2004).

Durante i primi anni di attività Noordin e i suoi seguaci appartenevano alla *Jemaah Islamiyah (JI)*, nota organizzazione militante alleata con *al-Qaeda*, avente come obiettivo l’insediamento del califfato Islamico nel sud-est asiatico; a causa dell’atteggiamento violento e spietato del gruppo vicino a Noordin, nacquerò attriti con le alte gerarchie della JI, e quindi in un secondo momento fondarono una fazione estremista a sé stante (ICG, 2006). Le autorità indonesiane lo definirono dotato di notevole capacità nella progettazione e gestione degli attacchi dinamitardi: nei casi di arresto di alcuni collaboratori, si dimostrò in grado di reclutarne in brevissimo tempo di nuovi. Degna di nota anche l’abilità del terrorista nel recuperare fondi, tramite occupazioni di copertura o rapine, nel trovare assistenza logistica per le missioni e nel persuadere giovani studenti a convertirsi ad un Islam estremo.

I dati utilizzati sono tratti da Roberts e Everton (2011), i quali li hanno codificati tramite rapporti ufficiali dell’*International Crisis Group’s*; le informazioni disponibili si basano principalmente su interrogazioni, documenti e operazioni di *intelligence* del governo indonesiano. Per quanto riguarda i dettagli sulla costruzione completa del dataset e delle informazioni sulla raccolta dei dati, si rimanda a Everton (2005).

Nello specifico, le informazioni riguardanti l’organizzazione terroristica che si intende analizzare sono state codificate sotto forma di 13 reti sociali indirette, una per ogni particolare rapporto (simmetrico) analizzato; si tratta quindi di un grafo multi-livello. Per ogni diversa relazione, i nodi rappresentano Noordin e i 78 terroristi in qualche modo a lui legati, mentre un arco tra due nodi indica la presenza di quella particolare relazione tra i due soggetti.

I rapporti disponibili sono i seguenti:

- Precedente appartenenza alla medesima organizzazione estremista
- Comunicazione frequente tramite telefono o internet
- Compagni di classe
- Compagni di scuola

- Legame di parentela
- Partecipazione alla medesima attività di addestramento militare
- Amicizia
- Lavoro nella medesima attività commerciale
- Appartenenza alla stessa Moschea o Sinagoga
- Affiliazione alla stessa istituzione religiosa nel medesimo periodo
- Ritrovo nel medesimo rifugio
- Partecipazione alla medesima operazione terroristica
- Partecipazione al medesimo incontro programmato
- Fornitura comune di materiale bellico di supporto

Ad esempio, nella matrice di adiacenza riferita alla variabile "Partecipazione al medesimo incontro programmato", se l'elemento $a_{ij} = 1 = a_{ji}$, il nodo i e il nodo j hanno partecipato insieme ad un ritrovo organizzato.

1.3.1 Analisi descrittive

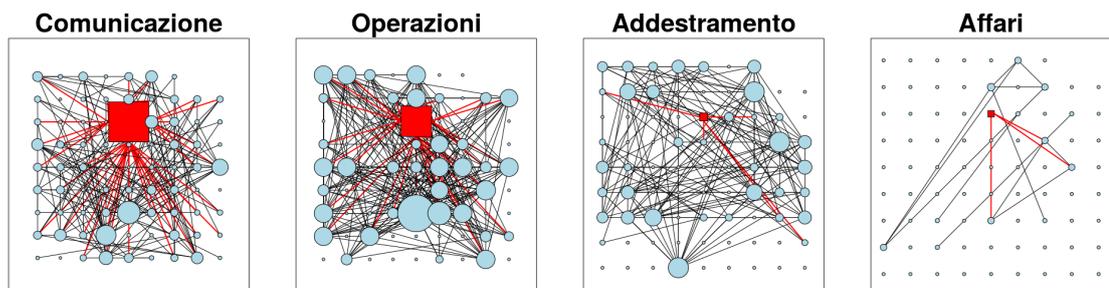


Figura 1.1: Rappresentazione grafica delle reti indicate. La dimensione di ogni nodo è proporzionale al grado; in evidenza il nodo e gli archi relativi al leader Noordin Mohamad Top.

Nella fig. 1.1 si può notare, graficamente, l'eterogeneità delle diverse reti. Le quattro rappresentazioni condividono la medesima posizione per tutti i nodi, così

da migliorarne la lettura. In alcune relazioni gli archi sono molto fitti, mentre in altri ve ne sono molti meno; gli stessi nodi, inoltre, formano molte connessioni in alcuni tipi di rapporto, mentre in altri relativamente pochi se non addirittura nessuno. Il leader Noordin, ad esempio, nonostante sia chiaramente un soggetto centrale nella rete relativa alla comunicazione, è praticamente periferico in quella riguardante l'addestramento. Per una valutazione grafica di tutte le reti disponibili, si rimanda alla fig. [A.1](#).

Tabella 1.1: Statistiche descrittive delle reti disponibili. Densità, transitività, media e deviazione standard della distribuzione dei gradi.

Rete	Statistiche di rete		Grado	
	Densità	Transitività	Media	Dev. std
Organizzazione	0.135	0.841	10.532	8.829
Comunicazione	0.065	0.265	5.063	5.774
Compagni classe	0.057	0.668	4.430	6.398
Compagni scuola	0.092	0.827	7.190	9.472
Parentela	0.005	0.667	0.405	0.689
Addestramento	0.048	0.715	3.722	4.849
Amicizia	0.030	0.204	2.304	2.604
Lavoro	0.005	0.889	0.380	0.965
Moschea o Sinagoga	0.004	0.882	0.304	0.806
Affiliazione religiosa	0.004	0.750	0.278	0.846
Rifugio	0.027	0.612	2.076	3.869
Operazioni	0.087	0.785	6.759	8.368
Incontro programmato	0.020	0.410	1.595	3.209
Fornitura Materiale	0.149	0.747	11.595	11.985

Si riportano nella tabella [1.1](#) alcune statistiche descrittive delle reti disponibili. Le reti analizzate appaiono abbastanza sparse, ossia composte da un numero piccolo di archi rispetto a tutti quelli che sarebbe possibile formare; la densità non è infatti mai superiore a 0.15. Questa caratteristica è molto diffusa in contesti di *Dark Network*, come testimoniato ad esempio da [Krebs \(2002\)](#). Alcuni autori sostengono che da un punto di vista sociale una bassa densità consenta comportamenti meno omogenei ([Granovetter, 2005](#)), e soprattutto una semplicità e velocità

maggiore nella sostituzione di membri. Nonostante il basso numero di connessioni, si registra una transittività importante: questa misura definisce la propensione di una rete a formare situazioni del tipo “un amico di un amico è anch’esso un amico”. Ad esempio, per come è definita la variabile "Lavoro", se il nodo i ha lavorato nella stessa attività di j e di k , allora è verosimile che anche il nodo j e il nodo k abbiano lavorato insieme. La distribuzione del grado medio di ogni rete è molto eterogenea, variando tra un massimo di 10 ad un minimo di soli 0.2 archi medi per nodo. Si nota inoltre come nelle reti con un grado medio molto elevato, anche la deviazione standard lo sia: alcuni soggetti, generalmente pochi, formano molti legami, mentre la maggior parte dei soggetti ne forma un numero modesto.

Capitolo 2

Modelli per reti

2.1 Motivazioni

Gran parte delle informazioni sulle *dark network* è ottenibile in modo agevole da parte dei servizi governativi, come ad esempio le informazioni biografiche riguardanti il passato e le relazioni legali tra individui (scuola frequentata, parentela, residenza); grazie alle potenzialità dei servizi di *intelligence*, è possibile ottenere in modo dettagliato anche informazioni sulla comunicazione via mezzi mediatici tra soggetti, quali cellulari, internet e mezzi di stampa. Altri tipi di relazione, invece, sono decisamente più "oscure", poiché riguardanti attività illecite o comunque difficilmente misurabili; fanno parte di questa seconda categoria, ad esempio, le dinamiche di scambio illegale di denaro tra individui, la collaborazione in operazioni criminali o la presenza di una relazione di amicizia o fiducia tra gli stessi. Ciò nonostante, è lecito aspettarsi che esista una relazione tra i diversi rapporti: ad esempio, la presenza di un legame di parentela tra due individui potrebbe rendere più verosimile uno scambio di denaro tra gli stessi, oppure il fatto che entrambi fossero compagni di scuola in passato rende più plausibile che tra gli stessi possa esistere anche una relazione di fiducia.

Coerentemente con queste riflessioni, in questo lavoro si è deciso di collocare il problema in un ambito di analisi supervisionata, considerando una rete particolare come *risposta* e le altre come *esplicative*; gli obiettivi principali saranno quelli di presentare strumenti in grado di fare inferenza sulla relazione esistente tra le reti

esplicative e la rete risposta, eventualmente quantificandone l'incertezza in modo opportuno, e fornire previsioni per osservazioni future.

Dopo un'attenta riflessione riguardo le reti disponibili, si è deciso che la rete più sensata ed interessante da utilizzare come risposta sia quella riferita alla co-partecipazione tra due terroristi nello stesso attentato. In ambito di analisi del terrorismo, questa particolare rete è cruciale per diversi motivi. Dal punto di vista della raccolta dati, fino al compimento degli attentati non è possibile avere informazioni oggettive a riguardo; allo stesso tempo, chiaramente, si vuole evitare che la tragedia si compia. In aggiunta, una stima del legame tra la co-partecipazione ad un'operazione terroristica e i diversi rapporti tra individui potrebbe fornire agli esperti una descrizione più chiara del fenomeno; in questo modo, ad esempio, i servizi di *intelligence* potrebbero ottimizzare le risorse disponibili per focalizzarsi unicamente sulle relazioni realmente utili per la previsione di un attacco terroristico, migliorando l'efficienza della raccolta dati ed attuando di conseguenza una prevenzione precoce. Si rimanda alla sezione 1.1 per riferimenti tecnici riguardanti la plausibilità di questo tipo di approccio.

Infine, un'applicazione desiderabile è quella di sfruttare l'informazione attualmente disponibile per fare delle previsioni sul comportamento di soggetti indagati in futuro: una volta codificate le informazioni esplicative dei nuovi terroristi indagati, ossia in questo caso specifico le relazioni degli stessi con i criminali attualmente presenti nel *dataset*, è di interesse stabilire se il nuovo soggetto potrebbe collaborare in un'operazione terroristica e con quali complici.

2.2 Contesto

Sia $v = |V|$ il numero di nodi nelle reti in oggetto. Siano quindi

$$\mathbf{Y} \in (0, 1)^{v \times v} \quad \mathbf{X} \in (0, 1)^{v \times v \times p}$$

rispettivamente, la matrice di adiacenza relative alla rete risposta e il tensore (*array*) composto dalle matrici di adiacenza delle p reti esplicative; le dimensioni di questa struttura saranno chiamate, rispettivamente, *righe*, *colonne*, *livelli*. Data la non direzionalità delle reti associate, le matrici sono simmetriche e con valori nulli sulla diagonale; per rappresentarle sono quindi sufficienti i $\frac{v(v-1)}{2}$ valori delle matrici triangolari inferiori. Le entità di base coinvolte in questo scenario sono

matrici, quindi durante la trattazione potrà essere necessaria l'introduzione di alcune operazioni particolari.

Gli obiettivi d'indagine e la natura dicotomica della matrice che codifica la rete risposta contestualizzano questo lavoro entro l'ambito dei problemi di *classificazione* binaria; per un'introduzione a questo tipo fondamentale di analisi statistica e ad alcune tecniche per affrontarla, si rimanda a [Azzalini e Scarpa \(2012, cap. 5\)](#).

L'ipotesi di indipendenza delle osservazioni alla base della maggior parte dei modelli statistici classici è decisamente poco verosimile in ambito di dati di rete. In questo tipo particolare di strutture, infatti, sono presenti diverse forme di dipendenza tra gli archi che, se non controllate in modo opportuno, possono fornire risultati fuorvianti ([Li e Loken, 2002](#)); ad esempio, elementi nella stessa colonna della matrice di adiacenza rappresentano archi collegati al medesimo nodo, ed è quindi lecito attendersi che non siano tra loro indipendenti. Nelle analisi descrittive della sezione [1.3.1](#), inoltre, è emerso chiaramente come i diversi attori coinvolti abbiano comportamenti molto diversi tra loro, che sarebbe riduttivo considerare identici.

2.3 Approccio

Gli approcci più recenti e di maggior successo a questo tipo di problemi considerano gli archi come variabili bernoulliane *condizionatamente* indipendenti; generalmente, si suppone l'esistenza di una qualche struttura latente, condizionandosi alla quale è possibile considerare indipendenti gli archi. Esempi noti sono modelli a distanze latenti ([Hoff et al., 2001](#)) e generalizzazioni ([Krivitsky et al., 2007](#)), i modelli a blocchi stocastici ad appartenenza singola ([Nowicki e Snijders, 2001](#)) e multipla ([Airoldi et al., 2008](#)) e il modello di mistura di fattorizzazioni di rango ridotto ([Durante et al., 2014](#)).

In ambito di modelli per reti a strutture latenti, le stime di massima verosimiglianza per i parametri coinvolti spesso risultano difficilmente ottenibili ed instabili ([Hoff, 2003](#)). A causa delle complicate strutture di dipendenza che questi modelli tentano di spiegare, i fattori latenti spesso sono specificati tramite formulazioni non banali, le quali causano instabilità numerica in fase di ottimizzazione; dal punto di vista dell'identificabilità del modello, inoltre, sono richiesti spesso vincoli stringenti che mettono in dubbio la coerenza degli stessi. In [Snijders e Nowicki \(1997\)](#),

ad esempio, gli autori illustrano come la massimizzazione diretta della verosimiglianza di un modello a blocchi stocastici sia praticamente impossibile, e anche ricorrendo ad un algoritmo EM non si riesca ad ottenere trattabilità computazionale se non con un numero ridotto di nodi. Un approccio bayesiano ai problemi in questione rappresenta di fatto la strada percorsa più frequentemente, permettendo di ottenere trattabilità computazionale e buoni risultati dal punto di vista interpretativo.

In questa sede non si intende eleggere l'approccio bayesiano (piuttosto che quello frequentista) come panacea ad ogni problema, e nemmeno entrare nel dibattito filosofico-epistemologico che coinvolge queste due culture. Il punto di vista di chi scrive è che la statistica sia il fondamento del metodo scientifico, e che quindi uno dei suoi obiettivi principali resti quello di estrarre informazione dai dati con strumenti opportuni e coerenti, per tentare di rispondere criticamente ad interrogativi utili. La scelta di utilizzare un approccio particolare rispetto ad un altro ha senso solo nel momento in cui ci si condiziona al problema specifico che si intende affrontare; alcune soluzioni presenteranno vantaggi - computazionali, interpretativi - rispetto alle altre, e all'analista spetterà la scelta di valutare quale strada sia la migliore da percorrere, "marginalizzando" rispetto a tutte quelle percorribili.

Ferme restando queste premesse, il modello proposto verrà collocato entro il contesto bayesiano; verrà quindi ipotizzato un processo generatore per i dati condizionato ad alcuni parametri, una distribuzione a priori per i parametri coinvolti, e l'obiettivo dell'inferenza sarà quindi la descrizione, analitica o ottenuta tramite simulazione, della distribuzione a posteriori dei parametri coinvolti dal modello. Per quanto riguarda le previsioni dei modelli proposti, il candidato naturale in quest'ambito è la *densità predittiva a posteriori*, ossia la densità delle nuove osservazioni condizionatamente ai dati attualmente disponibili, marginalizzando rispetto alla distribuzione a posteriori dei parametri coinvolti. Da un punto di vista bayesiano l'obiettivo del controllo a posteriori di un modello non è quello di falsificare lo stesso o il confronto tra modelli differenti, quanto piuttosto l'individuazione di quali particolari aspetti dei dati non sono stati colti in modo corretto dallo strumento adottato (Gelman e Shalizi, 2013), in modo tale da fornirne una specificazione migliore.

In questo lavoro, tuttavia, le previsioni del contesto bayesiano, ossia delle distribuzioni di probabilità, all'occorrenza saranno sintetizzate e utilizzate come se

fossero dei veri e propri stimatori frequentisti, ad esempio per confrontare modelli differenti, oppure per valutare la performance dei modelli su dei dati non osservati in fase di stima. Se questo modo di operare può sembrare scorretto o incompatibile, si citano alcuni esempi illuminanti. Nel caso di modelli con spazio parametrico finito, si dimostra come la consistenza dello stimatore di massima verosimiglianza implichi che lo stimatore bayesiano sia, rispetto ad appropriate funzioni di perdita, fortemente consistente e asintoticamente efficiente (Prakasa Rao, 1987, sez 3.4). In Hastie et al. (2009, sez. 8.4) viene mostrato come la distribuzione bootstrap (parametrica e non parametrica) degli stimatori coinvolti in certi modelli possa essere considerata un'approssimazione della distribuzione a posteriori degli stessi parametri, specificando opportune distribuzioni a priori limite.

I punti di contatto tra i due paradigmi sono molteplici, e quindi l'utilizzo ibrido dei due approcci, se coerente e critico, è decisamente giustificato; precisiamo tuttavia che, viste le premesse succitate, il presente lavoro si colloca globalmente in un ambito frequentista, sebbene utilizzi anche dei modelli appartenenti al paradigma bayesiano.

2.4 Modello ad effetti fissi

Il modello proposto suppone che gli archi della rete risposta siano delle variabili bernoulliane condizionatamente indipendenti, date le rispettive probabilità di successo; è quindi sensato specificare tali probabilità in funzione degli archi presenti nelle altre reti, per coglierne il contributo nella variazione media della rete risposta.

Il predittore lineare che collega la rete risposta e le reti esplicative è quindi scomposto in due componenti: la prima cattura la relazione tra reti esplicative e rete risposta, mentre la seconda tiene conto delle strutture di dipendenza che le variabili esplicative non sono in grado di cogliere, permettendo quindi di stimare la prima in modo adeguato.

Formalmente, la specificazione del modello è la seguente:

$$y_{ij} | \pi_{ij} \sim \text{Bin}(1, \pi_{ij}) \quad \pi_{ij} = \frac{\exp(\theta_{ij})}{1 + \exp(\theta_{ij})} \quad (2.1)$$

$$\theta_{ij} = \underbrace{\alpha + \sum_{k=1}^p \beta_k x_{ijk}}_{\text{Componente parametrica}} + \underbrace{\sum_{h=1}^{\infty} \lambda_h z_{ih} z_{jh}}_{\text{Componente non parametrica}}$$

Nella 2.1 $y_{ij} \in \{0, 1\}$ rappresenta il (i, j) -esimo elemento della matrice \mathbf{Y} , mentre $\theta_{ij} \in \mathbb{R}$ è il predittore lineare legato alla probabilità di successo tramite funzione logistica. La componente parametrica è formata da un'intercetta comune $\alpha \in \mathbb{R}$ e da un vettore formato da p coefficienti, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p) \in \mathbb{R}^p$, ognuno dei quali moltiplica la rispettiva matrice di adiacenza. L'inferenza su questo vettore permette di valutare direzione ed intensità della relazione tra le reti esplicative e la rete risposta.

La componente non parametrica, invece, scompone le deviazioni tra predittori lineari e componente parametrica in una struttura latente moltiplicativa; gli elementi (z_{i1}, z_{i2}, \dots) , $z_{ih} \in \mathbb{R}$ rappresentano le coordinate di ogni nodo i in uno spazio latente di dimensione infinita (da qui la dicitura di non parametrico), mentre $\lambda_h \in \mathbb{R}_{\geq 0}$ definisce il peso delle diverse dimensioni.

Definizione 5 (Prodotto contratto lungo la terza dimensione) *Sia \mathbf{M} un tensore tridimensionale di dimensione $(a \times b \times c)$ e \mathbf{n} un vettore di lunghezza c . Si definisce "prodotto contratto lungo la terza dimensione" l'operatore indicato con $\bar{\times}_3$ tale per cui $\mathbf{M} \bar{\times}_3 \mathbf{n} = \mathbf{G}$ è una matrice di dimensione $(a \times b)$ in cui il generico elemento $\mathbf{G}_{ij} = \sum_{k=1}^c m_{ijk} n_k$. Si rimanda a [Bader e Kolda \(2006\)](#) e [Lee e Cichocki \(2014\)](#) per ulteriori proprietà e generalizzazioni.*

La matrice $v \times v$ simmetrica Θ dei predittori lineari θ_{ij} può essere espressa anche tramite formulazione matriciale:

$$\Theta = \alpha + \mathbf{X} \bar{\times}_3 \boldsymbol{\beta} + \mathbf{Z} \boldsymbol{\Lambda} \mathbf{Z}^T \quad (2.2)$$

Nella formulazione matriciale 2.2, quindi, la matrice \mathbf{Z} è composta, per ogni riga, dalle infinitamente-numerabili coordinate latenti per ogni nodo, mentre $\boldsymbol{\Lambda}$ è la matrice diagonale con elementi $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots)$, $\lambda_h \in \mathbb{R}_{\geq 0}$. Si noti che in questo

modo è possibile conservare la struttura delle matrici di adiacenza, il che rende molto più intuitiva la lettura del modello.

Un approccio parametrico a questa impostazione prevederebbe, tendenzialmente, la stima di diversi modelli per diversi valori della dimensione dello spazio latente, confrontando poi sulla base di qualche criterio i modelli così stimati per sceglierne la complessità ottimale. Ricordando gli obiettivi di questo lavoro, non abbiamo alcuna garanzia che la dimensione "migliore" dello spazio latente, ritenuta tale sulla base dei dati attualmente disponibili, rimanga costante nel momento in cui si avranno a disposizione informazioni su nuovi soggetti; al contrario, è lecito supporre che al crescere del numero di nodi sia necessario uno spazio più complicato per descrivere in modo opportuno le intricate strutture di dipendenza tra i soggetti.

In letteratura esistono alcune tecniche in grado di trattare il *numero* di fattori latenti come un parametro ignoto; ad esempio, [Lopes e West \(2004\)](#) propongono un approccio in grado di variare la dimensione dello spazio parametrico durante la stima del modello, di fatto tramite un approccio di tipo *Reversible-jump* MCMC ([Brooks et al., 2011](#), cap. 3). Sebbene questa strategia permetta di tenere conto dell'incertezza derivante dalla stima della complessità del modello, si è rivelata particolarmente onerosa dal punto di vista computazionale; sono richieste, infatti, diverse *run* preliminari, e una regolazione non banale di alcuni passi del modello stesso.

Da un punto di vista teorico, un approccio non-parametrico permette di formulare un modello che sia sufficientemente generale, in grado di comprendere tutte le possibili soluzioni al problema specifico affrontato: per rappresentare adeguatamente *ogni* possibile sfumatura della situazione che si intende analizzare, si suppone che non sia sufficiente un insieme finito di parametri; in assenza di ulteriori informazioni, pare un'ipotesi più elegante rispetto ad assumere che il numero di parametri sia sempre lo stesso, oppure che sia una quantità ignota da stimare, anche quando il numero dei nodi tende a infinito. Da un punto di vista pratico, si concede allo stesso modello la flessibilità necessaria per adattarsi a situazioni molto differenti, evitando il rischio di semplificare (o complicare) eccessivamente il problema, ferma restando la medesima specificazione e procedimento di stima. Si rimanda a [Orbanz \(2012\)](#) per ulteriori riflessioni interessanti in questa direzione.

2.4.1 Distribuzioni a priori

La componente latente può anche essere parametrizzata come $\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T$, matrice simmetrica di dimensione $v \times v$, dove $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{\Lambda}^{\frac{1}{2}}$. Generalmente l'identificabilità dei fattori latenti si ottiene vincolando a 0 gli elementi triangolari superiori di $\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T$ (Geweke e Zhou, 1996); questa operazione tuttavia ha il difetto di imporre una sorta di gerarchia tra i nodi della rete, assunzione che noi vorremmo assolutamente evitare. L'ordine dei nodi, in questo problema, è perfettamente scambiabile: assumere che il particolare ordinamento in cui compaiono porti informazione pare decisamente forte come ipotesi. Da un punto di vista bayesiano, l'identificabilità di questa componente spesso non è indispensabile, soprattutto in casi come quello in oggetto, in cui l'inferenza riguarda la componente parametrica del modello. Per rigore, precisiamo che l'unico vincolo è che tutti gli elementi di $\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T$ siano finiti; si rimanda a Bhattacharya e Dunson (2011) per ulteriori precisazioni a questo riguardo.

Una specificazione bayesiana completa del modello 2.1 richiede la definizione delle distribuzioni a priori per i parametri coinvolti. Per quando riguarda $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ e α si sono scelte, rispettivamente una distribuzione a priori normale multivariata a componenti indipendenti e una distribuzione normale:

$$\boldsymbol{\beta} \sim N_p(\boldsymbol{\mu}_{0_\beta}, \boldsymbol{\Sigma}_{\beta_0}), \quad \boldsymbol{\mu}_{0_\beta} \in \mathbb{R}^p, \quad \boldsymbol{\Sigma}_{0_\beta} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2) \in \mathbb{R}_{\geq 0}^p$$

$$\alpha \sim N(\mu_{0_\alpha}, \sigma_{0_\alpha}^2), \quad \mu_{0_\alpha} \in \mathbb{R}, \quad \sigma_{0_\alpha}^2 \in \mathbb{R}_{\geq 0}$$

Come in ambito bayesiano parametrico le distribuzioni a priori su oggetti di dimensione finita sono delle variabili aleatorie, così in un contesto non-parametrico le misure di probabilità su oggetti casuali di dimensione infinita assumono la forma di processi stocastici. Adattando al caso in oggetto Bhattacharya e Dunson (2011), la distribuzione a priori su $\boldsymbol{\lambda}$, il vettore composto dagli elementi della matrice diagonale $\mathbf{\Lambda}$, è un processo stocastico gamma-inverso moltiplicativo.

Definizione 6 (Processo gamma-inverso moltiplicativo) *La sequenza numerabile di variabili aleatorie $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots)$ definisce un processo stocastico gamma-inverso moltiplicativo di parametri (a_1, a_2) se ogni generico elemento h ha distri-*

buzione gamma-inversa moltiplicativa:

$$\lambda_h = \prod_{m=1}^h \frac{1}{\delta_m}, \quad \delta_1 \sim \text{Gamma}(a_1, 1), \quad \delta_{m>1} \sim \text{Gamma}(a_2, 1)$$

con δ_m , $m = 1, \dots, \infty$ tra loro indipendenti. (*Bhattacharya e Dunson, 2011*)

La distribuzione a priori specificata nella definizione 6 permette di eliminare agevolmente le dimensioni ridondanti, favorendo un numero ridotto di componenti quando la vera struttura è sparsa. Nello specifico, la distribuzione a priori utilizzata sfavorisce modelli sovra-parametrizzati imponendo che gli elementi λ_h siano stocasticamente decrescenti verso 0 all'aumentare della dimensione dello spazio latente. Le proprietà della gamma-inversa moltiplicativa sono garantite finché $a_1 > 1$, $a_2 > 2$; si rimanda a *Bhattacharya e Dunson (2011)* per ulteriori proprietà e dimostrazioni, e a *Durante et al. (2014)* per un'applicazione elegante in un contesto di reti.

Per quanto riguarda la specificazione a priori delle coordinate latenti, ad ogni elemento z_{ih} della matrice \mathbf{Z} è stata assegnata una distribuzione a priori normale standard indipendente, $z_{ih} \sim N(0, 1)$, $i = 1, \dots, v$. Si noti che nella riparametrizzazione $\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T$ si ha che la distribuzione a priori condizionata $\tilde{z}_{ih}|\lambda_h \sim N(0, \lambda_h)$; questa proprietà è cruciale per mantenere coniugazione.

2.4.2 Distribuzione a posteriori

La distribuzione a posteriori per i parametri di interesse non è ricavabile analiticamente, ed è quindi necessario a ricorrere a metodi di simulazione appropriati. Nelle ultime decenni i metodi di tipo MCMC (*Markov Chain Monte Carlo*) hanno permesso di fare inferenza per modelli molto complessi, per i quali sarebbe praticamente impossibile calcolare la costante di normalizzazione coinvolta nella distribuzione a posteriori; si rimanda a *Albert (2009)* e *Brooks et al. (2011)* per un'introduzione e una trattazione adeguata su questo tipo di tecniche.

Chiaramente, per fini computazionali, è impossibile simulare un numero infinito di parametri, quindi è necessaria una strategia appropriata. Il processo gamma-inverso moltiplicativo può essere approssimato da una sequenza di H variabili omonime, per un certo valore H sufficientemente elevato (*Bhattacharya e Dunson, 2011*, sez. 2.2); nella trattazione seguente, dunque, si sfrutterà questo risultato

per ottenere trattabilità computazionale. Precisiamo che questo modo di operare è intrinsecamente diverso dalla stima di un modello in cui la dimensione dello spazio latente è *esattamente* pari ad H : questo valore rappresenta, infatti, solo un limite superiore conservativo per il numero di componenti; grazie alle proprietà del processo Gamma-inverso, inoltre, ci aspettiamo che quelle effettivamente utilizzate siano molte meno.

Per implementare un efficiente Gibbs sampling in grado di simulare dalla distribuzione a posteriori, si adatta l'approccio di [Polson et al. \(2013\)](#), il quale prevede una strategia di *data-augmentation* tramite l'introduzione di una variabile latente Pólya-Gamma.

Definizione 7 (Distribuzione Pólya-Gamma) *La variabile aleatoria X si distribuisce come una variabile Pólya-Gamma ([Polson et al., 2013](#)) di parametri $b > 0$ e $c \in \mathbb{R}$, indicata come $X \sim PG(b, c)$, se*

$$X \stackrel{D}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k - 1/2)^2 + c^2/(4\pi^2)}, \quad g_k \stackrel{iid}{\sim} \text{Gamma}(b, 1)$$

Il risultato fondamentale di questa famiglia di distribuzioni è che la verosimiglianza binomiale, parametrizzata tramite il *log-odd*, è equivalente ad una mistura di Normali rispetto ad una distribuzione Pólya-Gamma; quindi, se le distribuzioni a priori per i parametri di regressione di un modello logistico sono delle distribuzioni normali, condizionandosi ad un'opportuna variabile Pólya-Gamma si ottiene coniugazione. L'applicabilità del metodo proposto deriva dal fatto che gli autori propongono un valido algoritmo di accettazione-rifuto per generare agevolmente da queste distribuzioni. Si rimanda a [Polson et al. \(2013\)](#) per ulteriori proprietà teoriche su questa famiglia di variabili e sull'algoritmo generativo, e a [Choi et al. \(2013\)](#) per la dimostrazione dell'ergodicità uniforme delle catene così generate.

2.4.3 Gibbs sampling per il modello ad effetti fissi

La simulazione dalla distribuzione a posteriori si ottiene iterativamente tramite i seguenti passaggi:

Algoritmo 1 (Gibbs Sampling per il modello a effetti fissi)

1. **Aggiornamento della variabile latente Pólya-Gamma** Secondo la notazione definita in [Kolda e Bader \(2009, sez. 2\)](#), sia $\underline{\mathbf{X}}_{::k}$ il k -esimo livello dell'array $\underline{\mathbf{X}}$, ossia la matrice di adiacenza relativa alla k -esima rete esplicativa. Per alleggerire la notazione, indichiamo con $\bar{\mathbf{X}}$ la matrice $v(v-1)/2 \times p$ ottenuta affiancando gli elementi triangolari inferiori delle matrici $\underline{\mathbf{X}}_{::k}$, $k = 1, \dots, p$.

Si estraggono quindi dalla *full conditional* le variabili latenti Pólya-Gamma:

$$\omega_l | - \sim PG(1, \alpha + \bar{\mathbf{X}}_l \boldsymbol{\beta} + \mathcal{L}(\mathbf{Z} \boldsymbol{\Lambda} \mathbf{Z}^T)_l), \quad l = 1, \dots, \frac{v(v-1)}{2}$$

2. **Aggiornamento dei coefficienti parametrici** Sia $\boldsymbol{\Omega}$ la matrice diagonale $\frac{v(v-1)}{2} \times \frac{v(v-1)}{2}$ di elementi ω_l .

Si estrae il vettore $\boldsymbol{\beta}$ dalla *full-conditional* normale:

$$\boldsymbol{\beta} | - \sim N_p(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta), \quad \boldsymbol{\Sigma}_\beta = \left(\bar{\mathbf{X}}^T \boldsymbol{\Omega} \bar{\mathbf{X}} + \boldsymbol{\Sigma}_{0_\beta} \right)^{-1}$$

$$\boldsymbol{\mu}_\beta = \boldsymbol{\Sigma}_\beta \left(\bar{\mathbf{X}}^T \tilde{\mathbf{y}} - \bar{\mathbf{X}}^T \boldsymbol{\Omega} \left[\mathcal{L}(\mathbf{Z} \boldsymbol{\Lambda} \mathbf{Z}^T) + \alpha \right] + \boldsymbol{\Sigma}_{0_\beta}^{-1} \boldsymbol{\mu}_{0_\beta} \right)$$

con $\tilde{\mathbf{y}} = (y_1 - \frac{1}{2}, y_1 - \frac{1}{2}, \dots, y_{v(v-1)/2} - \frac{1}{2})$

3. **Aggiornamento dell'intercetta** Si estrae il valore α dalla *full-conditional* normale:

$$\alpha | - \sim N(\mu_\alpha, \sigma_\alpha^2), \quad \sigma_\alpha^2 = \frac{1}{\left(\sum_{i=1}^{v(v-1)/2} \omega_i + 1/\sigma_{0_\alpha}^2 \right)},$$

$$\mu_\alpha = \sigma_\alpha^2 \left(\sum_{l=1}^{v(v-1)/2} \left[\tilde{\mathbf{y}}_l - \boldsymbol{\Omega} \left[\bar{\mathbf{X}}_l \boldsymbol{\beta} + \mathcal{L}(\mathbf{Z} \boldsymbol{\Lambda} \mathbf{Z}^T)_l \right] \right] + \frac{\mu_{0_\alpha}}{\sigma_{0_\alpha}^2} \right)$$

4. **Aggiornamento delle coordinate latenti** L'idea di base è quella di sfruttare la struttura matriciale definita nella [2.2](#), per aggiornare ogni riga di $\tilde{\mathbf{Z}}$ indipendentemente. Ricordiamo, infatti, che condizionandosi alla struttura latente e alla componente parametrica, gli archi sono variabili bernoulliane indipendenti. Questo implica che per aggiornare le componenti latenti è sufficiente isolare il contributo delle righe i -esime di ogni matrice $v \times v$.

Siano, per $k = 1, \dots, v$, $\tilde{\mathbf{y}}_{(k)}$, $\tilde{\mathbf{Z}}_{(k)}$, $\bar{\mathbf{X}}_{(k)}$ e $\boldsymbol{\Omega}_{(k)}$ le matrici ottenute selezionando, al passo [1](#), gli elementi l riferiti alle coppie di nodi $\{i, j\}$ per le quali

$i = k$ oppure $j = k$, con $i < j$; questa operazione equivale, nella formulazione matriciale 2.2, ad estrarre separatamente le v righe delle matrici quadrate coinvolte. Definiamo, infine, $\tilde{\mathbf{Z}}_{(-k)}$ la matrice ottenuta rimuovendo la riga k -esima dalla matrice $\tilde{\mathbf{Z}}$.

Quindi, per ogni $k = 1, \dots, v$

$$\tilde{\mathbf{Z}}_{(k)} | - \sim N_H(\mu_{z_k}, \Sigma_{z_k}), \quad \Sigma_k = \left(\tilde{\mathbf{Z}}_{(-k)}^T \Omega_{(k)} \tilde{\mathbf{Z}}_{(-k)} + \Lambda^{-1} \right)^{-1}$$

$$\mu_{z_k} = \Sigma_k \left(\tilde{\mathbf{Z}}_{(-k)}^T \left[\tilde{\mathbf{y}}_{(k)} - \Omega_{(k)} \left[\bar{\mathbf{X}}_{(k)} \boldsymbol{\beta} + \alpha \right] \right] \right)$$

5. **Aggiornamento della matrice Λ** Si estraggono gli elementi $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_H)$ definiti nella definizione 6 dalle *full-conditional*

$$\delta_1 | - \sim \text{Gamma} \left(a_1 + \frac{VH}{2}, 1 + \frac{1}{2} \sum_{m=1}^H \theta_m^{(-1)} \sum_{k=1}^v \tilde{\mathbf{Z}}_{km}^2 \right)$$

$$\delta_{h>2} | - \sim \text{Gamma} \left(a_2 + \frac{V(H+1-h)}{2}, 1 + \frac{1}{2} \sum_{m=h}^H \theta_m^{(-h)} \sum_{k=1}^v \tilde{\mathbf{Z}}_{km}^2 \right)$$

dove $\theta_m^{(-i)} = \prod_{\substack{t=1 \\ t \neq i}}^m \delta_t$, con $\theta_1^{(-1)} = 1$. Quindi $\lambda_h = \prod_{m=1}^h \frac{1}{\delta_m}$

6. **Aggiornamento del predittore lineare** Si aggiorna infine la matrice simmetrica Θ come da definizione.

$$\Theta = \alpha + \mathbf{X} \bar{\mathbf{x}}_3 \boldsymbol{\beta} + \tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^T$$

2.5 Modello ad effetti casuali

Nel modello 2.1, le reti esplicative sono considerate, di fatto, come degli attributi relativi ai singoli archi. Questa assunzione delega totalmente alla struttura fattoriale latente il compito di rappresentare in modo adeguato le strutture di dipendenza che le variabili esplicative non sono riuscite a cogliere. Una formulazione più elegante potrebbe sfruttare meglio le reti esplicative, utilizzandole per spiegare una parte delle eterogeneità e dipendenze interne alla rete risposta. Dal punto

di vista della descrizione del fenomeno, una tale formulazione potrebbe aiutare gli esperti nell'individuazione dei ruoli ricoperti dai terroristi nei diversi rapporti; vorremmo identificare, ad esempio, i soggetti con i quali l'aver avuto una particolare relazione aumenti, rispetto alla media, la propensione a collaborarvi in un'operazione.

Per riuscire in questo intento ha senso formulare un modello con dei coefficienti casuali sulle reti esplicative, riferiti ad ogni nodo, che contribuiscono in forma additiva al predittore lineare. I modelli a coefficienti casuali sono ormai una pietra miliare della modellazione statistica di dati con dipendenze; si veda ad esempio [Gelman e Hill \(2006\)](#) per un'introduzione a queste tecniche.

Il modello si modifica nel modo seguente:

$$\begin{aligned}
 y_{ij} | \pi_{ij} &\sim \text{Bin}(1, \pi_{ij}) & \pi_{ij} &= \frac{\exp(\theta_{ij})}{1 + \exp(\theta_{ij})} \\
 \theta_{ij} &= \underbrace{\alpha + \sum_{k=1}^p [\beta_k + b_{ki} + b_{kj}] x_{kij}}_{\substack{\text{Componente parametrica:} \\ \text{effetti fissi e casuali}}} + \underbrace{\sum_{h=1}^{\infty} \lambda_h z_{ih} z_{jh}}_{\text{Componente non parametrica}} & (2.3) \\
 (b_{k1}, \dots, b_{kv}) &\sim N(0, \sigma_{bk}^2), \quad k = 1, \dots, p
 \end{aligned}$$

Il contributo di ogni rete esplicativa k può essere scomposto in due addendi: un effetto fisso comune, pari a β_k , e un effetto casuale indotto dalla coppia di nodi coinvolta nel (i, j) -esimo arco, ossia dalla somma tra l'effetto casuale relativo al nodo i , b_{ki} , e quello relativo al nodo j , b_{kj} ; gli effetti casuali di ogni nodo possono quindi essere interpretati come lo scostamento dall'effetto medio comune, pari a β_k .

La formulazione matriciale è in questo caso leggermente più pesante, poiché è necessario esplicitare in modo opportuno il tensore relativo agli effetti casuali.

Definizione 8 (Prodotto di Hadamard) *Siano \mathbf{A} e \mathbf{B} due tensori della stessa generica dimensione $(I_1 \times \dots \times I_n)$. Si definisce prodotto di Hadamard, o prodotto elemento-per-elemento, l'operatore indicato con \circ tale per cui $\mathbf{A} \circ \mathbf{B} = \mathbf{C}$, dove \mathbf{C} è ancora un tensore di dimensione $(I_1 \times \dots \times I_n)$, tale per cui il generico elemento $\mathbf{C}_{(i_1, \dots, i_n)} = \mathbf{A}_{(i_1, \dots, i_n)} \mathbf{B}_{(i_1, \dots, i_n)}$*

Sia $\underline{\mathbf{B}}$ il tensore di dimensione $(v \times v \times p)$, nel quale il generico elemento $\underline{B}_{ijk} = b_{ki} + b_{kj}$; per chiarimenti riguardo la costruzione operativa di questa quantità si rimanda all'appendice A.1. Moltiplicando elemento per elemento con il tensore $\underline{\mathbf{X}}$ e applicando l'operatore della definizione 5 si ottiene la matrice $(v \times v)$ formata dalla componente ad effetti casuali definita in 2.3

$$\Theta = \alpha + \mathbf{X} \bar{\times}_3 \beta + (\underline{\mathbf{B}} \circ \mathbf{X}) \bar{\times}_3 \mathbf{1}_p + \mathbf{Z} \Lambda \mathbf{Z}^T \quad (2.4)$$

2.5.1 Gibbs sampling per il modello ad effetti casuali

La simulazione dalla distribuzione a posteriori si ottiene iterativamente tramite i seguenti passaggi:

Algoritmo 2 (Gibbs Samping per il modello a effetti casuali)

1. **Aggiornamento della variabile latente Pólya-Gamma** Per alleggerire la notazione, sia $\widetilde{\mathbf{B}\mathbf{X}} = (\underline{\mathbf{B}} \circ \mathbf{X}) \bar{\times}_3 \mathbf{1}_p$, la matrice $(v \times v)$ contenente al suo interno predittori lineari riferiti alla componente ad effetti casuali.

Si estraggono quindi dalla *full conditional* le variabili latenti Pólya-Gamma:

$$\omega_l | - \sim PG(1, \alpha + \bar{\mathbf{X}}_l \beta + \mathcal{L}(\mathbf{Z} \Lambda \mathbf{Z}^T + \widetilde{\mathbf{B}\mathbf{X}})_l), \quad l = 1, \dots, \frac{v(v-1)}{2}$$

2. **Aggiornamento degli effetti fissi**

Si estrae il vettore β dalla *full-conditional* normale:

$$\beta | - \sim N_p(\mu_\beta, \Sigma_\beta), \quad \Sigma_\beta = \left(\bar{\mathbf{X}}^T \Omega \bar{\mathbf{X}} + \Sigma_{0_\beta} \right)^{-1}$$

$$\mu_\beta = \Sigma_\beta \left(\bar{\mathbf{X}}^T \tilde{\mathbf{y}} - \bar{\mathbf{X}}^T \Omega \left[\mathcal{L}(\mathbf{Z} \Lambda \mathbf{Z}^T + \widetilde{\mathbf{B}\mathbf{X}}) + \alpha \right] + \Sigma_{0_\beta}^{-1} \mu_{0_\beta} \right)$$

$$\text{con } \tilde{\mathbf{y}} = (y_1 - \frac{1}{2}, y_1 - \frac{1}{2}, \dots, y_{v(v-1)/2} - \frac{1}{2})$$

3. **Aggiornamento dell'intercetta** Si estrae il valore α dalla *full-conditional* normale:

$$\alpha | - \sim N(\mu_\alpha, \sigma_\alpha^2), \quad \sigma_\alpha^2 = \frac{1}{\left(\sum_{l=1}^{v(v-1)/2} \omega_l + 1/\sigma_{0_\alpha}^2\right)},$$

$$\mu_\alpha = \sigma_\alpha^2 \left(\sum_{l=1}^{v(v-1)/2} \left[\tilde{\mathbf{y}}_l - \Omega \left[\bar{\mathbf{X}}_l \boldsymbol{\beta} + \mathcal{L}(\mathbf{Z} \boldsymbol{\Lambda} \mathbf{Z}^T + \widetilde{\mathbf{B}} \mathbf{X})_l \right] \right] + \frac{\mu_{0_\alpha}}{\sigma_{0_\alpha}^2} \right)$$

4. **Aggiornamento degli effetti casuali** Sia \mathbf{B} la matrice di dimensione $(v \times p)$ ottenuta incolonnando i vettori composti dai p effetti casuali di ogni soggetto. Siano inoltre $\mathbf{B}_{(k)}$ e $\mathbf{B}_{(-k)}$, rispettivamente, la k -esima riga e la matrice nella quale è stata rimossa la riga k -esima della matrice \mathbf{B} , $k = 1, \dots, v$; sia infine $(\mathbf{Z} \boldsymbol{\Lambda} \mathbf{Z}^T)_{(k)}$ la k -esima riga della matrice $\mathbf{Z} \boldsymbol{\Lambda} \mathbf{Z}^T$ (privata dell'elemento in posizione k , non definito per costruzione). Sfruttando la formulazione matriciale è possibile aggiornare la matrice \mathbf{B} a blocchi, simulando cioè gli effetti casuali relativi ad ogni soggetto condizionatamente agli altri; si noti, infatti, che dalla 2.4, deriva che ogni riga della matrice dei predittori lineari Θ può essere riscritta come:

$$\Theta_{(k)} = \alpha + \bar{\mathbf{X}}_{(k)} \boldsymbol{\beta} + (\mathbf{Z} \boldsymbol{\Lambda} \mathbf{Z}^T)_{(k)} + \bar{\mathbf{X}}_{(k)} \mathbf{B}_{(k)} + (\bar{\mathbf{X}}_{(k)} \circ \mathbf{B}_{(-k)}) \bar{\times}_3 \mathbf{1}_p$$

Quindi, per ogni $k = 1, \dots, v$:

$$\mathbf{B}_{(k)} | - \sim N_p(\mu_{b_k}, \Sigma_{b_k}), \quad \Sigma_{b_k} = \left(\bar{\mathbf{X}}_{(k)}^T \boldsymbol{\Omega}_{(k)} \bar{\mathbf{X}}_{(k)} + \Sigma_{b_{k0}} \right)^{-1}$$

$$\mu_{b_k} = \Sigma_{b_k} \left(\bar{\mathbf{X}}_{(k)}^T \left[\tilde{\mathbf{y}}_{(k)} - \Omega_{(k)} \left[\alpha + \bar{\mathbf{X}}_{(k)} \boldsymbol{\beta} + (\mathbf{Z} \boldsymbol{\Lambda} \mathbf{Z}^T)_{(k)} + (\bar{\mathbf{X}}_{(k)} \circ \mathbf{B}_{(-k)}) \bar{\times}_3 \mathbf{1}_p \right] \right] \right)$$

5. **Aggiornamento delle coordinate latenti** Per ogni $k = 1, \dots, v$

$$\tilde{\mathbf{Z}}_{(k)} | - \sim N_H(\mu_{z_k}, \Sigma_{z_k}), \quad \Sigma_{z_k} = \left(\tilde{\mathbf{Z}}_{(-k)}^T \boldsymbol{\Omega}_{(k)} \tilde{\mathbf{Z}}_{(-k)} + \boldsymbol{\Lambda}^{-1} \right)^{-1}$$

$$\mu_{z_k} = \Sigma_{z_k} \left(\tilde{\mathbf{Z}}_{(-k)}^T \left[\tilde{\mathbf{y}}_{(k)} - \Omega_{(k)} \left[\alpha + \bar{\mathbf{X}}_{(k)} \boldsymbol{\beta} + \widetilde{\mathbf{B}} \mathbf{X}_{(k)} \right] \right] \right)$$

6. **Aggiornamento della matrice Λ** Si estraggono gli elementi $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_H)$ definiti nella definizione 6 dalle *full-conditional*

$$\delta_1 | - \sim \text{Gamma} \left(a_1 + \frac{VH}{2}, 1 + \frac{1}{2} \sum_{m=1}^H \theta_m^{(-1)} \sum_{k=1}^v \tilde{\mathbf{Z}}_{km}^2 \right)$$

$$\delta_{h>2} | - \sim \text{Gamma} \left(a_2 + \frac{V(H+1-h)}{2}, 1 + \frac{1}{2} \sum_{m=h}^H \theta_m^{(-h)} \sum_{k=1}^v \tilde{\mathbf{Z}}_{km}^2 \right)$$

$$\text{dove } \theta_m^{(-i)} = \prod_{\substack{t=1 \\ t \neq i}}^m \delta_t, \quad \text{con } \theta_1^{(-1)} = 1. \quad \text{Quindi } \lambda_h = \prod_{m=1}^h \frac{1}{\delta_m}$$

7. **Aggiornamento del predittore lineare** Si aggiorna infine la matrice simmetrica Θ come da definizione.

$$\Theta = \alpha + \mathbf{X} \bar{\times}_3 \boldsymbol{\beta} + (\underline{\mathbf{B}} \circ \mathbf{X}) \bar{\times}_3 \mathbf{1}_p + \tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^T$$

Capitolo 3

Simulazione

Per valutare la correttezza dei modelli proposti, è stato condotto un'appropriato studio di simulazione; lo studio è stato eseguito in diversi scenari di complessità crescente, i quali identificano diversi modelli stocastici generatori per le reti coinvolte.

3.1 Scenari di simulazione

Siano, per tutti gli scenari, v il numero di nodi e β il vettore di coefficienti sul quale si vuole fare inferenza, entrambi fissati dall'utente in fase di generazione dati.

1. Primo Scenario

Il modello generatore di reti più semplice è denominato modello di Erdős-Rényi, il quale prevede che tutti gli archi siano delle variabili bernoulliane indipendenti con probabilità di successo pari a π (Erdős e Rényi 1959). In questo scenario le reti esplicative e risposta vengono generate da un modello molto semplice secondo i seguenti passaggi:

- 1.1. Generazione di due reti esplicative con numero di nodi v e probabilità di formare un arco pari a 0.3 e 0.5, secondo il modello di Erdős-Rényi. Le reti sono forzate ad essere indirette, quindi le risultanti matrici di adiacenza saranno simmetriche per costruzione.

- 1.2. Le rispettive matrici di adiacenza vengono dunque "sovrapposte" per formare il tensore \mathbf{X} , il quale forma poi il predittore lineare secondo la seguente espressione: $\Theta = \mathbf{X} \bar{\times}_3 \beta$. Corrisponde ad un caso particolare del modello 2.1, nel quale la struttura "latente" e il parametro α sono nulli.
- 1.3. Ogni arco y_{ij} della rete risposta è generato indipendentemente come $\text{Bin}(1, 1/[1 + \exp(-\theta_{ij})])$, dove θ_{ij} è il generico elemento della matrice Θ ; per preservare simmetria matriciale, questa operazione si segue solo sulle rispettive triangolari inferiori, adattando i risultati di conseguenza.

2. Secondo Scenario

Corrisponde a generare le osservazioni dal modello 2.1. Mantenendo le stesse condizioni dello scenario 1, il secondo scenario genera la rete risposta esattamente dal modello 2.2. La struttura "latente" $\mathbf{Z}\Lambda\mathbf{Z}^T$ è stata costruita fissando $h = 2$, $\lambda = (3, 5)$, e generando le coordinate latenti da dalle normali standard; dopo aver calcolato il predittore lineare come in 2.2, gli archi della variabile risposta sono generati indipendentemente da dalle variabili bernoulliane. In questo modo si introduce una dipendenza tra i nodi nella rete risposta, poiché soggetti con coordinate latenti vicine avranno tendenzialmente una propensione diversa dalla media a formare una connessione, al netto del contributo derivante dalle reti esplicative.

3. Terzo Scenario

Per rendere la situazione più realistica, è sensato utilizzare modelli generatori più complessi per le reti esplicative, che tentino di riprodurre strutture di rete articolate. In questo modo si introducono nelle reti esplicative delle forme di dipendenza tra i nodi che si trasmettono, di conseguenza, nella rete risposta. Le reti di questo scenario sono state generate secondo due modelli:

- (a) **Modello geometrico** Dispone casualmente, in modo uniforme, i nodi in un quadrato bidimensionale unitario, e poi collega quelli la cui distanza risulta minore di una soglia k , definita dalla norma Euclidea. Se $k = 0$ si ottiene una rete senza connessioni, se $k \geq \sqrt{2}$ una rete con densità unitaria. Nella simulazione questo valore è stato fissato a $k = 0.3$

- (b) **Modello di Barabási–Albert.** Crea reti ad invarianza di scala, ossia caratterizzate da una distribuzione dei nodi che segue una legge esponenziale. Genera la rete in modo iterativo, aggiungendo ad ogni passo n un nodo aggiuntivo, il quale si collega ai nodi presenti nel grafo con una probabilità proporzionale a d_i^k , dove d_i , $i = 1, \dots, n - 1$ rappresenta i gradi dei nodi presenti, e k è una costante specificata dall'utente. Genera una struttura del tipo "ricco-sempre-più-ricco", in cui il parametro k determina la velocità di crescita di questa dinamica (Barabási e Albert 1999). Nella simulazione questo valore è stato fissato a $k = 3$

Utilizzando come esplicative reti generate da questi algoritmi, la rete risposta è stata simulata come nello scenario 2.

4. Quarto Scenario

Questo scenario corrisponde a generare le osservazioni dal modello ad effetti casuali 2.3; gli effetti casuali sui coefficienti β sono stati estratti dalle normali standard, mentre tutte le altre condizioni sono le stesse dello scenario 2.

5. Quinto Scenario

Per quest'ultimo scenario, infine, sono state mantenute le stesse condizioni dello scenario 4, utilizzando tuttavia le reti esplicative più realistiche generate nello scenario 3.

Ogni scenario è stato simulato per un numero di nodi pari a $v = (20, 50)$; i coefficienti sono stati fissati a valori pari a $\beta = (-2, 1)$.

3.2 Modelli competitori

I modelli proposti sono stati confrontati con due modelli classici per il contesto di classificazione: il modello logistico e la *random forest* (Breiman 2001). Questo secondo modello, molto sinteticamente, unisce un numero elevato di alberi di classificazione, ognuno dei quali stimato utilizzando dei sottoinsiemi delle variabili esplicative; si rimanda a Hastie et al. (2009, cap. 15) per una trattazione adeguata riguardo questo tipo di strumenti. Queste due tecniche particolari sono state scelte in quanto rappresentative delle due grandi culture dell'apprendimento statistico

(Breiman et al. 2001). Nel primo approccio si assume che i dati siano generati da un modello stocastico noto a meno di alcuni parametri, sulla stima dei quali si concentra l'inferenza; il secondo, invece, consiste in un approccio totalmente algoritmico al problema, secondo il quale il meccanismo generatore dei dati consiste in una "scatola nera", ignota e non interessante, poiché l'obiettivo dei modelli consiste unicamente la corretta previsione della variabile risposta. Senza né entrare nei meriti del dibattito, né definire se e quale dei due approcci sia il "migliore", si tiene conto del fatto che le culture non si collocano in universi paralleli separati, ma siano piuttosto gli estremi di un continuum nel quale ogni problema può trovare una o più collocazioni, eventualmente più o meno coerenti *dato* il problema stesso. Entrambi i modelli sono largamente applicati in contesti reali, ed è quindi sensato utilizzarli per confrontare le tecniche proposte in questo lavoro.

3.3 Applicazione ai dati simulati

Per quanto riguarda i modelli 2.1 e 2.2, è necessario specificare in modo opportuno le distribuzioni a priori. Nello specifico, gli iperparametri delle distribuzioni a priori per β ed α sono stati specificati come segue:

$$\mu_{0_\beta} = \mathbf{0}_p, \Sigma_{0_\beta} = \text{diag}(10, \dots, 10) \quad \mu_{0_\alpha} = 0, \sigma_{0_\alpha} = 10$$

Per quanto riguarda il processo gamma-inverso moltiplicativo, il limite superiore H è stato fissato a 10, mentre $a_1 = 2.5$ e $a_2 = 3.5$; per il modello a coefficienti casuali, inoltre, $\sigma_b^2 = 3$. La scelta degli iperparametri è giustificata come segue: per quanto riguarda i coefficienti β e α le distribuzioni a priori sono centrate in 0 e assegnano una probabilità pari a 95% a valori compresi tra $\{-20, 20\}$; in questo particolare contesto, le distribuzioni a priori possono considerarsi sufficientemente vaghe e coerenti con l'incertezza a priori riguardo queste quantità. Per la scelta dei parametri del processo gamma-inverso moltiplicativo, infine, sono stati selezionati dei valori tali da garantire le proprietà dello stesso, e che allo stesso tempo non penalizzassero in modo eccessivo la dimensionalità della struttura latente, come discusso ad esempio in Durante et al. (2014).

Il modello ad effetti fissi e quello ad effetti casuali sono stati applicati a tutti i diversi scenari, simulando ogni volta catene lunghe 10000 iterazioni, scartando le prime 2500 come *burn-in*. Il costo computazionale degli algoritmi proposti si

è rivelato non banale; la parte più onerosa è quella che riguarda l'inversione di matrici al passo 4 dei gibbs sampling: per ogni iterazione è necessario invertire $(v - 1)$ matrici di dimensione $(v - 1) \times (v - 1)$, operazione decisamente non banale al crescere di v . Nella specificazione ad effetti casuali, inoltre, è necessaria anche l'inversione di $(v - 1)$ matrici di dimensione $(p \times p)$, dove p rappresenta il numero di reti esplicative. In aggiunta, con numerosità campionarie elevate si sono incontrati problemi di instabilità numerica; per ovviare a queste situazioni, le matrici inverse sono state calcolate sfruttando la scomposizione in valori singolari: sebbene sia il metodo numericamente più stabile, è anche il più oneroso dal punto di vista computazionale.

Previa vettorizzazione delle matrici di adiacenza coinvolte, il modello logistico è stato stimato, invece, tramite la routine standard del software statistico R (R Core Team 2016); la *random forest* è stata stimata attraverso il pacchetto R `randomForest` (Liaw e Wiener 2002), calcolando 500 alberi per ogni modello, utilizzando tutte le variabili disponibili per ognuno.

3.4 Risultati

Per quanto riguarda i modelli bayesiani, sono state calcolate la media e deviazione standard delle distribuzioni a posteriori dei parametri β ; per il modello logistico, i parametri presentati sono le ordinarie stime di massima verosimiglianza, gli errori standard delle quali calcolati sfruttando l'approssimazione normale dello stimatore associato. La *random forest*, per definizione, non fornisce esplicitamente questo tipo di stime numeriche.

Per valutare la bontà dei modelli dal punto di vista dell'aderenza ai dati, invece, è stata calcolata l'AUC (*Area Under the ROC Curve*) tra i valori predetti del modello e quelli reali. Come previsioni per i modelli bayesiani è stata utilizzata la media della densità predittiva a posteriori; per il modello logistico, le previsioni ordinarie espresse in termini di probabilità di successo. Per la *random forest*, infine, sono state utilizzate le probabilità predette, calcolate cioè mediando i valori assegnati ad ogni osservazione per tutti gli alberi. Teniamo presente che un confronto opportuno e onesto a questo riguardo prevederebbe la valutazione della performance su un insieme di dati indipendente rispetto a quello sui quali i modelli sono stati stimati; in questa sede, tuttavia, siamo interessati a valutare la capacità

del modello di ricostruire proprietà dei dati generati secondo un meccanismo ben preciso, quindi anche un confronto di questo tipo pare sensato.

I risultati sono riportati nelle tabelle 3.1 e 3.2.

Come è lecito attendersi, al crescere della numerosità campionaria migliora anche l'accuratezza delle stime, valutata tramite la riduzione dei rispettivi errori standard; a parità di condizione, inoltre, il modello bayesiano presenta stime più variabili rispetto al modello logistico, di fatto poiché ogni distribuzione a posteriori tiene conto dell'incertezza derivante dalle altre e dal procedimento di stima.

Il modello logistico e la *random forest* registrano performance soddisfacenti solo negli scenari più semplici, ossia quelli nei quali le osservazioni sono tra loro indipendenti; in questi casi, i modelli sono anche più efficienti rispetto a quelli proposti. Negli scenari in cui i dati iniziano ad assumere caratteristiche di rete più complesse, la performance di questi strumenti cala vertiginosamente, mentre i modelli proposti riescono a trattare agevolmente con questo tipo di strutture, anche quando sono le stesse variabili esplicative a presentare strutture complicate. Grazie alla flessibilità della struttura latente non parametrica, i modelli sono in grado di cogliere le dipendenze introdotte nella rete risposta, permettendo un'inferenza corretta riguardo i parametri β . Le performance del modello ad effetti fissi e di quello ad effetti casuali sono tra loro molto simili; per quanto riguarda il secondo, tendenzialmente presenta stime con errori standard più alti, ma una capacità ad adattarsi in modo più soddisfacente quando anche le reti esplicative sono complesse.

Per concludere, i due approcci "classici" di analisi supervisionata non sembrano particolarmente adeguati in presenza di classificazione per reti. Nel momento in cui le matrici di adiacenza vengono vettorizzate e trattate come osservazioni indipendenti, le strutture di dipendenza derivanti dall'appartenenza a specifiche righe e colonne delle matrici originali vengono distrutte; a causa di questa procedura, sia l'inferenza su β che le previsioni siano decisamente scarse e fuorvianti. Tra i modelli proposti, invece, in questi scenari non sembra ci siano evidenze per propendere verso uno solo in particolare.

Tabella 3.1: Risultati dello studio di simulazione.

	v	Effetti Fissi				Effetti Casuali				GLM Binomiale			
		Media β_1	se β_1	Media β_2	se β_2	Media β_1	se β_1	Media β_2	se β_2	$\hat{\beta}_1$	se $\hat{\beta}_1$	$\hat{\beta}_2$	se $\hat{\beta}_2$
Scenario 1	20	-3,162	0.542	1,330	0.307	-2,595	1.065	1,383	0.866	-3.176	0.565	1.172	0.244
	50	-1,649	0.188	0,831	0.112	-3,646	0.593	1,454	0.572	-1.905	0.172	0.934	0.089
Scenario 2	20	-1,699	0.555	1,388	0.306	-1,978	0.687	1,752	0.531	-1.799	0.487	0.843	0.213
	50	-2,074	0.306	0,985	0.148	-2,627	0.426	1,370	0.256	-1.697	0.144	0.872	0.121
Scenario 3	20	-2,687	0.795	1,234	0.367	-2,829	1.130	1,288	0.909	-1.902	0.618	0.941	0.205
	50	-1,720	0.267	1,031	0.128	-2,454	0.634	1,177	0.592	-1.436	0.208	0.838	0.081
Scenario 4	20	-2,692	0.548	0,940	0.356	-2,754	0.966	0,972	0.844	-1.501	0.372	1.056	0.217
	50	-2,163	0.435	0,821	0.293	-2,252	0.841	1,079	0.670	-0.365	0.149	0.026	0.078
Scenario 5	20	-2,067	0.549	1,061	0.374	-2,337	0.930	0,613	0.870	-1.225	0.394	1.007	0.224
	50	-1,880	0.153	0,410	0.133	-1,874	0.996	1,324	0.674	-1.461	0.139	0.482	0.110

Tabella 3.2: Risultati dello studio di simulazione. Falsi negativi, falsi positivi, errore di classificazione, auc.

	v	Effetti Fissi				Effetti Casuali				GLM Binomiale				Random Forest			
		TFN	TFP	EC	AUC	TFN	TFP	EC	AUC	TFN	TFP	EC	AUC	TFN	TFP	EC	AUC
Scenario 1	20	0.354	0.293	0.326	0.753	0.524	0.451	0.491	0.526	0.296	0.228	0.263	0.789	0.296	0.228	0.263	0.712
	50	0.296	0.228	0.263	0.813	0.590	0.467	0.532	0.554	0.369	0.291	0.333	0.706	0.369	0.291	0.333	0.718
Scenario 2	20	0.270	0.356	0.311	0.815	0.255	0.378	0.334	0.805	0.325	0.344	0.326	0.694	0.310	0.344	0.326	0.688
	50	0.174	0.303	0.244	0.839	0.189	0.317	0.265	0.837	0.158	0.532	0.362	0.696	0.158	0.502	0.362	0.737
Scenario 3	20	0.205	0.192	0.200	0.889	0.561	0.183	0.41	0.662	0.282	0.452	0.347	0.645	0.274	0.808	0.479	0.594
	50	0.241	0.240	0.241	0.844	0.256	0.329	0.284	0.749	0.324	0.444	0.372	0.628	0.060	0.835	0.371	0.666
Scenario 4	20	0.184	0.207	0.195	0.903	0.097	0.013	0.063	0.99	0.345	0.338	0.342	0.525	0.106	0.597	0.305	0.548
	50	0.142	0.195	0.163	0.922	0.007	0.003	0.005	1.00	0.161	0.774	0.469	0.528	0.161	0.774	0.469	0.594
Scenario 5	20	0.564	0.463	0.517	0.588	0.126	0.07	0.105	0.978	0.429	0.394	0.416	0.599	0.092	0.746	0.337	0.611
	50	0.176	0.141	0.163	0.925	0.164	0.293	0.233	0.857	0.158	0.523	0.356	0.679	0.158	0.523	0.356	0.623

Capitolo 4

Applicazione ai dati reali

I modelli proposti sono stati applicati ai dati riferiti alla rete terroristica di Noordin Mohamad Top.

4.1 Simulazione dalla distribuzione a posteriori

Gli iperparametri delle distribuzioni a priori sono stati fissati agli stessi valori specificati in fase di simulazione, e giustificati nella sezione 3.3. I modelli proposti possono incorporare agevolmente delle iper-distribuzioni a priori anche su questi parametri: per quanto riguarda medie e varianze di β e α è possibile sfruttare delle distribuzioni a priori coniugate (normale, gamma - inversa), in modo tale da ottenere delle distribuzioni *full conditional* anche per queste quantità. Per a_1 e a_2 , invece, non esistono distribuzioni coniugate note, tuttavia è comunque possibile incorporare delle distribuzioni a priori opportune: per simulare i valori pseudo-casuali sarà necessario ricorrere, ad esempio, ad un passo Metropolis-Hasting all'interno del Gibbs sampling. I risultati ottenuti, comunque, si sono rivelati sufficientemente robusti rispetto a specificazioni diverse tra quelle proposte; in particolare, specificando alcuni valori

$$\sigma_{0_\beta} \in \{.1, 50\}, \sigma_{0_\alpha} \in \{.1, 50\}, \quad a_1 \in \{1.1, 5\}, a_2 \in \{2.1, 10\}$$

non si sono registrate variazioni sostanziali dei funzionali d'interesse a posteriori. Come tutti gli algoritmi di tipo iterativo, anche i metodi di tipo MCMC necessitano la specificazione di punti iniziali. A livello teorico, l'inizializzazione dell'algoritmo

non impatta sulla convergenza delle catene simulate, ma solo sulla velocità della stessa; tuttavia, la scelta di punti iniziali particolarmente "sbagliati" richiede un numero di iterazioni proibitivo prima di ottenere risultati adeguati. I metodi più efficienti prevedono di estrarre casualmente valori iniziali da opportune approssimazioni della distribuzione a posteriori; in alternativa, può aver senso generare i punti iniziali dalle distribuzioni a priori (Gelman et al., 2014). Per semplicità operativa è stata scelta questa seconda strada; inizializzando poi gli algoritmi da diverse configurazioni iniziali sufficientemente sparse, le catene sono sempre arrivate a convergenza negli stessi punti. Per le analisi seguenti sono state simulate catene di 25000 iterazioni, scartandone le prime 7500 come *burn-in*; l'analisi del *mixing* e dell'autocorrelazione non hanno fornito motivi per dubitare riguardo la convergenza delle stesse. Per ridurre la mole di dati da conservare, si è ricorso ad un'operazione di *thinning*, salvando un'iterazione ogni 5.

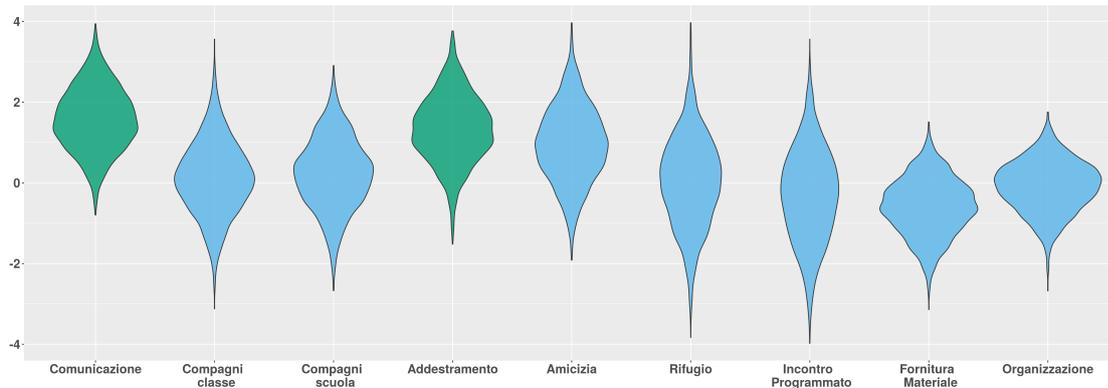
4.1.1 Modello ad effetti fissi: risultati

Si riportano nella tabella 4.1 le stime delle medie, deviazioni standard, e intervalli HPD al 90% delle distribuzioni simulate.

Tabella 4.1: Statistiche per le distribuzioni a posteriori dei parametri indicati. Modello ad effetti fissi con tutte le reti e modello con le reti a densità superiore a 0.01

Rete esplicativa	Modello con tutte le reti			Modello con reti a densità > 0.01		
	Media	Dev. Std.	HPD _{0.9}	Media	Dev. Std.	HPD _{0.9}
(Intercetta)	-5.88	0.60	(-6.75, -4.86)	-4.93	0.49	(-5.67, -4.01)
Comunicazione	1.76	0.89	(0.29, 3.09)	1.58	0.79	(0.32, 2.95)
Compagni di classe	0.15	1.06	(-1.55, 1.97)	0.12	0.92	(-1.34, 1.69)
Compagni scuola	0.22	0.95	(-1.41, 1.74)	0.22	0.88	(-1.14, 1.75)
Parentela	-0.35	1.88	(-3.27, 2.74)	-	-	-
Addestramento	1.65	0.97	(0.15, 3.35)	1.33	0.84	(0.02, 2.75)
Amicizia	1.09	1.07	(-0.75, 2.68)	0.96	0.94	(-0.46, 2.65)
Lavoro	0.70	1.82	(-2.17, 3.78)	-	-	-
Moschea o Sinagoga	-0.14	1.93	(-3.34, 2.91)	-	-	-
Affiliazione religiosa	-0.12	2.02	(-3.42, 3.23)	-	-	-
Incontro in rifugio	0.12	1.32	(-2.21, 2.10)	0.03	1.12	(-1.93, 1.67)
Incontro Programmato	-0.46	1.29	(-2.55, 1.75)	-0.32	1.15	(-2.28, 1.43)
Fornitura Materiale	-0.66	0.84	(-2.08, 0.72)	-0.60	0.71	(-1.77, 0.51)
Organizzazione	-0.14	0.72	(-1.27, 1.00)	-0.09	0.64	(-1.11, 0.96)

Figura 4.1: Grafici a violino delle distribuzioni a posteriori di β per il modello ad effetti fissi. In evidenza le distribuzioni per le quali la mediana a posteriori risulta maggiore di 1



Per i coefficienti riferiti alle reti "Lavoro", "Parentela", "Moschea o sinagoga" e "Affiliazione religiosa", le distribuzioni a posteriori risultano sostanzialmente piatte; le reti associate presentano delle densità inferiori a 0.005, quindi è molto difficile stimare queste quantità in modo coerente, soprattutto se il contributo che apportano non è particolarmente importante. Una soluzione per risolvere questo problema potrebbe essere, ad esempio, quella di lavorare su grafi ridotti, ottenuti campionando i nodi in modo tale da garantire una densità maggiore. In alternativa, è possibile utilizzare un approccio più elegante per selezionare le variabili, adattando ad esempio tecniche di regolarizzazione di tipo Lasso, note anche per avere una semplice interpretazione da un punto di vista bayesiano (Tibshirani, 1996). Nello specifico, utilizzando delle distribuzioni a priori per i coefficienti β con picchi in 0 e code pesanti (come ad esempio delle distribuzioni di Laplace), è possibile identificare le variabili realmente utili ai fini della previsione della rete risposta; questo tipo di strategia sarebbe cruciale se la dimensionalità del problema fosse importante, ossia se fossero disponibili un numero proibitivo di reti e l'interesse fosse nell'identificazione di quelle realmente utili. Tuttavia, in questo problema specifico, siamo molto più interessati ad interpretare la direzione e l'intensità del contributo delle diverse reti, più che a selezionare opportunamente le reti esplicative disponibili; per questi motivi, abbiamo ritenuto sufficiente rimuovere le reti con densità inferiore a 0.01, conservando tutte le altre all'interno del modello. È necessario tener presente che questo modo di operare non è, chiaramente, il più corretto dal punto di vista

statistico, tuttavia è da ritenersi adeguato per il particolare contesto applicativo in oggetto.

I risultati aggiornati sono riportati nella seconda parte della tabella 4.1; per valutare graficamente le distribuzioni a posteriori dei coefficienti, nella fig. 4.1 sono riportati dei grafici a violino. Questi grafici corrispondono a delle usuali stime di densità non parametriche di tipo *kernel* per ogni campione, riportate poi specularmente e affiancate orizzontalmente per confrontarle rapidamente; così facendo è possibile valutare diverse proprietà delle densità stimate, come ad esempio la variabilità o l'eventuale presenza di asimmetria e multi-modalità.

Alcune relazioni hanno decisamente un impatto notevole sulla rete relativa alla partecipazione ad un attentato terroristico: la presenza di un relazione di "Comunicazione" o "Addestramento" tra due individui aumenta il *log odd* di co-partecipare ad un attentato terroristico di circa 1.5 volte, ossia più di 4 volte in scala originale.

Le altre distribuzioni presentano degli intervalli di credibilità che comprendono un intorno di 0 con probabilità alta; per questo motivo il loro contributo si può ritenere non significativamente diverso da 0. Anche questo tipo di informazione è molto interessante per gli obiettivi di questo lavoro: informa, ad esempio, che aver soggiornato nel medesimo rifugio durante la latitanza non contribuisce, al netto del contributo derivante dalle altre reti esplicative, alla propensione a co-partecipare ad un attentato terroristico. Per orientare opportunamente le strategie dei servizi di *intelligence*, può essere cruciale sapere che questo tipo di rapporto, probabilmente molto difficile da codificare, non è particolarmente utile se l'obiettivo è la prevenzione precoce della relazione in oggetto.

Emerge inoltre che le reti riferite ai compagni di scuola, di classe e ad un legame di amicizia non contribuiscono significativamente alla variazione media della propensione a partecipare ad un attentato, al netto dell'effetto delle altre reti. Nonostante sia noto che Noordin reclutasse membri attraverso questi canali "diretti" (ICG, 2006, pag. 2), la sola condivisione di questo tipo di relazione non è sufficiente a spiegare la co-partecipazione ad un attacco. Questo risultato pare sensato poiché, di fatto, quasi tutti i collaboratori sono in qualche modo collegati tra loro da questo tipo di relazioni, mentre i partecipanti alla medesima operazione venivano scelti solo tra i più fedeli o dotati (ICG, 2006, pag. 9).

In fig. 4.2 sono riportati i box-plot delle distribuzioni a posteriori dei diversi λ_h .

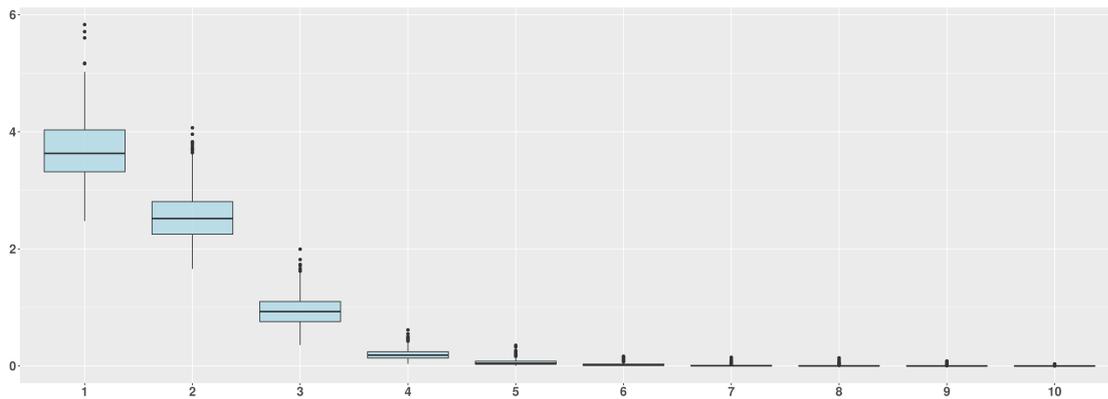


Figura 4.2: Boxplot delle distribuzioni a posteriori di λ . Modello a effetti fissi.

A posteriori, lo spazio latente collassa dalla quarta dimensione in poi, poiché sono sufficienti le dimensioni invocate a scomporre le dipendenze delle reti risposta non spiegate dalla reti esplicative.

4.1.2 Modello ad effetti casuali: risultati

Anche per il modello a coefficienti casuali sono state rimosse le stesse reti della controparte ad effetti fissi; per queste situazioni, infatti, non avrebbe senso formulare un modello a coefficienti casuali; i risultati per i coefficienti β sono riportati nella tabella 4.2 e nella fig. 4.3.

Rispetto alla specificazione precedente, le differenze principali riguardano i coefficienti riferiti alle reti "Organizzazione" e "Fornitura Materiale", le cui medie a posteriori risultano più intense, e le reti "Comunicazione" e "Addestramento", che invece presentano delle medie a posteriori più moderate. È lecito assumere, quindi, che gli effetti casuali additivi sui coefficienti colgano efficacemente delle eterogeneità nel contributo che queste reti apportano alla variazione della rete risposta. Una relazione di "Fornitura Materiale" influisce, in media, negativamente sulla propensione a co-partecipare ad un attentato terroristico; ricordiamo che la rete risposta codifica soltanto la co-partecipazione all'attentato vero e proprio, mentre le operazioni preliminari e successive sono rappresentate in altre reti (Roberts e Everton, 2011). In ogni operazione, alcuni membri avevano il compito di provvedere alla ricerca di luoghi sicuri in cui incontrarsi dopo un attacco, a fornire i mezzi per la

Tabella 4.2: Statistiche per le distribuzioni a posteriori dei parametri indicati. Modello ad effetti casuali con le reti a densità superiore a 0.01.

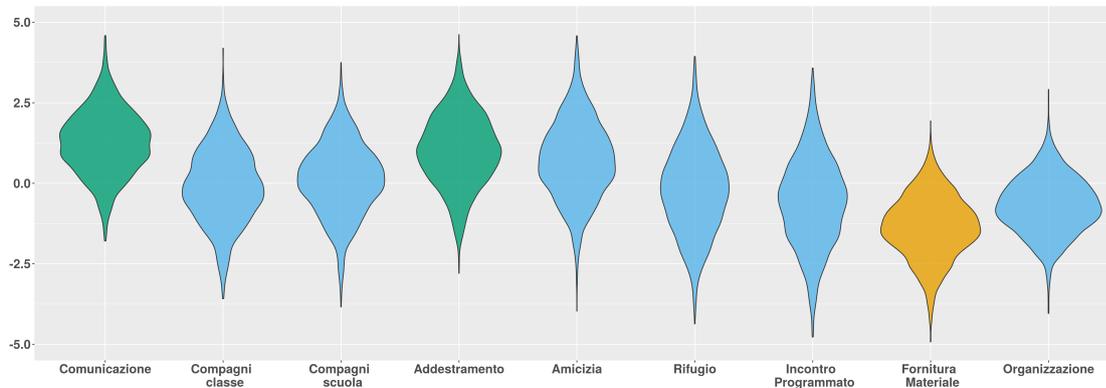
Variabile	Media	Dev. Std.	HPD _{0.9}
(Intercetta)	-5.77	0.77	(-6.92, -4.54)
Comunicazione	1.26	1.02	(-0.72, 3.24)
Compagni di classe	-0.02	1.20	(-1.93, 2.73)
Compagni di scuola	0.19	1.17	(-2.31, 2.25)
Addestramento	0.92	1.08	(-1.56, 2.76)
Amicizia	0.74	1.16	(-1.41, 3.05)
Incontro in rifugio	-0.26	1.37	(-2.63, 2.49)
Incontro Programmato	-0.52	1.38	(-3.15, 2.17)
Fornitura Materiale	-1.17	0.99	(-3.06, 0.66)
Organizzazione	-0.65	0.90	(-2.61, 0.95)

fuga e supporto logistico in generale; la rete in oggetto codifica esattamente questo tipo di relazione. Per questo motivo, se un terrorista ha offerto supporto logistico ad un altro, ha meno propensione a compiere l'attacco vero e proprio, poiché di fatto assegnato ad altri ruoli nella medesima operazione.

La fig. 4.4 rappresenta i boxplot dei coefficienti casuali relativi alle variabili che presentavano situazioni più eterogenee ed interessanti; le informazioni che è possibile trarre non sembrano banali, quindi ora verranno descritte. L'effetto medio della rete "Organizzazione", pari a -0.65 , comporta che tendenzialmente i soggetti precedentemente appartenenti alla stessa organizzazione abbiano una propensione minore, rispetto alla media, a compiere un attentato insieme. Questo risultato è coerente con le dinamiche ricostruite dalle forze dell'ordine, in base alle quali i membri fedeli a Noordin si separarono dalle varie rispettive organizzazioni per formare la fazione estremista da lui capeggiata. Quando, in un secondo momento, alcuni ex-compagni si unirono alla rete criminale, divennero membri con restrizioni; Noordin e i fedeli li trattarono sempre con diffidenza per non averli seguiti da subito (ICG, 2006), e quindi in un certo senso si giustifica il fatto che fossero meno propensi a compiere con loro degli attentati.

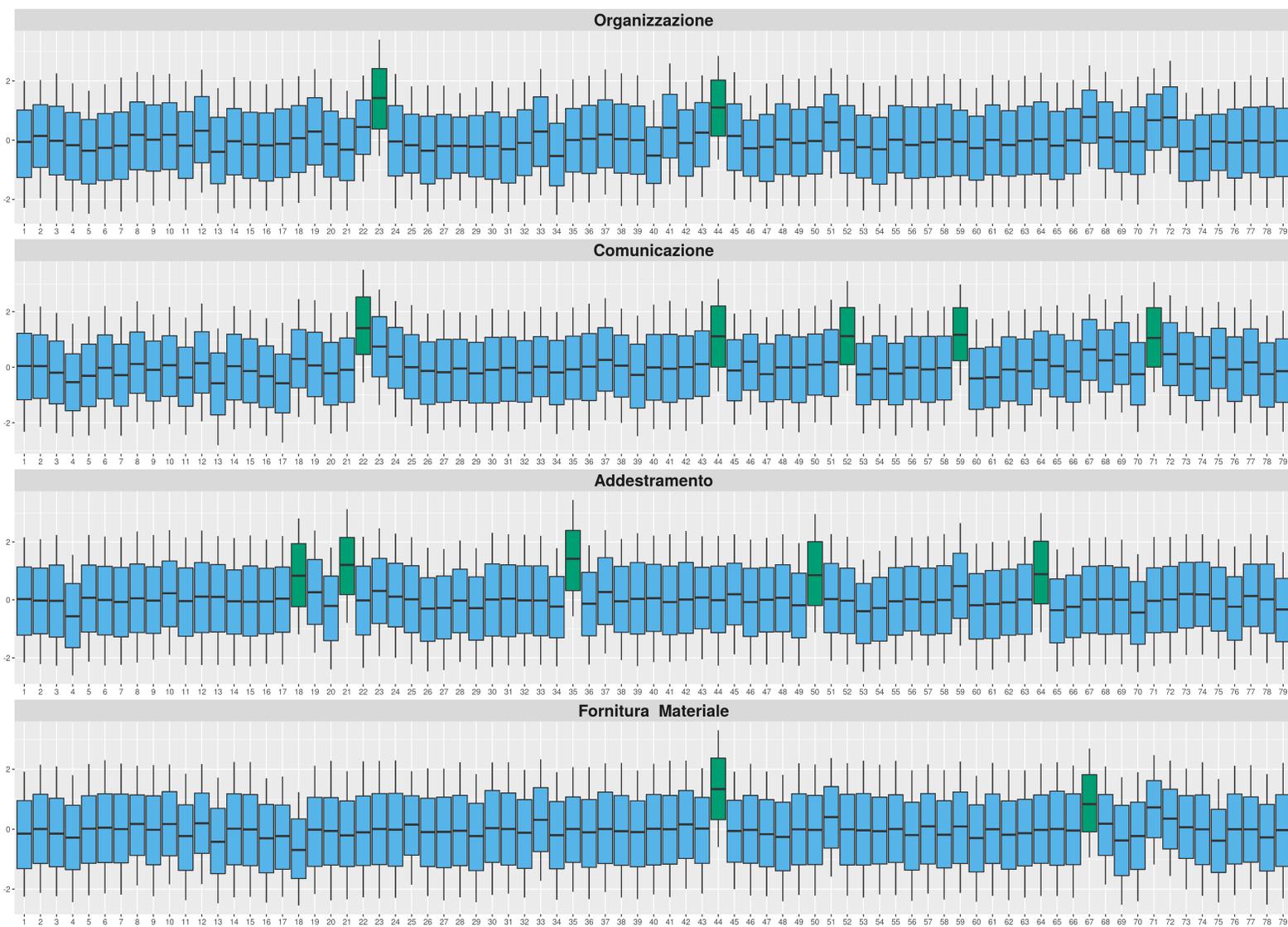
Si noti, tuttavia, come i coefficienti casuali di questa rete riferiti al nodo 23, tale "Azhari Husin" e il nodo 44, "Ismail", presentino chiare particolarità rispetto agli

Figura 4.3: Modello a effetti casuali: grafici a violino delle distribuzioni a posteriori di β . I colori distinguono le distribuzioni sulla base del valore della mediana a posteriori: maggiore di 1 (verde), compreso tra -1 ed 1 (azzurro), minore di -1 (arancione).



altri (fig. 4.4, primo grafico). Se un generico soggetto avesse fatto parte della medesima associazione di uno dei criminali succitati, allora la direzione della relazione sarebbe cambiata di segno, e la propensione a compiere un attentato in collaborazione sarebbe diventata *maggiore* rispetto a quella media (in realtà, per valutare l'effetto complessivo è necessario sommare anche l'effetto casuale riferito a questo individuo generico; tuttavia, il coefficiente associato ai due individui in oggetto è di un'intensità tale da rendere gli altri contributi irrilevanti). Secondo il modello stimato, quindi, questi due soggetti hanno un ruolo chiave nella scelta di complici per gli attentati; in accordo con questa ipotesi, si noti come Ismail presenti coefficienti particolari anche nella rete "Comunicazione" e "Fornitura Materiale". I resoconti degli interrogatori ai terroristi catturati dalle autorità sono in linea con queste congetture, confermando il ruolo centrale di queste individui soprattutto durante le prime operazioni del gruppo (ICG, 2006). Considerazioni analoghe possono essere fatte per tutti i nodi che presentano effetti particolari, che di fatto corrispondono a terroristi che hanno ricoperto ruoli particolari in quei determinati rapporti; per concludere, si noti ad esempio che per quanto riguarda la rete "Comunicazione" uno dei nodi con il coefficiente casuale più incisivo è il numero 59, ossia Noordin stesso.

Figura 4.4: Boxplot delle distribuzioni a posteriori dei coefficienti casuali per le variabili indicate. In riga, da 1 a 79, i nodi della rete. I colori distinguono le distribuzioni sulla base del valore della mediana a posteriori (in verde se maggiore di 1)



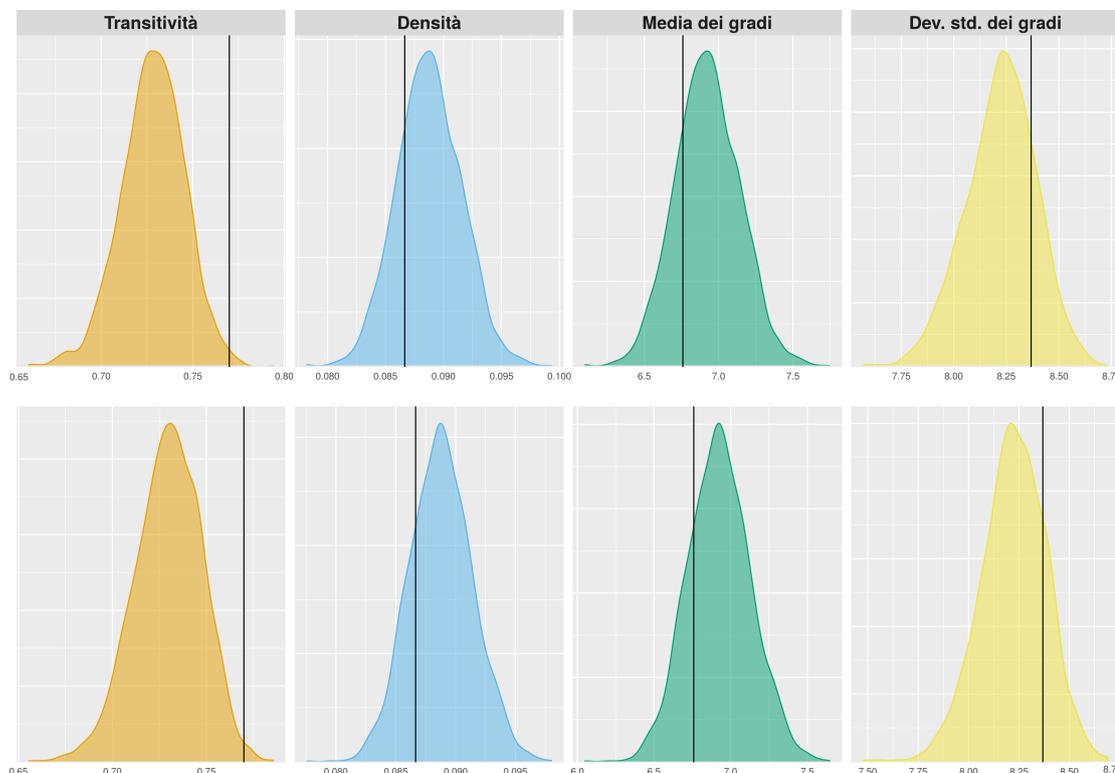
4.2 Confronti

Per valutare la bontà dei modelli, sono state applicate due strategie differenti; per una trattazione completa riguardo gli strumenti diagnostici e l'importanza di queste tecniche entro un'impostazione bayesiana si rimanda a [Gelman et al. \(2014, capitoli 6 e 7\)](#).

Il primo approccio che si presenta tenta di valutare l'accuratezza dei modelli rispetto ai dati dai quali sono stati stimati. Nel problema in oggetto, i campioni provenienti dalla densità predittiva a posteriori sono a loro volta delle reti, ed è quindi sensato un utilizzo di metriche tipiche di questo contesto per valutarne l'adeguatezza. Nello specifico, sia $T(\cdot)$ una generica statistica scalare di rete, come ad esempio quelle introdotte nella sezione 1.2; siano quindi $T(\mathbf{Y})$ il valore della statistica nel campione osservato e $T(\mathbf{Y}^{rep})$ la distribuzione della stessa calcolata rispetto alla densità predittiva a posteriori. Salvo semplici casi, tale distribuzione non è ricavabile analiticamente, tuttavia si può ricorrere ad approssimazioni Monte Carlo sfruttando campioni provenienti da \mathbf{Y}^{rep} ; grandi discrepanze tra le quantità in oggetto sono da interpretare come segnali di una scarsa capacità del modello di riprodurre le caratteristiche sintetizzate dalla statistica stessa. Per i due modelli, sono state valutate le distribuzioni a posteriori di transitività, densità, media e deviazione standard dei gradi dei nodi delle reti simulate; i risultati sono riportati in fig. 4.5. Non si notano forti segnali in direzione di un cattivo adattamento del modello ai dati osservati. L'unica caratteristica che sembra non essere colta in modo adeguato dal modello è la transitività, comunque riportata per correttezza.

L'obiettivo del secondo approccio è, invece, quello di valutare la capacità del modello nella previsione di osservazioni future, ossia non appartenenti al particolare campione osservato; a questo scopo è stata adattata una tecnica utilizzata, ad esempio, da [Hoff \(2008, sez. 3.2\)](#). Nello specifico, gli indici $\{i, j\}$ relativi agli elementi appartenenti alla triangolare inferiore della matrice \mathbf{Y} sono stati divisi casualmente in k parti di egual dimensione, definite s_1, \dots, s_k . Per $l = 1, \dots, k$, si considerano i valori $y_{ij} : \{i, j\} \in s_l$ come osservazioni mancanti, si stima il modello utilizzando tutti i dati a disposizione, e si confrontano, tramite un qualche indicatore di performance, delle opportune previsioni $\hat{p}_{ij} = \mathbb{P}[y_{ij} = 1 | -] : \{i, j\} \in s_l$ con i veri valori rimossi. Il procedimento viene ripetuto per k volte, fissando ad ogni ciclo un *fold* differente come mancante; i k diversi valori della statistica di

Figura 4.5: Densità predittive a posteriori delle statistiche indicate. In evidenza il valore della statistica nel campione osservato. I grafici nella prima riga sono riferiti al modello ad effetti fissi; nella seconda, a quello ad effetti casuali.



bontà di adattamento vengono infine riassunti tramite, ad esempio, la media aritmetica. Gli algoritmi di stima non permettono la gestione diretta di osservazioni mancanti, tuttavia ai passi 6 e 7 dei rispettivi Gibbs sampling è possibile sostituire iterativamente a questi valori stime coerenti, simulandole dalla rispettiva densità predittiva a posteriori.

I risultati di questo approccio, con $k = 4$, sono riportati nella tabella 4.3. Come stima a posteriori per \hat{p}_{ij} è stata utilizzata la media della densità predittiva a posteriori per y_{ij} ; data la natura della variabile risposta, questi valori sono compresi tra 0 ed 1, e quindi interpretabili come la propensione ad osservare una arco in quella specifica relazione della rete risposta. In ambito di analisi supervisionata, è sensato fornire anche previsioni di tipo dicotomico, in modo tale da poter classificare chiaramente le osservazioni come eventuali "successi" (Azzalini e Scarpa,

2012); per fare questo, la media a posteriori è stata trasformata in una variabile dicotomica, pari ad 1 se i valori sono maggiori della densità della rete risposta, e 0 altrimenti.

Come misura della bontà di adattamento, sono stati utilizzati la proporzione di falsi positivi, di falsi negativi, l'errore globale di classificazione e l'*auc*, che ricordiamo essere indipendente dalla particolare soglia utilizzata.

Tabella 4.3: Risultati della bontà di adattamento. Tasso di falsi positivi (TFP), tasso di falsi negativi (TFN), errore di classificazione complessivo (EC), *auc*.

	TFN	TFP	EC	AUC
Modello ad effetti fissi	0.0045	0.0319	0.0272	0.99871
Modello ad effetti casuali	0.0073	0.0471	0.0295	0.99875

In questo problema specifico, un valore cruciale è la proporzione di *falsi negativi*, ossia il rapporto tra le osservazioni classificate come insuccessi, che in realtà non lo sono (falsi negativi), e il numero totale di successi osservati; questi errori rappresentano un costo oneroso, poiché indicano la proporzione di collaborazioni ad attentati terroristici che i modelli non sono in grado di identificare come tali. Idealmente, è desiderabile che questo valore sia uguale a 0; nei modelli stimati, assume valori molto piccoli.

Si supponga ora che sia un'intera riga della matrice di adiacenza \mathbf{Y} ad essere mancante. Questa situazione simula il caso in cui si volessero fare previsioni sul comportamento di un nuovo soggetto, per il quale sono disponibili solo informazioni relative alle reti esplicative, e si volesse fare inferenza sulla sua propensione a compiere un attentato con qualcuno dei soggetti presenti. Per i modelli proposti è relativamente semplice fornire delle previsioni per questi soggetti, di fatto in modo analogo a quanto svolto nei confronti precedenti. Dopo avere ottenuto previsioni opportune per la riga rimossa, è possibile confrontare i valori previsti con quelli reali per valutare ulteriormente la bontà dei modelli. Si presentano nella tabella 4.4 i risultati ottenuti rimuovendo per ogni modello il nodo indicato nella prima colonna.

In questo scenario, le bontà di adattamento dei modelli proposti risultano, tendenzialmente, peggiori rispetto ai confronti precedenti; queste differenze, tuttavia, non sono sorprendenti. Nello scenario precedente, per alcuni soggetti - nel peggiore dei casi - erano disponibili solo delle osservazioni parziali, sulle quali i modelli pote-

Tabella 4.4: Risultati della bontà di adattamento dei modelli, rimuovendo le osservazioni riferite al nodo indicato. Si riportano il tasso di falsi negativi (TFN), il tasso di falsi positivi (TFP), l'errore di classificazione (EC) e l'auc.

Nodo rimosso	Modello a effetti fissi				Modello a effetti casuali			
	TFN	TFP	EC	AUC	TFN	TFP	EC	AUC
Abu Fida	0.00	0.31	0.24	0.99	0.00	0.08	0.06	0.99
Baharudin Soleh	0.00	0.25	0.18	0.96	0.09	0.18	0.15	0.96
Irun Hidayat	0.00	0.08	0.06	0.99	0.00	0.15	0.12	0.99
Urwah (Pranoto)	0.29	0.16	0.19	0.89	0.00	0.20	0.15	0.91
Son Hadi	0.00	0.05	0.04	0.99	0.00	0.18	0.14	0.99
Misno	0.29	0.34	0.33	0.62	0.25	0.32	0.31	0.69
Azhari Husin	0.16	0.00	0.08	0.98	0.16	0.00	0.08	0.98
Noordin Mohammed Top	0.03	0.12	0.09	0.98	0.13	0.15	0.14	0.98
Anif Solchanudin	0.14	0.34	0.32	0.82	0.10	0.34	0.32	0.83
Subur Sugiarto	0.43	0.32	0.33	0.65	0.21	0.30	0.36	0.72
Media	0.13	0.19	0.19	0.89	0.09	0.18	0.18	0.91

vano fare affidamento per ricostruire quelle ignote, ferme restando le informazioni provenienti dagli altri nodi. In questo secondo scenario, invece, i dati mancanti non sono più disposti in modo casuale, ma fortemente sistematico; per i soggetti rimossi, tutte le osservazioni della rete risposta devono essere sostituite con delle stime, quindi tendenzialmente è lecito registrare *performance* peggiori. Sempre per queste ragioni, probabilmente, il modello ad effetti casuali ottiene un vantaggio in questo scenario: grazie alla specificazione particolare, riesce in un "prestito" di informazioni tra i diversi nodi, che concede stime appropriate anche per i soggetti mancanti.

4.3 Riflessioni finali

Per concludere, gli approcci proposti suggeriscono diverse strategie per affrontare le minacce terroristiche. Come suggerito dagli esperti [Roberts e Everton \(2011\)](#) e richiamato nel capitolo 1 si conferma che approcci di tipo "non cinetico", come quello presentato, possono essere fondamentali nell'individuazione precoce della co-partecipazione ad un attentato terroristico; nello specifico, utilizzando infor-

mazioni riferite a terroristi indagati, i modelli proposti in questo lavoro sono in grado di fornire previsioni affidabili sul possibile comportamento di soggetti per i quali sono disponibili solo delle caratteristiche esplicative (sotto forma di rete). E' possibile quantificarne la propensione a co-partecipare ad un attacco con ognuno dei soggetti indagati, tenendo conto dell'incertezza derivante dal processo di stima; i servizi di *intelligence* possono utilizzare queste informazioni, ad esempio, per orientare le proprie strategie verso la prevenzione di contatti tra gli individui a propensione più elevata.

L'approccio presentato ha confermato anche il ruolo cruciale dell'identificazione di relazioni di comunicazione tra terroristi, e in buona parte anche il controllo dei campi di addestramento militare di questi gruppi estremisti. Gli altri tipi di relazione, se l'interesse si concentra nella previsione di un attentato terroristico, non si sono rivelati particolarmente utili.

Da un punto di vista statistico, la bontà d'adattamento dei due modelli può considerarsi identica. La decisione di quale delle due strategie sia preferibile è compito dei servizi incaricati, mentre per coerenza espositiva si sono presentati entrambi gli approcci.

Appendice A

Materiale aggiuntivo

A.1 Costruzione della matrice dei coefficienti casuali

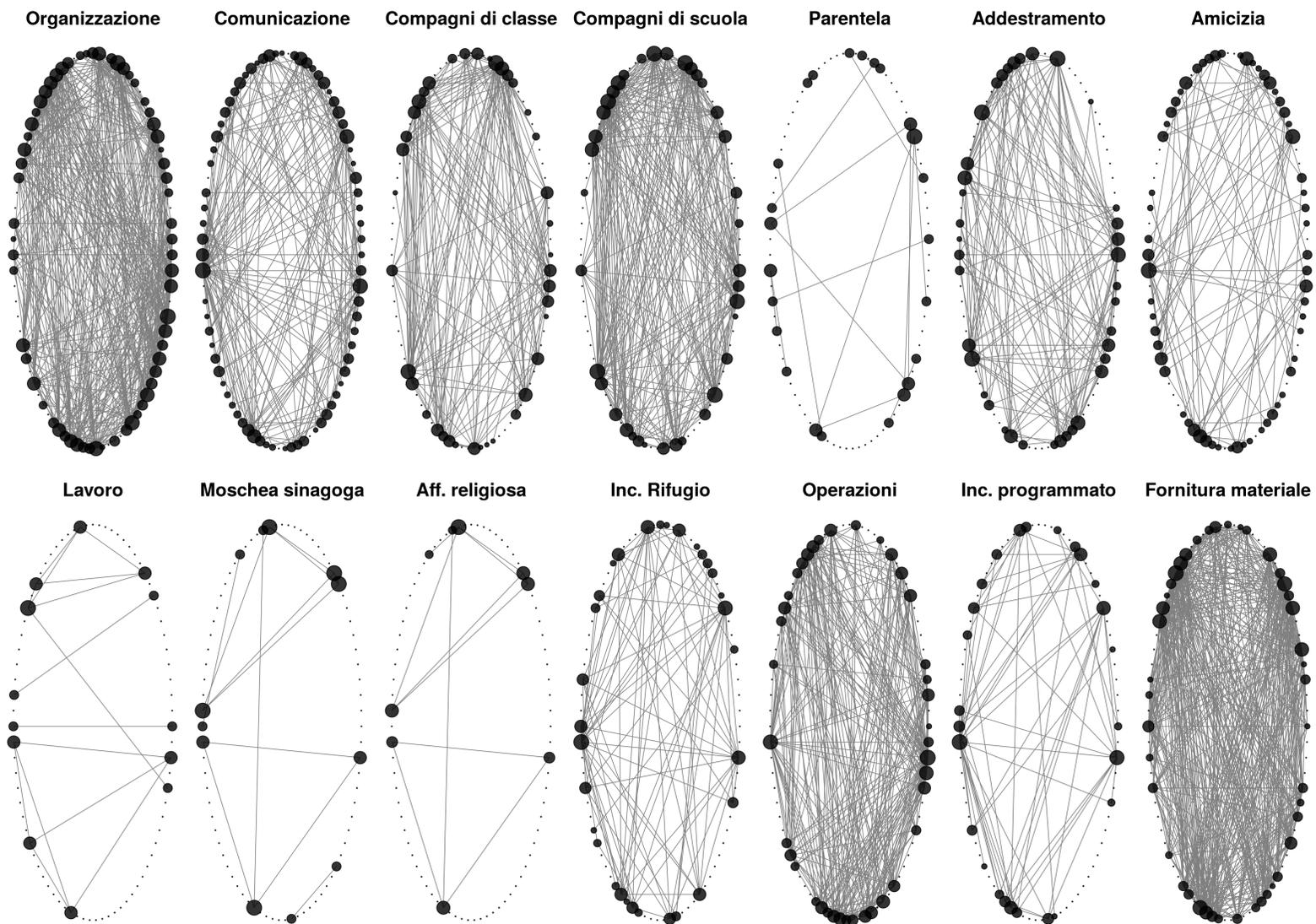
Verrà ora precisata la costruzione "operativa" della formulazione matriciale 2.4. Questi chiarimenti sono utili per la comprensione sia del modello da un punto di vista teorico, che per l'algoritmo di stima.

Siano b_{k1}, \dots, b_{kv} i coefficienti casuali della rete k -esima, $k = 1, \dots, p$, riferiti ai v nodi delle reti in oggetto. Per $k = 1, \dots, p$, sia \mathbf{A}_k la matrice di dimensione $(v \times v)$ ottenuta replicando e quindi incolonnando v volte il vettore (b_{k1}, \dots, b_{kv}) . A questo punto, per ogni k , sia $\mathbf{C}_k = \mathbf{A}_k + \mathbf{A}_k^T$; si noti come le matrici così risultanti siano simmetriche per costruzione, e che il generico elemento di posizione ij sia pari a $b_{ki} + b_{kj}$. Sovrapponendo le matrici così definite si ottiene il tensore \mathbf{B} della 2.4.

Per quanto riguarda l'aggiornamento di questa quantità nel Gibbs sampling, si noti come l'effetto casuale riferito al generico nodo i -esimo compaia, nella formulazione matriciale, solo negli elementi di riga i -esima o di colonna i -esima; è questa particolarità a permettere l'aggiornamento descritto al passo 4.

A.2 Rappresentazione grafica delle reti coinvolte

Figura A.1: Rappresentazione grafica delle reti disponibili. La disposizione dei nodi è la stessa in ogni rete, così da migliorare la lettura; la imensione di ogni nodo è proporzionale al grado relativo.



Bibliografia

- Airoldi, Edoardo M et al. (2008). «Mixed membership stochastic blockmodels». In: *Journal of Machine Learning Research* 9.Sep, pp. 1981–2014.
- Albert, Jim (2009). *Bayesian computation with R*. Springer Science & Business Media.
- Azzalini, Adelchi e Bruno Scarpa (2012). *Data analysis and data mining: An introduction*. OUP USA.
- Bader, Brett W e Tamara G Kolda (2006). «Algorithm 862: MATLAB tensor classes for fast algorithm prototyping». In: *ACM Transactions on Mathematical Software (TOMS)* 32.4, pp. 635–653.
- Barabási, Albert-László e Réka Albert (1999). «Emergence of scaling in random networks». In: *science* 286.5439, pp. 509–512.
- Bauman, Zygmunt (2016). *A Few Comments On The Mis-Imagined War On Terrorism*. <https://www.socialeurope.eu/2016/03/comments-mis-imagined-war-terrorism/>.
- Bhattacharya, A e D B Dunson (2011). «Sparse Bayesian infinite factor models». In: *Biometrika* 98.2, pp. 291–306.
- Breiman, Leo (2001). «Random forests». In: *Machine learning* 45.1, pp. 5–32.
- Breiman, Leo et al. (2001). «Statistical modeling: The two cultures (with comments and a rejoinder by the author)». In: *Statistical Science* 16.3, pp. 199–231.
- Brooks, Steve et al. (2011). *Handbook of Markov Chain Monte Carlo*. CRC press.

- Chen, Hsinchun (2012). *Dark Web: Exploring and Data Mining the Dark Side of the Web*. 1^a ed. Integrated Series in Information Systems 30. Springer-Verlag New York.
- Choi, Hee Min, James P Hobert e Others (2013). «The Polya-Gamma Gibbs sampler for Bayesian logistic regression is uniformly ergodic». In: *Electronic Journal of Statistics* 7, pp. 2054–2064.
- Durante, Daniele, David B Dunson e Joshua T Vogelstein (2014). «Nonparametric Bayes Modeling of Populations of Networks». In: *arXiv preprint arXiv:1406.7851*, pp. 1–29. arXiv: [1406.7851](https://arxiv.org/abs/1406.7851).
- Erdős, Paul e Alfréd Rényi (1959). «On Random Graphs I.» In: *Publicationes Mathematicae (Debrecen)* 6, pp. 290–297.
- Everton, Sean F. (2005). «Tracking, Destabilizing and Disrupting Dark Networks with Social Networks Analysis». In: *Simulation* May, pp. 2069–2072.
- Flynn, Michael T, Matthew F Pottinger e Paul D Batchelor (2010). *Fixing intel: A blueprint for making intelligence relevant in Afghanistan*. Rapp. tecn. DTIC Document.
- Gelman, Andrew e Jennifer Hill (2006). *Data analysis using regression and multi-level/hierarchical models*. Cambridge University Press.
- Gelman, Andrew e Cosma Rohilla Shalizi (2013). «Philosophy and the practice of Bayesian statistics». In: *British Journal of Mathematical and Statistical Psychology* 66.1, pp. 8–38.
- Gelman, Andrew et al. (2014). *Bayesian data analysis*. Vol. 2. Chapman & Hall/CRC Boca Raton, FL, USA.
- Geweke, John e Guofu Zhou (1996). «Measuring the pricing error of the arbitrage pricing theory». In: *Review of Financial Studies* 9.2, pp. 557–587.
- Granovetter, Mark (2005). «The impact of social structure on economic outcomes». In: *The Journal of economic perspectives* 19.1, pp. 33–50.

- Hastie, Trevor, Robert Tibshirani e Jerome Friedman (2009). *The elements of statistical learning*. Vol. 2. New York: Springer.
- Hoff, Peter D (2003). «Bilinear Mixed Effects Models for Dyadic Data». In: *None* 2002, pp. 1–20.
- (2008). «Multiplicative latent factor models for description and prediction of social networks». In: *Computational and Mathematical Organization Theory* 15.4, pp. 261–272.
- Hoff, Peter D et al. (2001). «Latent Space Approaches to Social Network Analysis». In: *Journal of the American Statistical Association* 97, pp. 1090–1098.
- ICG, International Crisis Group (2006). «Terrorism in Indonesia: Noordin's group». In:
- Kolda, Tamara G e Brett W Bader (2009). «Tensor decompositions and applications». In: *SIAM review* 51.3, pp. 455–500.
- Krebs, Valdis E (2002). «Mapping networks of terrorist cells». In: *Connections* 24.3, pp. 43–52.
- Krivitsky, Pavel N et al. (2007). «Representing Degree Distributions, Clustering, and Homophily in Social Networks With Latent Cluster Random Effects Models». In:
- Lee, Namgil e Andrzej Cichocki (2014). «Fundamental tensor operations for large-scale data analysis in tensor train formats». In: *arXiv preprint arXiv:1405.7786*.
- Li, Heng e Eric Loken (2002). «A Unified Theory of Statistical Analysis and Inference for Variance Component Models for Dyadic Data». In: *Statistica Sinica* 12, pp. 519–535.
- Liaw, Andy e Matthew Wiener (2002). «Classification and Regression by random-Forest». In: *R News* 2.3, pp. 18–22. URL: <http://CRAN.R-project.org/doc/Rnews/>.

- Lopes, Hedibert Freitas e Mike West (2004). «Bayesian model assessment in factor analysis». In: *Statistica Sinica*, pp. 41–67.
- McGehee, Ralph W (1983). *Deadly deceits: My 25 years in the CIA*. Open Road Media.
- Meter, Karl M Van (2002). «Terrorists/Liberators: Researching and Dealing with Adversary Social Networks». In: *Social Networks* 24.3, pp. 66–78.
- Nowicki, Krzysztof e Tom A B Snijders (2001). «Estimation and prediction for stochastic blockstructures». In: *Journal of the American Statistical Association* 96.455, pp. 1077–1087.
- Orbanz, Peter (2012). «Lecture Notes on Bayesian Nonparametrics». In: *Journal of Mathematical Psychology* 56, pp. 1–12.
- Pedahzur, Ami e Arie Perliger (2006). «The changing nature of suicide attacks: a social network perspective». In: *Social forces* 84.4, pp. 1987–2008.
- Polson, Nicholas G., James G. Scott e Jesse Windle (2013). «Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables». In: *Journal of the American Statistical Association* 108.504, pp. 1339–1349. arXiv: [arXiv : 1205.0310v3](https://arxiv.org/abs/1205.0310v3).
- Prakasa Rao, Bhagavatula Lakshmi Surya (1987). *Asymptotic theory of statistical inference*. John Wiley & Sons, Inc.
- R Core Team, R (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.
- Roberts, Nancy e Sean F Everton (2011a). «Roberts and Everton Terrorist Data: Noordin Top Terrorist Network (Subset). [Machine-readable data file]». In:
- (2011b). «Strategies for Combating Dark Networks». In: *Journal of Social Structure* 12, p. 2.
- Rodríguez, Josep A e José A Rodríguez (2005). «The March 11 th terrorist network: In its weakness lies its strength». In:

Sampson, Samuel F (1968). *A novitiate in a period of change: An experimental and case study of social relationships*. Cornell University.

Snijders, Tom AB e Krzysztof Nowicki (1997). «Estimation and prediction for stochastic blockmodels for graphs with latent block structure». In: *Journal of classification* 14.1, pp. 75–100.

Tibshirani, Robert (1996). «Regression shrinkage and selection via the lasso». In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288.

Wikipedia (2004). *Noordin Mohammad Top - Wikipedia, The Free Encyclopedia*.