

UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Scuola di Economia e Statistica

Corso di laurea Magistrale in

SCIENZE STATISTICHE ED ECONOMICHE



**UNA GENERALIZZAZIONE DELLA DISTRIBUZIONE DI
ERDŐS-RÉNYI CENTRATA PER L'ANALISI DI DATI NETWORK**

Relatore: Prof. Bernardo Nipoti

Correlatore: Prof. Tommaso Rigon

Tesi di Laurea di:

Sara Capozio

Matr. N. 853600

Anno Accademico 2023/2024

Ringraziamenti

Non avrei mai potuto raggiungere questo traguardo senza avere vicino persone che hanno creduto in me e che profondamente ringrazio.

Al mio relatore, il Professore Bernardo Nipoti per l'immensa disponibilità, la preziosa collaborazione e il costante supporto dimostrati durante lo svolgimento del lavoro.

Ai miei familiari, in particolare a mia mamma, per avermi sempre incoraggiato ad affrontare ogni sfida con determinazione e coraggio e per avermi dato la possibilità di realizzare serenamente i miei obiettivi.

Agli amici di sempre, compagni di avventure e di ricordi preziosi per non avermi mai fatto sentire sola e ai miei compagni di corso con cui ho condiviso grandi gioie e fatiche in questi due anni di crescita.

Infine un ringraziamento speciale a Luca, per esserci sempre stato anche nei momenti più difficili e per avermi aiutato a vedere la vita a colori.

Indice

Introduzione	1
1 Modello di Erdős–Rényi centrato generalizzato	3
1.1 Aspetti preliminari	3
1.2 Introduzione al modello Generalizzato di Erdős–Rényi	5
2 Proprietà del modello GCER	10
2.1 Proprietà relative al parametro di concentrazione	10
2.2 Proprietà relative al parametro di scala	16
3 Inferenza bayesiana per il modello GCER	20
4 Inferenza bayesiana per il modello CER con prior GCER	22
4.1 Distribuzione GCER come prior coniugata in un modello CER	22
4.2 Parametro di scala α non noto	25
4.2.1 Prior con componenti dipendenti	25
4.2.2 Prior con componenti indipendenti	28
5 Analisi discriminante come esempio applicativo	30
5.1 Ipotesi di omoschedasticità	31
5.2 Ipotesi di eteroschedasticità	32
6 Studio di simulazione	34
7 Analisi dei dati Human Brain Networks	39
Conclusioni	49
Bibliografia	51

Introduzione

Le reti multiple rappresentano una tipologia di dati in grado di rappresentare interazioni o connessioni tra elementi di un sistema più o meno complesso. Esse consistono in un insieme di grafi caratterizzati dallo stesso numero di nodi messi in relazione tra loro tramite interazioni. A definire gli elementi distintivi di ciascun grafo sono le connessioni tra i nodi, ovvero la struttura degli archi.

Le reti multiple trovano applicazione in molti ambiti disciplinari, in particolare quello delle neuroscienze. In tale campo, una popolazione di reti può essere usata per rappresentare le interazioni tra varie regioni cerebrali per più individui, la cui analisi permette di identificare la presenza di eventuali disturbi neurologici.

Per modellare questa tipologia di dati sono stati introdotti diversi modelli probabilistici, tra cui il modello di Erdős–Rényi (ER) che si basa sull'idea che ogni connessione tra due nodi distinti sia indipendente e identicamente distribuita come una Bernoulli. Nell'articolo di Lunagómez et al. (2021) è stata introdotta un'estensione del modello ER, ossia il modello Erdős–Rényi centrato (CER). Rispetto al modello ER, il modello CER appare più flessibile in quanto supera l'assunto di identica distribuzione delle connessioni tra coppie di nodi distinti. La probabilità di osservare un arco tra due nodi non è più identica per tutte le coppie ma dipende da due parametri: un parametro di locazione e un parametro di scala. Il parametro di locazione coincide con la moda della distribuzione e rappresenta una misura di centralità, mentre il parametro di scala misura il grado di concentrazione delle osservazioni attorno alla misura di locazione.

Sebbene il modello CER sia caratterizzato da una forma analitica facilmente interpretabile, la sua struttura probabilistica risulta essere piuttosto rigida. Questo può rappresentare un limite nel momento in cui si analizzano strutture d'interazione complesse che richiedono approcci più flessibili.

Nella presente trattazione viene introdotta una generalizzazione del modello di Erdős–Rényi centrato che prende il nome di modello di Erdős–Rényi centrato generalizzato, in forma abbreviata, GCER. Questo nuovo modello presenta la stessa struttura parametrica del modello CER, ma, a differenza di quest'ultimo, il parametro di locazione è rappresentato da un grafo pesato definito su uno spazio più ampio rispetto allo spazio dei grafi binari.

La tesi si pone l'obiettivo di analizzare nel dettaglio il nuovo modello probabilistico GCER

andando a delinearne le proprietà principali, le strategie inferenziali ed eventuali applicazioni.

La prima parte del lavoro si concentra sulla definizione del modello e sulle proprietà che lo caratterizzano. In questa sezione viene esaminata la relazione tra il modello CER e il GCER, evidenziando come alcune proprietà caratteristiche del modello più rigido possono essere generalizzate al caso più flessibile. Questa analisi consente di mettere in luce non solo le somiglianze tra i due approcci ma anche i vantaggi derivanti dall'utilizzo del modello GCER. La trattazione prosegue con l'analisi delle strategie inferenziali bayesiane utilizzate per stimare i parametri del modello di Erdős–Rényi centrato generalizzato, mostrando le limitazioni analitiche e computazionali che questo approccio presenta.

Le problematiche riscontrate utilizzando il modello GCER come funzione di verosimiglianza non si presentano quando questo modello viene utilizzato come distribuzione a priori per il modello CER. Infatti, il modello GCER gode della proprietà di coniugatezza rispetto al modello CER a condizione che il parametro di scala sia noto. Questa proprietà permette di derivare facilmente da un punto di vista analitico la distribuzione a posteriori per il parametro di locazione, semplificando così l'inferenza. Tuttavia, la proprietà di coniugatezza non è rispettata nel caso in cui il parametro di scala non è noto, per tale ragione risulta necessario adottare tecniche algoritmiche simulate, come il metodo di campionamento Gibbs Sampling.

La seconda parte del lavoro si concentra su uno studio simulativo, con lo scopo di mostrare la flessibilità del modello GCER rispetto al modello CER e sull'applicazione delle metodologie inferenziali introdotte al dataset reale 'Human Brain Networks', al fine di modellare le caratteristiche strutturali a livello cerebrale di trenta pazienti.

Lo studio si conclude con un esempio applicativo basato sull'analisi discriminante che porta alla teorizzazione della rispettiva regola decisionale e alla valutazione della bontà classificativa del modello CER sotto ipotesi di eteroschedasticità delle osservazioni.

1. Modello di Erdős–Rényi centrato generalizzato

1.1. Aspetti preliminari

Un grafo binario etichettato non orientato semplice $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ è formato da un insieme di nodi etichettati $\mathcal{V} = \{1, \dots, N\}$ e da un insieme di archi $\mathcal{E} \subseteq \{\mathcal{E} \subset \mathcal{V} : |\mathcal{E}| = 2\}$. Il numero totale di nodi è dunque denotato con $N = |\mathcal{V}|$, mentre

$$\mathcal{G}_{[N]} := \{\mathcal{G} = (\mathcal{V}, \mathcal{E}) : \mathcal{V} = \{1, \dots, N\}\},$$

rappresenta l'insieme dei grafi binari semplici etichettati non orientati con nodi \mathcal{V} , chiamato anche spazio dei grafi. Qualora si prendano in esame soltanto grafi semplici, il numero totale di grafi ottenibili è dato da $|\mathcal{G}_{[N]}| = 2^M$, dove $M = N(N - 1)/2$ rappresenta il numero massimo di connessioni possibili per un grafo con N nodi. Ogni grafo \mathcal{G} può essere rappresentato tramite una matrice di adiacenza $A_{\mathcal{G}}$ di dimensioni $N \times N$ tale che:

$$A_{\mathcal{G}}[i, j] = \begin{cases} 1 & \text{se } [i, j] \in \mathcal{E} \\ 0 & \text{altrimenti} \end{cases}$$

dove $A_{\mathcal{G}}[i, j]$ rappresenta l'entrata nella posizione $[i, j]$ della matrice di adiacenza $A_{\mathcal{G}}$ relativa al grafo \mathcal{G} . Poichè vengono presi in considerazione grafi non orientati, la matrice di adiacenza $A_{\mathcal{G}}$ risulta essere simmetrica.

Per poter definire il modello probabilistico di base su $\mathcal{G}_{[N]}$, che costituisce il fondamento della presente trattazione, è opportuno introdurre il concetto di distanza di Hamming. La distanza di Hamming tra due grafi, definita come:

$$d_{\text{H}}(\mathcal{G}_1, \mathcal{G}_2) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N |A_{\mathcal{G}_1}[i, j] - A_{\mathcal{G}_2}[i, j]|,$$

rappresenta una misura di similarità definita su $\mathcal{G}_{[N]}$ che coincide con il numero di connessioni differenti tra i due grafi \mathcal{G}_1 e \mathcal{G}_2 . Tale quantità è utile a quantificare quanto due grafi differiscono in termini delle loro strutture di connessione. Il modello CER è un modello probabilistico definito in termini di distanza di Hamming. Il modello CER è parametrizzato da un parametro di locazione \mathcal{G}^m definito su $\mathcal{G}_{[N]}$, che coincide con la moda della distribuzione, e un parametro di scala α che misura la variabilità dei grafi attorno al parametro di centralità \mathcal{G}^m . Per definire il parametro di locazione, viene sfruttata la nozione di media di Fréchet, una generalizzazione del concetto di media che si estende al contesto non Euclideo. Nella distribuzione CER, la media di Fréchet \mathcal{G}^m e la moda coincidono. Il modello si basa sull'assunto che ogni arco si distribuisce in modo indipendente, ma non identico, secondo una distribuzione bernoulliana. La probabilità che ci sia un arco tra una coppia di nodi è pari a α o $1 - \alpha$ a seconda che l'arco connetta la stessa coppia di nodi del grafo centrale \mathcal{G}^m . Formalmente, il modello può essere definito specificando la probabilità di osservare un arco tra i nodi $i \neq j$ come:

$$\Pr(A_{\mathcal{G}}[i, j] = A_{\mathcal{G}^m}[i, j]) = 1 - \alpha.$$

Riscrivendo il modello in forma globale, si ottiene una distribuzione discreta p_{CER} su $\mathcal{G}_{[N]}$ definita come:

$$p_{\text{CER}}(\mathcal{G}; \mathcal{G}^m, \alpha) = \alpha^{d_{\text{H}}(\mathcal{G}, \mathcal{G}^m)} (1 - \alpha)^{M - d_{\text{H}}(\mathcal{G}, \mathcal{G}^m)}. \quad (1.1)$$

Per indicare che un grafo aleatorio \mathcal{G} è distribuito secondo la distribuzione in 1.1, si utilizza la notazione $\mathcal{G} \sim \text{CER}(\mathcal{G}^m, \alpha)$.

Nei paragrafi successivi, laddove necessario, si preferirà introdurre la notazione più compatta $p_{\text{CER}}(\cdot)$ per $p_{\text{CER}}(\cdot; \mathcal{G}^m, \alpha)$.

Nonostante la distribuzione (1.1) sia definita per ogni $\alpha \in (0, 1)$, il dominio del parametro di scala è ristretto all'intervallo $(0, 1/2)$ al fine di garantire l'unimodalità della distribuzione. Dalla formula (1.1) si deduce che i grafi più concentrati attorno alla moda \mathcal{G}^m hanno maggior probabilità di essere osservati. In particolare se $d_{\text{H}}(\mathcal{G}_1, \mathcal{G}^m) > d_{\text{H}}(\mathcal{G}_2, \mathcal{G}^m)$, allora $p_{\text{CER}}(\mathcal{G}_2) > p_{\text{CER}}(\mathcal{G}_1)$.

1.2. Introduzione al modello Generalizzato di Erdős–Rényi

Il modello CER può essere visto come una generalizzazione del modello più semplice, il modello di Erdős–Rényi, in cui si ipotizza che ogni connessione si distribuisca in modo indipendente e sia identicamente distribuita secondo una distribuzione bernoulliana. Nel modello CER la probabilità di successo risulta essere funzione di un parametro di locazione e uno di scala. In questo caso cade l'assunto di identica distribuzione degli archi in quanto la probabilità che si presenti un arco dipende dal valore assunto da \mathcal{G}^m in corrispondenza degli stessi nodi.

La tesi considera un'ulteriore generalizzazione del modello CER tramite l'introduzione di una nuova misura di locazione rappresentata da un grafo etichettato non orientato semplice pesato W . Questo nuovo parametro è definito su uno spazio più ampio rispetto allo spazio dei grafi binari, denominato $\mathcal{W}_{[N]}$. Lo spazio $\mathcal{W}_{[N]}$ è formalmente definito come:

$$\mathcal{W}_{[N]} = \{W \in [0, 1]^{N \times N} : W[i, j] = W[j, i] \text{ e } W[i, i] = 0 \text{ per ogni } i, j \in \{1, \dots, N\}\}$$

e rappresenta l'insieme di tutti i grafi pesati etichettati non orientati semplici con N nodi, in cui ogni elemento $W[i, j]$ della matrice W rappresenta il peso dell'arco tra i nodi i e j .

Definizione 1.1. Un grafo aleatorio $\mathcal{G} \in \mathcal{G}_{[N]}$ ha distribuzione di Erdős–Rényi centrata generalizzata (GCER) se per ogni $i \neq j$,

$$\Pr(\mathcal{G}[i, j] = 1; W, \alpha) = \left(1 + \left(\frac{\alpha}{1 - \alpha}\right)^{2W[i, j] - 1}\right)^{-1}. \quad (1.2)$$

Per indicare che un grafo aleatorio \mathcal{G} è distribuito secondo tale modello si utilizza la notazione $\mathcal{G} \sim \text{GCER}(W, \alpha)$.

Come nel caso della distribuzione CER, ogni arco si distribuisce in modo indipendente, ma non identico, secondo una distribuzione bernoulliana con probabilità di successo definita in (1.2), formalizzabile come:

$$\mathcal{G}[i, j] \sim \text{Bernoulli} \left(\left(1 + \left(\frac{\alpha}{1 - \alpha} \right)^{2W[i, j] - 1} \right)^{-1} \right).$$

Il valore atteso dell'arco tra la coppia di nodi i e j coincide con la probabilità di osservare una connessione tra di essi. Pertanto, si ha:

$$\mathbb{E}[\mathcal{G}[i, j]] = \left(1 + \left(\frac{\alpha}{1 - \alpha} \right)^{2W[i, j] - 1} \right)^{-1}. \quad (1.3)$$

Il modello probabilistico GCER risulta essere più flessibile del modello CER in cui, la probabilità di osservare una connessione tra due nodi può assumere solo due valori, in base alla presenza di un arco che collega gli stessi nodi in \mathcal{G}^m . Nel nuovo modello, l'introduzione di un grafo pesato come matrice di locazione, comporta il superamento della struttura rigida del CER permettendo una maggior flessibilità nel modellare le probabilità di connessione. Questa nuova distribuzione può essere vista come un compromesso tra il modello CER e il modello più generale in cui ogni arco si distribuisce in modo indipendente con una probabilità di successo definita da un parametro specifico per ogni coppia di nodi. Quest'ultimo caso, pur rappresentando il massimo livello di flessibilità, presenta come svantaggio la perdita di interpretabilità dei parametri e la conveniente distinzione tra parametro di locazione e parametro di scala che caratterizza il modello CER. Al contrario il modello GCER preserva questa proprietà grazie alla sua struttura parametrica. Nel modello GCER, infatti, la probabilità di successo è funzione di due parametri distinti ciascuno con un'interpretazione ben specifica: un parametro di locazione W e un parametro di scala α .

Il parametro W è un parametro di locazione definito sullo spazio $\mathcal{W}_{[N]}$, rappresentato da una matrice simmetrica di pesi di dimensioni $N \times N$. Dall'equazione (1.3) risulta possibile scrivere $W[i, j]$ in funzione del parametro α e del valore atteso $\mathbb{E}[i, j]$, ottenendo:

$$W[i, j] = \frac{1}{2} \left(\frac{\log \left(\frac{1 - \mathbb{E}[\mathcal{G}[i, j]]}{\mathbb{E}[\mathcal{G}[i, j]]} \right)}{\log \left(\frac{\alpha}{1 - \alpha} \right)} + 1 \right).$$

Ogni elemento $W[i, j] \in [0, 1]$ misura l'intensità della connessione tra la coppia di nodi i e j , influenzando la probabilità di osservare un arco tra di essi. Dalla Figura 1.1 si osserva che quanto più $W[i, j]$ è prossimo al suo valore massimo, tanto maggiore è la forza della connessione tra la coppia di nodi i e j e conseguentemente la probabilità di osservare un arco tra di essi. Al contrario, quanto più $W[i, j]$ è vicino allo zero, minore sarà la forza della connessione.

Il parametro α , come nel caso della distribuzione CER, è un parametro di scala che assume valori nell'intervallo $(0, 1/2)$ e quantifica la variabilità del grafo attorno al parametro di localizzazione W . Come si evince in Figura 1.1 per valori bassi di α , la concentrazione attorno a W è massima, al contrario per valori elevati di α , la probabilità di osservare un arco tra i nodi i e j si concentra attorno a $1/2$ per qualunque valore di $W[i, j]$, indicando una maggior variabilità nella struttura del grafo.

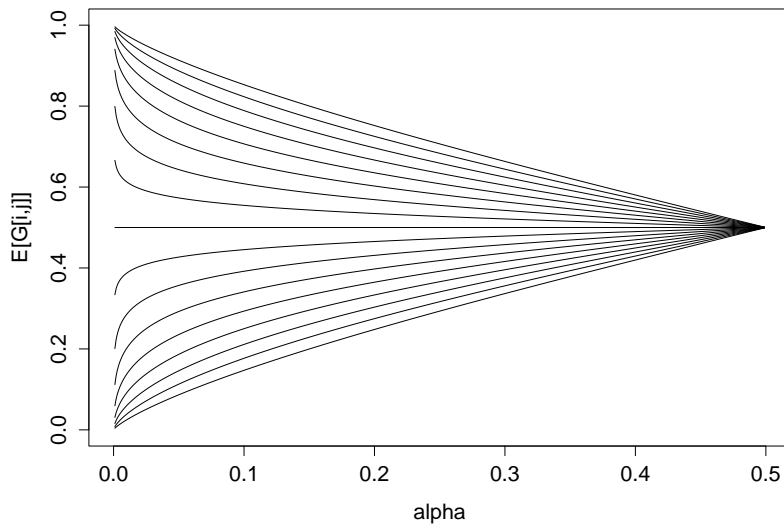


Figura 1.1: Valore atteso $\mathbb{E}[\mathcal{G}[i, j]]$ come funzione di $\alpha \in (0, 1/2)$ per diversi valori di $W[i, j] \in \{0.1, \dots, 0.9\}$ (dal basso verso l'alto).

Questa nuova distribuzione prende il nome di distribuzione di Erdős–Rényi centrata generalizzata in quanto comprende la distribuzione CER come caso particolare. Infatti, fissando $W = \mathcal{G}^m \in \mathcal{G}_{[N]}$:

$$\Pr(\mathcal{G}[i, j] = 1; \mathcal{G}^m, \alpha) = \begin{cases} 1 - \alpha & \text{se } \mathcal{G}^m[i, j] = 1 \\ \alpha & \text{se } \mathcal{G}^m[i, j] = 0 \end{cases}$$

Se la matrice di locazione W appartiene allo spazio $\mathcal{G}_{[N]}$, allora la distribuzione GCER si riduce alla distribuzione CER e il parametro di locazione W dunque coincide con la moda della distribuzione.

Per comprendere al meglio l'andamento della probabilità di osservare una connessione tra la coppia di nodi i e j al variare dei parametri α e W , definita nell'equazione (1.2), si analizza come la distribuzione di un grafo aleatorio $\mathcal{G} \sim \text{GCER}(W, \alpha)$ si comporta nei casi limite:

- Per $\alpha \rightarrow 0$ e $W[i, j] > 1/2$, $\Pr(\mathcal{G}[i, j] = 1) = 1$. La presenza di un arco tra i nodi i e j del grafo \mathcal{G} , quando α tende a zero, è certa se $W[i, j] > 1/2$. Viceversa, considerando valori di α molto piccoli e pesi inferiori a $1/2$, la probabilità di osservare una connessione tra la coppia di nodi i e j risulta essere nulla. In questi due scenari la distribuzione di probabilità è degenerare in quanto, sulla base dei valori assunti da $W[i, j]$ la probabilità di osservare o non osservare un arco tra le coppie di nodi i e j raggiunge il suo valore limite.
- Per $\alpha \in (0, 1/2)$, $W[i, j] = 1/2$ e per $\alpha \rightarrow 1/2$, $W[i, j] \in [0, 1]$, $\Pr(\mathcal{G}[i, j] = 1) = 1/2$. Questi due scenari rappresentano le situazioni di massima incertezza in cui $\mathcal{G}[i, j] \sim \text{Bernoulli}(1/2)$. In questi casi la distribuzione GCER si riduce alla distribuzione classica di Erdős–Rényi.

Per $g \in \{0, 1\}$, è possibile riscrivere la distribuzione GCER in modo esteso:

$$\begin{aligned} \Pr(\mathcal{G}[i, j] = g; W, \alpha) &= \left[1 + \left(\frac{\alpha}{1-\alpha} \right)^{2W[i, j]-1} \right]^{-g} \left[1 - \left(1 + \left(\frac{\alpha}{1-\alpha} \right)^{2W[i, j]-1} \right)^{-1} \right]^{1-g} \\ &= \left[1 + \left(\frac{\alpha}{1-\alpha} \right)^{2W[i, j]-1} \right]^{-g} \left[\frac{\left(\frac{\alpha}{1-\alpha} \right)^{2W[i, j]-1}}{1 + \left(\frac{\alpha}{1-\alpha} \right)^{2W[i, j]-1}} \right]^{1-g} \\ &= \left(\frac{\alpha}{1-\alpha} \right)^{g(1-2W[i, j])} \left[\frac{\left(\frac{\alpha}{1-\alpha} \right)^{2W[i, j]-1}}{1 + \left(\frac{\alpha}{1-\alpha} \right)^{2W[i, j]-1}} \right]. \end{aligned}$$

Questo porta a fornire una definizione della distribuzione GCER equivalente a quella data in Definizione 1.1.

Definizione 1.2. Un grafo aleatorio $\mathcal{G} \in \mathcal{G}_{[N]}$ ha distribuzione di Erdős–Rényi centrata generalizzata (GCER) se

$$p(\mathcal{G}; W, \alpha) = \left(\frac{\alpha}{1 - \alpha} \right)^{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}[i,j](1-2W[i,j])} \frac{1}{\prod_{i=1}^{N-1} \prod_{j=i+1}^N \left(1 + \left(\frac{\alpha}{1-\alpha} \right)^{1-2W[i,j]} \right)}$$

o equivalentemente, a meno di una costante di normalizzazione,

$$p(\mathcal{G}; W, \alpha) \propto \left(\frac{\alpha}{1 - \alpha} \right)^{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}[i,j](1-2W[i,j])} .$$

Risulta importante sottolineare che, sebbene il modello sia ben definito per $\alpha \in (0, 1)$, il dominio del parametro viene ristretto all'intervallo $(0, 1/2)$ al fine di garantire alcune proprietà della distribuzione, che verranno enunciate nel Capitolo 2, e soprattutto preservare l'interpretabilità del modello. Se si considerassero valori di $\alpha \in (1/2, 1)$ verrebbero, infatti, favorite configurazioni di grafo distanti dalla misura di centralità W .

2. Proprietà del modello GCER

In questa sezione vengono esposte le proprietà che caratterizzano il modello probabilistico GCER e la sua relazione con il modello CER. L'analisi di tali proprietà consente di definire il ruolo dei parametri di locazione W e di scala α . In primo luogo viene analizzato il parametro W evidenziando il suo legame con il parametro di locazione del modello CER \mathcal{G}^m e il suo ruolo nel rappresentare una misura di centralità per la distribuzione. Successivamente, l'analisi si concentra sul parametro di scala α dimostrandone la capacità di quantificare la variabilità della distribuzione.

2.1. Proprietà relative al parametro di concentrazione

Al fine di studiare il ruolo che il parametro W ricopre nel modello GCER, risulta necessario introdurre la seguente nozione:

Definizione 2.1. Sia Y una variabile aleatoria definita sullo spazio \mathcal{Y} e sia $d_H(\cdot, \cdot)$ la distanza di Hamming definita su \mathcal{Y} . La soluzione del problema di minimo:

$$\phi^m = \operatorname{argmin}_{\phi \in \mathcal{Y}} \mathbb{E}[d_H^2(Y, \phi)]$$

è chiamata media di Fréchet.

Proposizione 2.1. : Sia $\mathcal{G} \sim \text{GCER}(W, \alpha)$ un grafo aleatorio definito su $\mathcal{G}_{[N]}$, allora la media di Fréchet della distribuzione può essere interpretata come una discretizzazione del grafo pesato W :

$$\mathcal{G}^m = \lfloor 2W \rfloor,$$

dove $\lfloor 2W \rfloor$ si ottiene prendendo la parte intera di ogni elemento $2W[i, j]$ della matrice $2W$.

Dimostrazione. Sia Y un grafo aleatorio definito su $\mathcal{G}_{[N]}$, la media di Fréchet è ottenuta come soluzione del problema di minimo:

$$\mathcal{G}^m = \operatorname{argmin}_{\mathcal{G} \in \mathcal{G}_{[N]}} \mathbb{E}[d_{\text{H}}^2(Y, \mathcal{G})],$$

dove il quadrato della distanza di Hamming è definita come:

$$d_{\text{H}}^2(Y, \mathcal{G}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N (|Y[i, j] - \mathcal{G}[i, j]|)^2 = \sum_{i=1}^{N-1} \sum_{j=i+1}^N |Y[i, j] - \mathcal{G}[i, j]|.$$

Il valore atteso della distanza di Hamming al quadrato è esprimibile nel seguente modo:

$$\mathbb{E}[d_{\text{H}}^2(Y, \mathcal{G})] = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathbb{E}[|Y[i, j] - \mathcal{G}[i, j]|].$$

Sotto l'ipotesi di esistenza di un arco tra la coppia di nodi i e j , ovvero $\mathcal{G}[i, j] = 1$, l'errore quadratico atteso è dato da $\mathbb{E}[d_{\text{H}}^2(Y[i, j], 1)] = 1 - p_{i,j}$. Al contrario in assenza di connessione tra la coppia di nodi i e j l'errore quadratico medio assume valore $\mathbb{E}[d_{\text{H}}^2(Y[i, j], 0)] = p_{i,j}$. Si può notare che se $p_{i,j} < 1/2$, l'errore quadratico medio è minimizzato scegliendo $\mathcal{G}[i, j] = 0$. Viceversa, se $p_{i,j} > 1/2$, l'errore risulta minimo se $\mathcal{G}[i, j] = 1$. Pertanto la media di Fréchet \mathcal{G}^m può essere riscritta come:

$$\mathcal{G}^m[i, j] = \begin{cases} 1 & \text{se } p_{i,j} > 1/2 \\ 0 & \text{se } p_{i,j} < 1/2 \end{cases}$$

dove

$$p_{i,j} = \Pr(\mathcal{G}[i, j] = 1) = \left(1 + \left(\frac{\alpha}{1 - \alpha} \right)^{(2W[i,j]-1)} \right)^{-1}.$$

La probabilità di osservare un arco tra la coppia di nodi i e j può essere riscritta in modo

esplicito considerando la condizione $p_{i,j} > 1/2$, come segue:

$$p_{i,j} > 1/2 \Leftrightarrow \left(1 + \left(\frac{\alpha}{1-\alpha}\right)^{(2W[i,j]-1)}\right)^{-1} > 1/2 \Leftrightarrow \left(\frac{\alpha}{1-\alpha}\right)^{(2W[i,j]-1)} < 1$$

Per $0 < \alpha < 1/2$, la condizione sopra riportata è rispettata se $W[i, j] > 1/2$. La media di Fréchet, quindi, può essere riscritta come:

$$\mathcal{G}^m[i, j] = \begin{cases} 1 & \text{se } W_{i,j} > 1/2 \\ 0 & \text{se } W_{i,j} < 1/2 \end{cases}$$

oppure, in forma compatta come:

$$\mathcal{G}^m[i, j] = \lfloor 2W[i, j] \rfloor.$$

□

Proposizione 2.2. La moda e la media di Fréchet coincidono per il modello G-CER.

Dimostrazione. La dimostrazione della Proposizione 2.2 è suddivisa in due parti. La prima parte della dimostrazione coincide con la prima parte della Proposizione 3.6 dell'articolo di Lunagómez et al. (2021).

Parte 1: Siano a e b due vettori con s elementi; inoltre sia w tale che $w_i > 0$ e $\sum_i w_i = 1$ e sia a tale che $a_i \geq 0$. Sia σ una permutazione di $\{1, 2, \dots, s\}$ tale che $w_{\sigma[i]} \geq w_{\sigma[j]}$ per ogni coppia $[i, j]$ con $i < j$. Sia τ una permutazione di $\{1, 2, \dots, s\}$ tale che $\tau_{\sigma[i]} \leq \tau_{\sigma[j]}$ per ogni coppia $[i, j]$ con $i < j$. Pertanto τ soddisfa:

$$\tau = \operatorname{argmin}_{\phi \in \operatorname{Sym}(s)} \sum a_{\phi[i]} \times w_{\sigma[i]}$$

dove $\operatorname{Sym}(s)$ denota l'insieme di tutte le permutazioni degli indici s . Questa affermazione viene dimostrata per induzione:

per $s = 2$: si consideri il caso $a_{\tau[1]} < a_{\tau[2]}$ e $w_{\tau[1]} > w_{\tau[2]}$. Poiché gli elementi del vettore w sono non negativi e sommano a uno, segue che $w_{\sigma[1]} > 1/2 > w_{\sigma[2]}$. Allora la quantità

$$a_{\tau[1]} \times w_{\sigma[1]} + a_{\tau[2]} \times w_{\sigma[2]}$$

risulta essere più vicino a $a_{\tau[1]}$ rispetto a $a_{\tau[2]}$. Se si permutano gli indici a_{τ} , per ottenere un nuovo vettore $a_{\tau'}$ allora

$$a_{\tau'[1]} \times w_{\sigma[1]} + a_{\tau'[2]} \times w_{\sigma[2]}$$

è più vicino a $a_{\tau[2]}$ rispetto a $a_{\tau[1]}$. Poiché $a_{\tau[1]} < a_{\tau[2]}$, la condizione risulta soddisfatta

per $s = k$: si assume che il risultato valga per $s = k$

per $s = k + 1$: si considerino due casi, definiti in termini dell'esistenza di punti fissi di τ .

Caso 1. Almeno un elemento di a_{τ} rimane fisso. Senza perdita di generalità, si può assumere che l'elemento di a_{τ} che rimane invariato è il $(k + 1)$ -esimo. Risulta possibile rinormalizzare i primi k elementi di w_{σ} definendo:

$$w'_{\sigma[i]} = \frac{w_{\sigma[i]}}{1 - w_{\sigma[k+1]}}.$$

Poiché $w'_{\sigma[i]} > w'_{\sigma[j]}$ per ogni $1 \leq i < j \leq k$, si può applicare l'ipotesi di induzione alle prime k entrate di a_{τ} , $a_{\tau'}$ e $w'_{\sigma[i]}$ per ottenere:

$$\sum_{i=1}^k a_{\tau[i]} \times w'_{\sigma[i]} \leq \sum_{i=1}^k a_{\tau'[i]} \times w'_{\sigma[i]} \quad (2.1)$$

L'equazione (2.1) è valida per qualsiasi permutazione τ' che lasci l'elemento $k + 1$ invariato rispetto a τ . Pertanto vale:

$$\sum_{i=1}^k a_{\tau[i]} \times w_{\sigma[i]} \leq \sum_{i=1}^k a_{\tau'[i]} \times w_{\sigma[i]},$$

poichè è sufficiente moltiplicare $w'_{\sigma[i]}$ per una costante positiva, $1 - w_{\sigma[k+1]}$. Assumendo $a_{\tau[k+1]} = a_{\tau'[k+1]}$, segue che

$$\sum_{i=1}^{k+1} a_{\tau[i]} \times w_{\sigma[i]} \leq \sum_{i=1}^{k+1} a_{\tau'[i]} \times w_{\sigma[i]}.$$

Pertanto τ minimizza $a_{\tau[i]} \times w_{\sigma[i]}$ per tutte le permutazioni che lasciano invariato almeno un elemento rispetto a τ .

Caso 2: Si consideri ora il caso in cui nessun elemento di a_{τ} sia rimasto invariato con un nuovo ordinamento τ' . Sia a_{τ^*} il vettore ottenuto permutando due elementi di $a_{\tau'}$, in modo tale che l'elemento $(k + 1)$ -esimo di a_{τ^*} , coincida con l'elemento $(k + 1)$ -esimo di a_{τ} . Applicando un argomento analogo al caso $k = 2$ si ottiene:

$$\sum_{i=1}^{k+1} a_{\tau^*[i]} \times w_{\sigma[i]} \leq \sum_{i=1}^{k+1} a_{\tau'[i]} \times w_{\sigma[i]}$$

Poiché almeno l'elemento $(k + 1)$ -esimo di a_{τ^*} coincide con l'elemento $(k + 1)$ -esimo di a_{τ} , si ottiene:

$$\sum_{i=1}^{k+1} a_{\tau[i]} \times w_{\sigma[i]} \leq \sum_{i=1}^{k+1} a_{\tau^*[i]} \times w_{\sigma[i]},$$

pertanto, la condizione è valida anche per questo caso.

Parte 2. Si dimostri che il numero di grafi in $\mathcal{G}_{[N]}$ tale per cui $d_H(\mathcal{G}, \mathcal{G}^m) = h$, con $h \in (1, 2, \dots, N_e)$, è costante rispetto a \mathcal{G}^m . Una distanza $d_H(\mathcal{G}, \mathcal{G}^m) = h$ implica che h entrate della matrice di adiacenza di \mathcal{G}^m sono modificate. Poiché i grafi sono etichettati, il numero di grafi tale per cui $d_H(\mathcal{G}, \mathcal{G}^m) = h$ coincide con la seguente quantità:

$$\binom{N_e}{h},$$

che risulta costante rispetto a \mathcal{G}^m . Il valore atteso $\mathbb{E}[d_H^2(\mathcal{G}, \mathcal{G}_k)]$ per un modello GCER con

parametri (W, α) è definito come segue:

$$\mathbb{E}[d_{\text{H}}^2(\mathcal{G}, \mathcal{G}_k)] = \sum_{i \in I} d_{\text{H}}^2(\mathcal{G}_i, \mathcal{G}_k) p_{\text{GCER}}(\mathcal{G}_i; W, \alpha)$$

dove I è un insieme di indici che rappresenta le configurazioni \mathcal{G}_i in $\mathcal{G}_{[N]}$, ordinati in modo tale che $p_{\text{GCER}}(\mathcal{G}_i; W, \alpha) \geq p_{\text{GCER}}(\mathcal{G}_j; W, \alpha)$ per $i < j$. Poiché il numero di grafi in $\mathcal{G}_{[N]}$ con distanza $d_{\text{H}}(\mathcal{G}, \mathcal{G}^m) = h$ è costante rispetto a \mathcal{G}^m , per ogni $h \in \{1, 2, \dots, N_e\}$, il vettore delle distanze di Hamming rispetto a \mathcal{G}_k può essere ottenuto permutando le entrate del vettore delle distanze di Hamming rispetto a \mathcal{G}^m . Questo implica che

$$(d_{\text{H}}^2(\mathcal{G}_1, \mathcal{G}_k), d_{\text{H}}^2(\mathcal{G}_2, \mathcal{G}_k), \dots, d_{\text{H}}^2(\mathcal{G}_{|\mathcal{G}_{[N]}|}, \mathcal{G}_k))$$

è ottenuto permutando le entrate del vettore

$$(d_{\text{H}}^2(\mathcal{G}_1, \mathcal{G}^m), d_{\text{H}}^2(\mathcal{G}_2, \mathcal{G}^m), \dots, d_{\text{H}}^2(\mathcal{G}_{|\mathcal{G}_{[N]}|}, \mathcal{G}^m)).$$

Applicando il risultato della Parte (1) della dimostrazione, si può affermare che l'aspettativa della distanza di Hamming al quadrato è minimizzata per il grafo \mathcal{G}^m e quindi \mathcal{G}^m coincide con la media di Fréchet.

$$\mathbb{E}[d_{\text{H}}^2(\mathcal{G}, \mathcal{G}_k)] = \sum_{i \in I} d_{\text{H}}^2(\mathcal{G}_i, \mathcal{G}^m) p_{\text{GCER}}(\mathcal{G}_i; W, \alpha)$$

Questo risultato è valido in quanto, i grafi più vicini al parametro di locazione W sono anche quelli più prossimi alla moda della distribuzione in quanto \mathcal{G}^m è una discretizzazione del parametro W . Di conseguenza i grafi meno distanti da \mathcal{G}^m verranno pesati maggiormente in quanto avranno probabilità maggiore di essere osservati. \square

Proposizione 2.3. Siano \mathcal{G}_1 e \mathcal{G}_2 due grafi definiti sullo spazio $\mathcal{G}_{[N]}$. Se i due grafi \mathcal{G}_1 e \mathcal{G}_2 sono generati dal modello GCER con $W \in \mathcal{W}_{[N]}$ e se $\alpha \in (0, 1/2)$ allora $d_{\text{H}}(\mathcal{G}_1, W) >$

$d_H(\mathcal{G}_2, W)$ implica che $p_{\text{GCER}}(\mathcal{G}_2; W, \alpha) > p_{\text{GCER}}(\mathcal{G}_1; W, \alpha)$, mentre $d_H(\mathcal{G}_1, W) = d_H(\mathcal{G}_2, W)$ implica che $p_{\text{GCER}}(\mathcal{G}_1; W, \alpha) = p_{\text{GCER}}(\mathcal{G}_2; W, \alpha)$.

Dimostrazione. Siano \mathcal{G}_1 e \mathcal{G}_2 in $\mathcal{G}_{[N]}$ e W in $\mathcal{W}_{[N]}$.

$$\begin{aligned} \frac{p_{\text{GCER}}(\mathcal{G}_1; W, \alpha)}{p_{\text{GCER}}(\mathcal{G}_2; W, \alpha)} &= \frac{\left(\frac{1-\alpha}{\alpha}\right)^{-\sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}_1[i,j](1-2W[i,j])}}{\left(\frac{1-\alpha}{\alpha}\right)^{-\sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}_2[i,j](1-2W[i,j])}} \\ &= \left(\frac{1-\alpha}{\alpha}\right)^{\left(\sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}_2[i,j] - 2\sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}_2[i,j]W[i,j] + W[i,j]^2\right)} \\ &\quad \times \left(\frac{1-\alpha}{\alpha}\right)^{-\left(\sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}_1[i,j] - 2\sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}_1[i,j]W[i,j] + W[i,j]^2\right)} \\ &= \left(\frac{1-\alpha}{\alpha}\right)^{\left[d_H^2(\mathcal{G}_2, W) - d_H^2(\mathcal{G}_1, W)\right]} \end{aligned}$$

Allora, $\frac{p_{\text{GCER}}(\mathcal{G}_1; W, \alpha)}{p_{\text{GCER}}(\mathcal{G}_2; W, \alpha)} > 1$ se e solo se $[d_H^2(\mathcal{G}_2, W) - d_H^2(\mathcal{G}_1, W)] \log\left(\frac{1-\alpha}{\alpha}\right) > 0$.

Per $0 < \alpha < 1/2$, la quantità $\log\left(\frac{1-\alpha}{\alpha}\right)$ è positiva, tale per cui:

$$\frac{p_{\text{GCER}}(\mathcal{G}_1; W, \alpha)}{p_{\text{GCER}}(\mathcal{G}_2; W, \alpha)} > 1 \Leftrightarrow d_H^2(\mathcal{G}_2, W) > d_H^2(\mathcal{G}_1, W) \Leftrightarrow d_H(\mathcal{G}_2, W) > d_H(\mathcal{G}_1, W).$$

Seguendo lo stesso procedimento, è possibile dimostrare il caso $d_H(\mathcal{G}_2, W) = d_H(\mathcal{G}_1, W)$, condizione che implica $p_{\text{GCER}}(\mathcal{G}_1; W, \alpha) = p_{\text{GCER}}(\mathcal{G}_2; W, \alpha)$. \square

2.2. Proprietà relative al parametro di scala

Per comprendere al meglio l'effetto del parametro di scala α sulla distribuzione di probabilità, è utile studiare l'impatto che esso esercita sulla variabilità della distribuzione. Dato che la distribuzione del singolo arco è nota, risulta possibile definirne la varianza come sotto riportato

$$\begin{aligned}\text{Var}(\mathcal{G}[i, j]) &= \left(1 + \left(\frac{\alpha}{1-\alpha}\right)^{2W[i,j]-1}\right)^{-1} \left(1 - \left(1 + \left(\frac{\alpha}{1-\alpha}\right)^{2W[i,j]-1}\right)^{-1}\right) \\ &= \frac{\left(\frac{\alpha}{1-\alpha}\right)^{2W[i,j]-1}}{\left(1 + \left(\frac{\alpha}{1-\alpha}\right)^{2W[i,j]-1}\right)^2}.\end{aligned}\quad (2.2)$$

Dall'equazione (2.2) si può osservare che la varianza di un singolo arco è minima, ovvero $\text{Var}(\mathcal{G}[i, j]) = 0$, per $\alpha \rightarrow 0$ e $W[i, j] \in [0, 1] \setminus \{1/2\}$. Come precedentemente affermato, quando il parametro di scala va a zero, la distribuzione che ne risulta è degenera. Al contrario, ogni connessione tra i nodi i e j assume variabilità massima pari a $\text{Var}(\mathcal{G}[i, j]) = \frac{1}{4}$ per $\alpha \rightarrow 1/2$, $W \in (0, 1)$, e quando $\alpha \in (0, 1/2)$, $W = 1/2$, ovvero negli scenari di massima incertezza su ogni arco.

Come misura di variabilità globale della distribuzione GCER è stata utilizzata come metrica l'entropia. L'entropia è una misura di variabilità che quantifica l'incertezza associata alla distribuzione di probabilità e, nel contesto della famiglia $p_{\text{GCER}}(\cdot; W, \alpha)$, essa può essere interpretata come una misura di dispersione attorno al parametro di locazione W . L'entropia può essere definita nel seguente modo:

$$H_{\mathcal{G}} = - \sum_{\mathcal{G} \in \mathcal{G}_{[N]}} p_{\text{GCER}}(\mathcal{G}; W, \alpha) \log(p_{\text{GCER}}(\mathcal{G}; W, \alpha)).$$

Questa espressione può essere anche riscritta come:

$$H_{\mathcal{G}} = - \sum_{i=1}^{N-1} \sum_{j=i+1}^N H(\mathcal{G}[i, j]; W, \alpha) \quad (2.3)$$

dove $H(\mathcal{G}[i, j]; W, \alpha)$, rappresenta l'entropia per ogni coppia di nodi i e j .

Riscrivendo $H(\mathcal{G}[i, j]; W, \alpha)$ in forma estesa, si ottiene la seguente espressione:

$$\begin{aligned}
H(\mathcal{G}[i, j]) &= \left(1 - \left[1 + \left(\frac{\alpha}{1-\alpha} \right)^{2W[i,j]-1} \right]^{-1} \right) \log \left(1 - \left[1 + \left(\frac{\alpha}{1-\alpha} \right)^{2W[i,j]-1} \right]^{-1} \right) \\
&\quad + \left[1 + \left(\frac{\alpha}{1-\alpha} \right)^{2W[i,j]-1} \right]^{-1} \log \left(\left[1 + \left(\frac{\alpha}{1-\alpha} \right)^{2W[i,j]-1} \right]^{-1} \right) \\
&= \frac{\left(\frac{\alpha}{1-\alpha} \right)^{2W[i,j]-1}}{1 + \left(\frac{\alpha}{1-\alpha} \right)^{2W[i,j]-1}} \log \left(\frac{\alpha}{1-\alpha} \right)^{2W[i,j]-1} - \log \left(1 + \left(\frac{\alpha}{1-\alpha} \right)^{2W[i,j]-1} \right).
\end{aligned}$$

Sostituendo questa quantità all'interno della definizione di entropia in (2.3) si ottiene l'entropia globale della distribuzione GCER, ossia:

$$\begin{aligned}
H_{\mathcal{G}} &= - \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left[\frac{\left(\frac{\alpha}{1-\alpha} \right)^{2W[i,j]-1}}{1 + \left(\frac{\alpha}{1-\alpha} \right)^{2W[i,j]-1}} \log \left(\frac{\left(\frac{\alpha}{1-\alpha} \right)^{2W[i,j]-1}}{1 + \left(\frac{\alpha}{1-\alpha} \right)^{2W[i,j]-1}} \right) \right. \\
&\quad \left. + \frac{1}{1 + \left(\frac{\alpha}{1-\alpha} \right)^{2W[i,j]-1}} \log \left(\frac{1}{1 + \left(\frac{\alpha}{1-\alpha} \right)^{2W[i,j]-1}} \right) \right]. \tag{2.4}
\end{aligned}$$

Per studiare come impatta il parametro di scala α sull'entropia della distribuzione, si analizza il comportamento dell'entropia nei casi limite di α . Dalla formula (2.4) si osserva che per $\alpha \rightarrow 0$ e $W[i, j] \in [0, 1] \setminus \{1/2\}$

$$H_{\mathcal{G}} = 0.$$

In altre parole, quando α tende a zero, i grafi presentano concentrazione massima attorno al parametro di locazione W .

Al contrario per $\alpha \rightarrow 1/2$, $W[i, j] \in (0, 1)$ e per $W[i, j] = 1/2$, $\alpha \in (0, 1/2)$

$$H_{\mathcal{G}} = \frac{N(N-1)}{2}.$$

In questo scenario, la dispersione attorno al grafo pesato W è massima, in quanto su ogni arco $H[\mathcal{G}[i, j]] = 1/2$.

Proposizione 2.4. Per ogni parametro di locazione fisso $W \in \mathcal{W}_{[M]}$, l'entropia associata alla famiglia $p_{\text{GCER}}(\cdot; W, \alpha)$ è una funzione monotona di α .

Dimostrazione. Per dimostrare la Proposizione 4.2.1 si procede tramite il calcolo della derivata dell'equazione (2.4) rispetto al parametro di scala α

$$\frac{dH_{\mathcal{G}}}{d\alpha} = -\log\left(\frac{\alpha}{1-\alpha}\right) \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{(2W[i,j]-1)^2}{\left(1 + \left(\frac{\alpha}{1-\alpha}\right)^{(2W[i,j]-1)}\right)^2} \left(\frac{\alpha}{1-\alpha}\right)^{(2W[i,j]-1)-1}.$$

Poichè $\alpha \in (0, 1/2)$, la derivata dell'entropia risulta essere positiva, il che implica che l'entropia aumenta all'aumentare di parametro di scala α . □

3. Inferenza bayesiana per il modello GCER

Definiti gli aspetti teorici e le proprietà che caratterizzano il modello GCER, l'attenzione si concentra sulla stima dei parametri α e W attraverso tecniche di inferenza bayesiana.

Siano $\{\mathcal{G}_1, \dots, \mathcal{G}_n\}$ un insieme di osservazioni definite su $\mathcal{G}_{[N]}$ e sia W definito su $\mathcal{W}_{[N]}$. La distribuzione di probabilità a posteriori che si desidera stimare è la seguente:

$$p(W, \alpha \mid \mathcal{G}_1, \dots, \mathcal{G}_n) \propto p(W \mid \alpha) p(\alpha) \prod_{i=1}^n p(\mathcal{G}_i \mid W, \alpha), \quad (3.1)$$

dove $p(\mathcal{G}_i \mid W, \alpha)$ coincide con la funzione di verosimiglianza del modello GCER.

Per semplicità, si ipotizza che il parametro di scala α sia noto concentrando così l'analisi inferenziale solamente sul parametro di locazione W . In assenza di conoscenze a priori specifiche sulla struttura di W , viene attribuita ad ogni sua componente una distribuzione a priori uniforme definita sull'intervallo $[0, 1]$. Questa scelta permette di semplificare il calcolo della distribuzione a posteriori che può essere espressa come:

$$p(W \mid \mathcal{G}_1, \dots, \mathcal{G}_n) \propto \prod_{i=1}^n p(\mathcal{G}_i \mid W, \alpha). \quad (3.2)$$

Riscrivendo l'equazione (3.2) in forma estesa, si ottiene:

$$p(W \mid \mathcal{G}_1, \dots, \mathcal{G}_n) \propto \left(\frac{\alpha}{1-\alpha} \right)^{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \sum_{k=1}^n \mathcal{G}_k[i,j] (1-2W[i,j])} \prod_{i=1}^{N-1} \prod_{j=i+1}^N \left(1 + \left(\frac{\alpha}{1-\alpha} \right)^{1-2W[i,j]} \right)^{-n}. \quad (3.3)$$

Dall'equazione (3.3) si osserva che la distribuzione a posteriori non appartiene ad una famiglia di distribuzioni nota e presenta una forma analitica alquanto complessa da trattare. Per campionare il parametro di locazione W dalla sua distribuzione a posteriori risulta quindi necessario ricorrere a metodi algoritmici simulativi, come il Metropolis–Hastings Random Walk, un algoritmo Monte Carlo a catena di Markov utilizzato per campionare da distribuzioni a posteriori complesse.

L'idea di base del metodo consiste nel proporre ad ogni iterazione un nuovo candidato di $W'[i, j]$ ottenuto campionando da una distribuzione normale centrata nel valore di $W[i, j]$ all'iterazione precedente e varianza fissata pari a σ^2 . Il nuovo candidato viene accettato o rifiutato sulla base del valore assunto dal rapporto di accettazione definito come:

$$r = \min\left\{1, \frac{p(W[i, j] | W'[i, j])}{p(W'[i, j] | W[i, j])}\right\}$$

dove $W[i, j]$ rappresenta il valore corrente e $W'[i, j]$ il nuovo candidato.

La scelta del parametro σ^2 del kernel gaussiano gioca un ruolo cruciale, in quanto incide direttamente sulla probabilità di accettazione dei valori di $W[i, j]$ proposti nel metodo Metropolis–Hastings Random Walk. Se si considerano valori troppo elevati di σ^2 , il tasso di accettazione rischia di essere troppo basso in quanto l'algoritmo, effettuando step troppo grandi, non riesce ad esplorare tutto lo spazio parametrico, viceversa, se si considerano valori troppo bassi di σ^2 il tasso di accettazione rischia di essere molto elevato in quanto l'algoritmo, effettuando passi molto brevi, tende ad accettare sempre gli stessi valori.

Questi due scenari risultano essere problematici al fine di garantire la convergenza dell'algoritmo; per tale ragione risulta necessario fissare il parametro di variabilità in modo efficiente.

4. Inferenza bayesiana per il modello CER con prior GCER

Questa sezione ha lo scopo di introdurre diversi modelli bayesiani gerarchici al fine di condurre un'analisi inferenziale sui parametri di locazione \mathcal{G}^m e scala α che caratterizzano il modello CER. Tale analisi viene condotta sotto l'assunzione che la distribuzione campionaria e la distribuzione a priori fissata per la media di Fréchet abbiano forme funzionali differenti. Per quanto riguarda la funzione di verosimiglianza, si ipotizza che il campione segua una distribuzione CER, mentre al parametro di locazione \mathcal{G}^m viene assegnata come distribuzione a priori la distribuzione GCER.

Sia $N \in \mathbb{N}$ e d_H la distanza di Hamming definita su $\mathcal{G}_{[N]}$. Siano $\{\mathcal{G}_1, \dots, \mathcal{G}_n\}$ un insieme di osservazioni definite su $\mathcal{G}_{[N]}$. La distribuzione d'interesse da stimare è la seguente:

$$p(\mathcal{G}^m, \alpha \mid \mathcal{G}_1, \dots, \mathcal{G}_n) = p(\mathcal{G}^m \mid \alpha) p(\alpha) \prod_{i=1}^n p(\mathcal{G}_i \mid \mathcal{G}^m, \alpha) \quad (4.1)$$

dove:

$$p(\mathcal{G}_i \mid \mathcal{G}^m, \alpha) = \alpha^{d_H(\mathcal{G}_i, \mathcal{G}^m)} (1 - \alpha)^{\frac{(N-1)N}{2} - d_H(\mathcal{G}_i, \mathcal{G}^m)}. \quad (4.2)$$

Le espressioni (4.1) e (4.2) rappresentano gli elementi chiave per definire differenti specificazioni di modelli gerarchici bayesiani.

4.1. Distribuzione GCER come prior coniugata in un modello CER

Proposizione 4.1. La distribuzione GCER è prior coniugata per il parametro di locazione \mathcal{G}^m del modello CER con parametro di scala α noto.

Prima di passare alla dimostrazione, è necessario precisare alcuni aspetti sulla struttura del modello gerarchico preso in considerazione. Nello specifico, il modello di riferimento è il seguente:

$$\begin{aligned}\mathcal{G}_i | \mathcal{G}^m, \alpha &\stackrel{\text{iid}}{\sim} \text{CER}(\mathcal{G}^m, \alpha) \\ \mathcal{G}^m &\sim \text{GCER}(W_0, \tau_{\nu_0}(\alpha)),\end{aligned}$$

dove $\nu_0 > 0$ e $W_0 \in \mathcal{W}_{[N]}$ e la funzione $\tau_{\nu_0} : [0, 1/2] \rightarrow [0, 1/2]$ è definita come:

$$\tau_{\nu_0}(\alpha) = \frac{\alpha^{\nu_0}}{(1 - \alpha)^{\nu_0} + \alpha^{\nu_0}}.$$

Tramite τ_{ν_0} il parametro α viene trasformato per definire il parametro di scala della distribuzione a priori GCER per \mathcal{G}^m in modo che lo stesso non coincida necessariamente con quello del kernel CER, caso che si verifica fissando $\nu_0 = 1$.

La quantità ν_0 può essere interpretata come la grandezza del campione ideale su cui si basa l'opinione a priori su \mathcal{G}^m . Si osserva che per valori elevati di ν_0 , il parametro di scala diventa sempre più piccolo e di conseguenza le realizzazioni risultano essere molto concentrate attorno alla misura base a priori W_0 .

Dimostrazione. Sotto l'ipotesi che il parametro di scala α sia noto, la distribuzione a posteriori di \mathcal{G}^m è proporzionale al prodotto tra la funzione di verosimiglianza e la distribuzione a priori per \mathcal{G}^m . Formalmente:

$$p(\mathcal{G}^m | \mathcal{G}_1, \dots, \mathcal{G}_n) \propto p(\mathcal{G}_1, \dots, \mathcal{G}_n | \mathcal{G}^m)p(\mathcal{G}^m).$$

La funzione di verosimiglianza è data dall'equazione (4.2).

Per quanto riguarda la distribuzione a priori essa assume la seguente forma:

$$p(\mathcal{G}^m) \propto \left(\frac{\tau_{\nu_0}(\alpha)}{1 - \tau_{\nu_0}(\alpha)} \right)^{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}^m[i,j](1-2W_0[i,j])}$$

dove, sviluppando la quantità $\tau_{\nu_0}(\alpha)$ si può riscrivere:

$$p(\mathcal{G}^m) \propto \left(\frac{\alpha}{1-\alpha} \right)^{\nu_0 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}^m[i,j](1-2W_0[i,j])} .$$

Definite le quantità necessarie per calcolare la distribuzione a posteriori per \mathcal{G}^m , si illustrano i passaggi analitici per determinarla.

$$\begin{aligned} p(\mathcal{G}^m | \mathcal{G}_1, \dots, \mathcal{G}_n) &\propto p(\mathcal{G}_1, \dots, \mathcal{G}_n | \mathcal{G}^m) p(\mathcal{G}^m) \\ &= \alpha^{\nu_0 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}^m[i,j](1-2W_0[i,j]) + \sum_{k=1}^n d_H(\mathcal{G}^m, \mathcal{G}_k)} \\ &\quad \times (1-\alpha)^{\frac{nN(N-1)}{2} - \sum_{k=1}^n d_H(\mathcal{G}^m, \mathcal{G}_k) - \nu_0 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}^m[i,j](1-2W_0[i,j])} \\ &\propto \left(\frac{\alpha}{1-\alpha} \right)^{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}^m[i,j](\nu_0 - 2\nu_0 W_0[i,j] + n - 2 \sum_{k=1}^n \mathcal{G}_k[i,j])} \\ &= \left(\frac{\alpha}{1-\alpha} \right)^{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}^m[i,j](\nu_0 + n) \left(1 - 2 \frac{\nu_0 W_0[i,j] + \sum_{k=1}^n \mathcal{G}_k[i,j]}{\nu_0 + n} \right)} \\ &= \left(\frac{\alpha}{1-\alpha} \right)^{\nu' \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}^m[i,j](1-2W'[i,j])} . \end{aligned}$$

Dalla formula della distribuzione a posteriori di \mathcal{G}^m si può osservare che la distribuzione GCER è prior coniugata per il modello CER. Per tale ragione si può affermare che:

$$\mathcal{G}^m | \mathcal{G}_1, \dots, \mathcal{G}_n \sim \text{GCER}(W', \tau_{\nu'}(\alpha)) \quad (4.3)$$

dove

$$\tau_{\nu'}(\alpha) = \tau_{\nu_0+n}(\alpha)$$

e

$$W' = \frac{\nu_0}{\nu_0 + n} W_0 + \frac{n}{\nu_0 + n} \frac{\sum_{k=1}^n \mathcal{G}_k}{n} = \frac{\nu_0}{\nu_0 + n} W_0 + \frac{n}{\nu_0 + n} \bar{\mathcal{G}}.$$

□

Dalla struttura della distribuzione a posteriori di \mathcal{G}^m si può osservare che il nuovo parametro di scala $\tau_{\nu'}(\alpha)$ dipende sia dalla numerosità campionaria n che da ν_0 .

Per quanto riguarda il parametro di locazione, esso è definito come combinazione lineare della matrice W_0 , che rappresenta i pesi a priori associati ad ogni coppia di nodi i e j , e della matrice $\bar{\mathcal{G}}$, i cui elementi $\bar{\mathcal{G}}[i, j]$ corrispondono alla media campionaria delle connessioni per ogni coppia di nodi. I pesi di questa combinazione lineare sono proporzionali rispettivamente al parametro ν_0 e alla grandezza campionaria n .

Grazie alla proprietà di coniugatezza di cui gode la distribuzione GCER, risulta possibile definire la probabilità a posteriori per \mathcal{G}^m di osservare una connessione tra la coppia di nodi i e j . Tale probabilità è definita nel seguente modo:

$$\Pr(\mathcal{G}^m[i, j] = 1 \mid \mathcal{G}_1, \dots, \mathcal{G}_n) = \left[1 + \left(\frac{\alpha}{1 - \alpha} \right)^{\nu'(2W'[i, j] - 1)} \right]^{-1}.$$

Di conseguenza, la variabile $\mathcal{G}^m[i, j]$ condizionatamente alle osservazioni $\mathcal{G}_1, \dots, \mathcal{G}_n$ segue una distribuzione bernoulliana della seguente forma:

$$\mathcal{G}^m[i, j] \mid \mathcal{G}_1, \dots, \mathcal{G}_n \sim \text{Bernoulli} \left(\left[1 + \left(\frac{\alpha}{1 - \alpha} \right)^{\nu'(2W'[i, j] - 1)} \right]^{-1} \right).$$

4.2. Parametro di scala α non noto

Assumendo che il parametro α non sia noto, si possono definire due specificazioni distinte per le distribuzioni a priori per i parametri (\mathcal{G}^m, α) di un modello CER.

4.2.1. Prior con componenti dipendenti

Il modello di riferimento è il seguente:

$$\begin{aligned} \mathcal{G}_i \mid \mathcal{G}^m &\stackrel{\text{iid}}{\sim} \text{CER}(\mathcal{G}^m, \alpha), \\ \mathcal{G}^m \mid \alpha &\sim \text{GCER}(W_0, \tau_{\nu_0}(\alpha)), \\ \alpha &\sim \text{TBeta}(1/2, a_0, b_0). \end{aligned}$$

In questo scenario si è scelta come distribuzione a priori per il parametro di scala α una

distribuzione Beta troncata definita sull'intervallo $[0, 1/2]$, in quanto agevole sia da un punto di vista analitico sia per garantire che il supporto del parametro α risulti compreso tra 0 e $1/2$.

La distribuzione a posteriori congiunta per i parametri (\mathcal{G}^m, α) è proporzionale al prodotto tra la funzione di verosimiglianza, la distribuzione a priori assegnata al parametro di localizzazione \mathcal{G}^m e la distribuzione a priori assegnata al parametro di scala α . Formalmente:

$$p(\mathcal{G}^m, \alpha \mid \mathcal{G}_1, \dots, \mathcal{G}_n) \propto p(\mathcal{G}_1, \dots, \mathcal{G}_n \mid \mathcal{G}^m, \alpha) p(\mathcal{G}^m \mid \alpha) p(\alpha).$$

La funzione di verosimiglianza è data dall'equazione (4.2), mentre la distribuzione a priori per \mathcal{G}^m coincide con:

$$p(\mathcal{G}^m \mid \alpha) = \left(\frac{\alpha}{1-\alpha} \right)^{\nu_0 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}^m[i,j](1-2W_0[i,j])} \frac{1}{\prod_{i=1}^{N-1} \prod_{j=i+1}^N \left(1 + \left(\frac{\alpha}{1-\alpha} \right)^{\nu_0(1-2W_0[i,j])} \right)}.$$

Per quanto riguarda il parametro di scala α , ad esso viene assegnata la seguente distribuzione a priori Beta troncata:

$$p(\alpha) = \begin{cases} \frac{\alpha^{a_0-1}(1-\alpha)^{b_0-1}}{\int_0^{1/2} \alpha^{a_0-1}(1-\alpha)^{b_0-1} d\alpha}, & \text{se } \alpha \in [0, 1/2], \\ 0 & \text{altrimenti.} \end{cases}$$

Definite le quantità necessarie per calcolare la distribuzione a posteriori congiunta per i parametri (\mathcal{G}^m, α) , si illustrano i passaggi analitici per determinarla:

$$\begin{aligned} p(\mathcal{G}^m, \alpha \mid \mathcal{G}_1, \dots, \mathcal{G}_n) &\propto p(\mathcal{G}_1, \dots, \mathcal{G}_n \mid \mathcal{G}^m, \alpha) p(\mathcal{G}^m \mid \alpha) p(\alpha) \\ &= \alpha^{\sum_{k=1}^n d_H(\mathcal{G}^m, \mathcal{G}_k) + \nu_0 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}^m[i,j](1-2W_0[i,j]) + a_0 - 1} \\ &\quad \times (1-\alpha)^{\frac{nN(N-1)}{2} - \sum_{k=1}^n d_H(\mathcal{G}^m, \mathcal{G}_k) - \nu_0 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}^m[i,j](1-2W_0[i,j]) + b_0 - 1} \\ &\quad \times \frac{1}{\prod_{i=1}^{N-1} \prod_{j=i+1}^N \left(1 + \left(\frac{\alpha}{1-\alpha} \right)^{\nu_0(1-2W_0[i,j])} \right)} \end{aligned}$$

$$\begin{aligned}
&= \left(\frac{\alpha}{1-\alpha} \right)^{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}^m[i,j](\nu_0+n)} \left(1 - 2 \frac{\nu_0 W_0[i,j] + \sum_{k=1}^n \mathcal{G}_k[i,j]}{\nu_0+n} \right) \\
&\quad \times \alpha^{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}_k[i,j] + a_0 - 1} \\
&\quad \times (1-\alpha)^{\frac{nN(N-1)}{2} - \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}_k[i,j] + b_0 - 1} \\
&\quad \times \frac{1}{\prod_{i=1}^{N-1} \prod_{j=i+1}^N \left(1 + \left(\frac{\alpha}{1-\alpha} \right)^{\nu_0(1-2W_0[i,j])} \right)}.
\end{aligned}$$

Dalla formula della distribuzione a posteriori congiunta, si osserva che nello scenario in cui α non è noto, la proprietà di coniugatezza non è soddisfatta. Per tale ragione, per campionare dalla distribuzione a posteriori di (\mathcal{G}^m, α) , è necessario ricorrere a metodi algoritmici simulativi, come il metodo Gibbs Sampler combinato con Metropolis–Hastings Random Walk. L'idea base dell'algorithm Gibbs Sampler è quella di campionare dalla distribuzione congiunta a posteriori utilizzando le distribuzioni condizionate (full conditional) dei parametri coinvolti.

Dall'equazione della distribuzione a posteriori, risulta possibile ricavare le distribuzioni full conditional per i parametri \mathcal{G}^m e α . La distribuzione full conditional per \mathcal{G}^m coincide con il modello GCER con parametri aggiornati, mentre la distribuzione full conditional del parametro di scala α presenta una struttura più complessa dalla quale risulta difficile campionare direttamente. La sua forma funzionale è la seguente:

$$\begin{aligned}
p(\alpha \mid \mathcal{G}_1, \dots, \mathcal{G}_n, \mathcal{G}^m) &\propto \alpha^{\sum_{k=1}^n d_H(\mathcal{G}^m, \mathcal{G}_k) + \nu_0 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}^m[i,j](1-2W_0[i,j]) + a_0 - 1} \\
&\quad \times (1-\alpha)^{\frac{nN(N-1)}{2} - \sum_{k=1}^n d_H(\mathcal{G}^m, \mathcal{G}_k) - \nu_0 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}^m[i,j](1-2W_0[i,j]) + b_0 - 1} \\
&\quad \times \frac{1}{\prod_{i=1}^{N-1} \prod_{j=i+1}^N \left(1 + \left(\frac{\alpha}{1-\alpha} \right)^{\nu_0(1-2W_0[i,j])} \right)}.
\end{aligned}$$

Nell'algorithm Gibbs Sampling, per aggiornare ad ogni iterazione la matrice di adiacenza associata a \mathcal{G}^m , risulta sufficiente campionare dalla distribuzione GCER

$$p(\mathcal{G}^m \mid \mathcal{G}_1, \dots, \mathcal{G}_n, \alpha) \propto \left(\frac{\alpha}{1-\alpha} \right)^{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}^m \left(1 - 2 \frac{\nu_0 W_0[i,j] + \sum_{k=1}^n \mathcal{G}_k[i,j]}{\nu_0+n} \right)},$$

con parametri aggiornati $\tau_{\nu'}(\alpha) = \tau_{\nu_0+n}(\alpha)$ e $W' = \frac{\nu_0}{\nu_0+n}W_0 + \frac{n}{\nu_0+n}\bar{\mathcal{G}}$.

Per aggiornare il parametro di scala α , invece, viene utilizzato il metodo Metropolis–Hastings Random Walk. Ad ogni iterazione viene proposto un nuovo valore di α generato da una distribuzione normale centrata sul valore corrente, con una certa varianza σ^2 . Il nuovo parametro proposto viene accettato o rifiutato sulla base del valore assunto dal rapporto di accettazione, così definito:

$$r = \min\left(1, \frac{p(\alpha^{\text{proposta}} | \mathcal{G}_1, \dots, \mathcal{G}_n)}{p(\alpha^{\text{corrente}} | \mathcal{G}_1, \dots, \mathcal{G}_n)}\right).$$

La scelta del parametro σ^2 del kernel gaussiano gioca un ruolo cruciale, in quanto incide direttamente sulla probabilità di accettazione dei valori di α proposti nel metodo Metropolis–Hastings Random Walk. Per garantire la convergenza dell’algoritmo, il parametro σ^2 deve essere scelto in modo che il tasso di accettazione sia compreso all’interno di un intervallo ottimale. Studi simulativi suggeriscono un tasso di accettazione ideale compreso tra il 20% e 30%. Si veda anche Roberts & Rosenthal (2001).

4.2.2. Prior con componenti indipendenti

In questo contesto, si assume che il parametro di scala che parametrizza la funzione di verosimiglianza sia differente da quello della distribuzione a priori per \mathcal{G}^m .

Di conseguenza, il modello può essere esplicitato come segue:

$$\begin{aligned} \mathcal{G}_i | \mathcal{G}^m &\stackrel{\text{iid}}{\sim} \text{CER}(\mathcal{G}^m, \alpha), \\ \mathcal{G}^m &\sim \text{GCER}(W_0, \alpha_0), \\ \alpha &\sim \text{TBeta}(1/2, a_0, b_0). \end{aligned}$$

Si definisce la distribuzione a posteriori congiunta per i parametri (\mathcal{G}^m, α) che si è interessati a stimare nel seguente modo:

$$p(\mathcal{G}^m, \alpha | \mathcal{G}_1, \dots, \mathcal{G}_n) \propto p(\mathcal{G}_1, \dots, \mathcal{G}_n | \mathcal{G}^m, \alpha)p(\mathcal{G}^m)p(\alpha),$$

che riscritta in forma estesa assume la seguente forma funzionale:

$$\begin{aligned}
p(\mathcal{G}^m, \alpha \mid \mathcal{G}_1, \dots, \mathcal{G}_n) &\propto \alpha^{\sum_{k=1}^n d_H(\mathcal{G}^m, \mathcal{G}_k) + a_0 - 1} \\
&\times (1 - \alpha)^{\frac{nN(N-1)}{2} - \sum_{k=1}^n d_H(\mathcal{G}^m, \mathcal{G}_k) + b_0 - 1} \\
&\times \left(\frac{\alpha_0}{1 - \alpha_0} \right)^{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}^m[i,j] (1 - 2W_0[i,j])} \\
&\times \frac{1}{\prod_{i=1}^{N-1} \prod_{j=i+1}^N \left(1 + \left(\frac{\alpha_0}{1 - \alpha_0} \right)^{(1 - 2W_0[i,j])} \right)}.
\end{aligned}$$

Anche in questo caso la distribuzione a priori non risulta coniugata con la funzione di verosimiglianza. Per tale ragione per campionare dalla distribuzione a posteriori di (\mathcal{G}^m, α) è necessario ricorrere a metodi algoritmici simulativi, come il metodo Gibbs Sampler.

Dall'equazione della distribuzione a posteriori, risulta possibile ricavare le distribuzioni full conditional per i parametri \mathcal{G}^m e α .

La distribuzione full conditional per \mathcal{G}^m presenta la seguente forma funzionale:

$$\begin{aligned}
p(\mathcal{G}^m \mid \mathcal{G}_1, \dots, \mathcal{G}_n, \alpha) &\propto \left(\frac{\alpha}{1 - \alpha} \right)^{n \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}^m[i,j] \left(1 - 2 \sum_{k=1}^n \frac{\mathcal{G}_k[i,j]}{n} \right)} \\
&\times \left(\frac{\alpha_0}{1 - \alpha_0} \right)^{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathcal{G}^m[i,j] (1 - 2W_0[i,j])}.
\end{aligned}$$

La distribuzione full conditional per il parametro di scala α assume invece una forma nota e coincide con una distribuzione Beta troncata con parametri aggiornati:

$$\alpha \mid \mathcal{G}_1, \dots, \mathcal{G}_n, \mathcal{G}^m \sim \text{TBeta} \left(1/2, \sum_{k=1}^n d_H(\mathcal{G}^m, \mathcal{G}_k) + a_0, \frac{nN(N-1)}{2} - \sum_{k=1}^n d_H(\mathcal{G}^m, \mathcal{G}_k) + b_0 \right).$$

5. Analisi discriminante come esempio applicativo

Il contesto delle reti multiple dà origine a diversi risvolti applicativi, uno dei quali riguarda la definizione di regole decisionali al fine di classificare, sulla base delle loro proprietà strutturali, nuovi grafi, di cui non si conosce la classe di appartenenza.

In questa fase della trattazione ci si concentra sulla classificazione di osservazioni che si distribuiscono secondo il modello CER, analizzando i casi in cui le osservazioni si presentano sotto ipotesi di omoschedasticità e di eteroschedasticità.

A tale scopo, il classificatore bayesiano definisce la regola decisionale su cui si basa l'analisi discriminante per cui un nuovo soggetto appartiene alla categoria a cui corrisponde la massima probabilità a posteriori. Formalmente, assumendo che le osservazioni siano divise in K categorie:

$$\Pr(Y = k | \mathcal{G}) = \max_{k \in \{1, \dots, K\}} \Pr(Y = k | \mathcal{G}),$$

dove il termine $\Pr(Y = k | \mathcal{G})$ rappresenta la probabilità a posteriori che il grafo \mathcal{G} appartenga alla classe k . Questa quantità viene calcolata tramite il teorema di Bayes e può essere scritta come:

$$\Pr(Y = k | \mathcal{G}) = \frac{\Pr(Y = k)p(\mathcal{G} | Y = k)}{\sum_{k=1}^K \Pr(Y = k)p(\mathcal{G} | Y = k)}.$$

Il classificatore di Bayes dipende dalla distribuzione a priori $\Pr(Y = k)$, ossia dalla probabilità a priori di appartenere alla k -esima classe e dalla funzione di verosimiglianza $p(\mathcal{G} | Y = k)$ che descrive come i dati sono distribuiti all'interno di ogni classe.

Quest'ultima quantità coincide con il modello CER, tale per cui la funzione di verosimiglianza può essere riscritta in modo esteso come:

$$p(\mathcal{G} | Y = k) = \alpha^{d_H(\mathcal{G}_k^m, \mathcal{G})} (1 - \alpha)^{\frac{N(N-1)}{2} - d_H(\mathcal{G}_k^m, \mathcal{G})}.$$

5.1. Ipotesi di omoschedasticità

In questo scenario si ipotizza che il modello CER sia parametrizzato da un parametro di locazione \mathcal{G}^m specifico per ogni k -esima classe e un parametro di scala α comune per tutte le classi. Quando il grado di dispersione delle osservazioni attorno al parametro di locazione non dipende dalla classe di appartenenza, si verifica la condizione di omoschedasticità. In notazione, per un'osservazione \mathcal{G} assegnata alla k -esima classe:

$$\mathcal{G} \sim \text{CER}(\mathcal{G}_k^m, \alpha).$$

Si ipotizzi inizialmente che le osservazioni siano suddivise in due classi ($K = 2$). Un nuovo soggetto apparterrà al primo gruppo se vale la seguente condizione:

$$\Pr(Y = 1)p(\mathcal{G} | Y = 1) > \Pr(Y = 2)p(\mathcal{G} | Y = 2),$$

la quale, sotto ipotesi di uniformità delle distribuzioni a priori, può essere riscritta come segue:

$$p(\mathcal{G} | Y = 1) > p(\mathcal{G} | Y = 2).$$

Sostituendo le opportune quantità all'interno della disequazione e applicando per semplicità analitica la trasformazione logaritmica, si ottiene:

$$d_{\text{H}}(\mathcal{G}_1^m, \mathcal{G}) \log(\alpha) - d_{\text{H}}(\mathcal{G}_1^m, \mathcal{G}) \log(1 - \alpha) > d_{\text{H}}(\mathcal{G}_2^m, \mathcal{G}) \log(\alpha) - d_{\text{H}}(\mathcal{G}_2^m, \mathcal{G}) \log(1 - \alpha).$$

La probabilità che un grafo appartenga alla classe $k = 1$ è maggiore della probabilità che appartenga alla classe $k = 2$ se la distanza di Hamming tra il grafo e \mathcal{G}_1^m è minore di quella tra il grafo stesso e \mathcal{G}_2^m , ossia:

$$d_H(\mathcal{G}_1^m, \mathcal{G}) < d_H(\mathcal{G}_2^m, \mathcal{G}).$$

Estendendo il risultato ottenuto al caso in cui si considerino un numero di classi pari a K , la regola decisionale può essere così definita:

$$Class(\mathcal{G}) = \underset{k}{\operatorname{argmin}} d_H(\mathcal{G}, \mathcal{G}_k^m). \quad (5.1)$$

In conclusione si può affermare che un nuovo grafo viene assegnato alla classe il cui parametro di concentrazione minimizza la distanza di Hamming rispetto al grafo stesso.

5.2. Ipotesi di eteroschedasticità

Si consideri lo scenario in cui i grafi si distribuiscono seguendo un modello CER con parametri di locazione e di scala specifici per ogni classe, ossia:

$$\mathcal{G} \sim \operatorname{CER}(\mathcal{G}_k^m, \alpha_k).$$

In questo caso le osservazioni risultano essere eteroschedastiche, per cui il loro grado di dispersione attorno al parametro \mathcal{G}^m non è più comune per tutte le classi ma è definito specificatamente per ognuna di esse.

Si considerino inizialmente due classi $K = 2$, un nuovo grafo apparterrà alla prima classe se vale la seguente condizione:

$$\Pr(Y = 1)p(\mathcal{G} | Y = 1) > \Pr(Y = 2)p(\mathcal{G} | Y = 2).$$

Sotto ipotesi di uniformità delle distribuzioni a priori e riscrivendo la condizione in modo esplicito, si ottiene:

$$\alpha_1^{d_H(\mathcal{G}_1^m, \mathcal{G})} (1 - \alpha_1)^{\frac{N(N-1)}{2} - d_H(\mathcal{G}_1^m, \mathcal{G})} > \alpha_2^{d_H(\mathcal{G}_2^m, \mathcal{G})} (1 - \alpha_2)^{\frac{N(N-1)}{2} - d_H(\mathcal{G}_2^m, \mathcal{G})}.$$

Se il risultato ottenuto viene generalizzato al caso in cui si considerino un numero maggiore di classi pari a K , la regola decisionale, così definita:

$$Class(\mathcal{G}) = \underset{k}{\operatorname{argmax}} \alpha_k^{d_H(\mathcal{G}_k^m, \mathcal{G})} (1 - \alpha_k)^{\frac{N(N-1)}{2} - d_H(\mathcal{G}_k^m, \mathcal{G})}, \quad (5.2)$$

esplicita che un nuovo grafo apparterrà a quella classe a cui corrisponde la funzione di verosimiglianza massima.

6. Studio di simulazione

Una volta definiti gli aspetti teorici e le principali tecniche utilizzate per stimare i parametri del modello CER e GCER, l'analisi prosegue con la presentazione di uno studio simulativo che ha lo scopo di confrontare tre configurazioni diverse di modelli, al fine di valutare la flessibilità e l'adattabilità del modello GCER rispetto al modello CER.

I modelli gerarchici bayesiani presi in considerazione sono i seguenti:

- **Modello CER con prior CER su \mathcal{G}^m :**

$$\mathcal{G}_i | \mathcal{G}^m, \alpha \sim \text{CER}(\mathcal{G}^m, \alpha), \quad \mathcal{G}^m \sim \text{CER}(\mathcal{G}_0, \alpha).$$

In questo primo scenario viene utilizzata la stessa forma funzionale per modellare le osservazioni e il parametro di scala \mathcal{G}^m .

- **Modello CER con prior GCER su \mathcal{G}^m :**

$$\mathcal{G}_i | \mathcal{G}^m, \alpha \sim \text{CER}(\mathcal{G}^m, \alpha), \quad \mathcal{G}^m \sim \text{GCER}(W_0, \alpha).$$

In questo secondo scenario vengono utilizzate forme funzionali differenti per modellare le osservazioni e il parametro di scala \mathcal{G}^m .

- **Modello GCER con prior uniforme su W :**

$$\mathcal{G}_i | W, \alpha \sim \text{GCER}(W, \alpha), \quad W \sim \mathcal{U}(0, 1).$$

In questo ultimo caso, le osservazioni vengono modellate tramite il modello generalizzato GCER assegnando al parametro di locazione W una prior non informativa.

Nelle tre specificazioni appena introdotte si assume che il parametro di scala α sia noto. Per stimare il seguente parametro in modo efficiente è stato utilizzato un approccio “Em-

pirical Bayes”. L’idea di base consiste nello scegliere il parametro α che massimizza la distribuzione congiunta dei dati, o verosimiglianza marginale, formalmente:

$$\hat{\alpha} = \operatorname{argmax}_{\alpha \in (0,1/2)} p(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n).$$

La distribuzione congiunta dei dati $p(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n)$ varia sulla base del modello in esame:

- **Modello CER con prior CER o modello CER con prior GCER**

Nel caso di un modello CER la distribuzione congiunta è ottenuta marginalizzando rispetto a \mathcal{G}^m

$$p(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n) = \sum_{\mathcal{G}^m \in \mathcal{G}_{[N]}} p(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n | \mathcal{G}^m) p(\mathcal{G}^m). \quad (6.1)$$

- **Modello GCER con prior uniforme su W**

Nel caso di un modello GCER la distribuzione congiunta è ottenuta marginalizzando rispetto al grafo pesato W

$$p(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n) = \int_{[0,1]^{N(N-1)/2}} p(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n | W) p(W) dW. \quad (6.2)$$

Nel caso di un modello CER ricavare analiticamente la distribuzione congiunta dall’equazione (6.1) risulta essere molto complesso, in quanto marginalizzare rispetto a \mathcal{G}^m consiste nel sommare tutte le possibili configurazioni del parametro di locazione, il cui numero cresce esponenzialmente all’aumentare di N ed è pari a $2^{\frac{N(N-1)}{2}}$. Anche nel caso del modello GCER risolvere l’espressione (6.2) risulta essere molto complesso per via dell’elevata dimensionalità dello spazio di integrazione. Per tale ragione, in entrambi i casi, si procede a stimare la distribuzione congiunta tramite l’applicazione di metodi algoritmici simulativi di tipo Monte Carlo.

L’approccio Monte Carlo consiste nell’approssimare le equazioni (6.1) e (6.2) andando a campionare s volte, con s sufficientemente grande, i parametri \mathcal{G}^m e W dalle rispettive distribuzioni a priori. Nel caso del modello CER, la distribuzione congiunta viene approssimata come segue:

$$p(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n) \approx \frac{1}{s} \sum_{i=1}^s p(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n | \mathcal{G}_i^m)$$

dove $\mathcal{G}_1^m, \dots, \mathcal{G}_s^m$ sono i campioni generati dalla distribuzione a priori $p(\mathcal{G}^m)$. Analogamente, per il modello GCER:

$$p(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n) \approx \frac{1}{s} \sum_{i=1}^s p(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n | W_i)$$

dove $\mathcal{G}_1^m, \dots, \mathcal{G}_s^m$ sono i campioni generati dalla distribuzione a priori $p(W)$. In entrambi i casi l'approssimazione Monte Carlo ottenuta è una quantità che dipende dal parametro di scala α . Tramite metodi di ottimizzazione numerica si identifica il valore del parametro α che massimizza la distribuzione $p(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n)$.

Concentrandosi ora sulla stima dei parametri di locazione \mathcal{G}^m e W , nelle tre configurazioni sono stati utilizzati i seguenti approcci:

- **Modello CER con prior CER**

Il parametro di locazione \mathcal{G}^m è stato stimato implementando un algoritmo Gibbs Sampling con un numero di iterazioni pari a 10000 di cui 1000 di burn in.

- **Modello CER con prior GCER**

Nel caso di un modello CER con prior GCER, il parametro \mathcal{G}^m è stato stimato sfruttando la proprietà di coniugatezza di cui gode la distribuzione GCER rispetto al modello CER.

- **Modello GCER con prior uniforme su W**

Nel caso di un modello GCER, il parametro W è stato stimato implementando un Metropolis–Hastings Random Walk con un numero di iterazioni pari a 1000 di cui 100 di burn in.

Definite le strategie di stima utilizzate per stimare i parametri delle tre configurazioni di modelli, si procede con la simulazione di due dataset con $n = 1000$ osservazioni e $N = 10$ nodi. I due dataset sono stati simulati considerando due scenari differenti: nel primo caso le osservazioni sono state generate da un modello CER e successivamente da un modello GCER.

L'obiettivo dell'analisi consiste nel valutare le performance di adattamento dei tre modelli sulla base dei due scenari simulativi. Dal confronto tra le tre configurazioni modellistiche applicate ai due dataset, ci si attende che il modello CER presenti la miglior capacità di adattamento ai dati generati dalla stessa forma funzionale, pur mantenendo performance simili a quelle del modello GCER con prior uniforme. Al contrario, nel caso di dati generati dal modello GCER, ci si aspetta un divario significativo tra le performance del modello GCER e quelle del modello CER.

Per valutare la bontà di adattamento dei tre modelli è stata scelta come metrica la Log Pseudo Marginal Likelihood (Geisser & Eddy, 1979). La Log Pseudo Marginal Likelihood è un criterio bayesiano utilizzato per effettuare model selection ed è definito nel seguente modo:

$$\text{LPML} = \sum_{i=1}^n \log(\text{CPO}_i),$$

dove CPO_i coincide con $f_i(y_i | y_{-i}, M)$, ossia con la densità predittiva stimata sulla base di tutte le osservazioni, dato il modello M , escludendone l' i -esima. Per calcolare la LPML è dimostrabile che campionando dalla distribuzione a posteriori del vettore dei parametri θ , la quantità CPO_i può essere definita come:

$$(\text{CPO}_i)^{-1} = \frac{1}{s} \sum_{k=1}^s \frac{1}{f_i(y_i | \theta^k, M)},$$

dove:

- $\theta = (\theta_1, \dots, \theta_m)$ rappresenta il vettore dei parametri del modello con m componenti
- θ^k è il valore del vettore dei parametri al passo k della catena di campionamento
- $f_i(y_i | \theta^k, M)$ rappresenta la funzione di verosimiglianza per l'osservazione i -esima, dato il modello M
- s coincide con il numero di campioni a posteriori generati.

Valori più elevati di LPML indicano che il modello presenta una migliore bontà di adattamento ai dati, per tale ragione si sceglierà il modello con la Log Pseudo Marginal Likelihood

maggiore.

Le performance di adattamento delle tre configurazioni modellistiche applicate ai due scenari simulativi sono riportate nella Tabella 6.1.

Origine Dataset	Modello	LPML
Simulato da modello CER	CER con prior CER	-22562.06
	CER con prior GCER	-22562.06
	GCER con prior uniforme	-22581.36
Simulato da modello GCER	CER con prior CER	-30018.25
	CER con prior GCER	-30010.25
	GCER con prior uniforme	-29551.02

Tabella 6.1: Valori di LPML per le tre configurazioni di modelli, in base all'origine del dataset.

I risultati ottenuti in tabella sono coerenti con quanto atteso. Si può osservare che nel primo scenario in cui i dati sono generati dal modello CER, le migliori performance in termini di LPML sono ottenute adottando lo stesso modello generativo dei dati. Le performance relative al modello GCER, tuttavia, risultano prossime a quelle del modello CER, evidenziandone la sua flessibilità.

Per quanto riguarda il secondo scenario in cui i dati sono simulati dal GCER, si osserva che il modello migliore corrisponde al modello generativo dei dati con prior uniforme. In questo caso, le performance del modello CER si discostano significativamente da quelle del modello GCER confermandone la rigidità strutturale e la limitata capacità di adattarsi a scenari più complessi. Si può inoltre osservare che l'utilizzo di una distribuzione a priori più flessibile per modellare il parametro di concentrazione del modello CER consente di catturare meglio la struttura dei dati.

7. Analisi dei dati Human Brain Networks

I risultati teorici esposti nei capitoli precedenti sono stati applicati al dataset ‘Human Brain Networks’. Questo dataset analizza la connettività cerebrale di trenta individui ognuno dei quali è stato sottoposto fino ad un massimo di dieci misurazioni attraverso scansioni di imaging a risonanza magnetica per diffusione (dMRI), per un totale di $n = 266$ osservazioni.

La connettività cerebrale, in particolare quella strutturale, fornisce informazioni sulla struttura di connessioni fisiche tra le diverse aree cerebrali. Analizzare la connettività strutturale è fondamentale per individuare anomalie che possono indicare l’esistenza di eventuali disturbi neurologici e psichiatrici. Ogni osservazione del dataset, come mostrato in Figura 7.1, coincide con un grafo binario etichettato non orientato con 48 nodi. Ogni nodo rappresenta una regione cerebrale, mentre gli archi binari identificano se sia presente o meno una connessione tra le differenti aree del cervello.

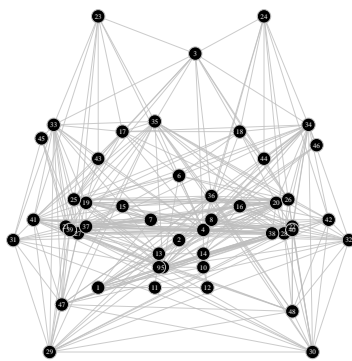


Figura 7.1: Osservazione i -esima del dataset Human Brain Networks

Lo scopo dell’analisi consiste nell’applicazione dell’inferenza bayesiana descritta in precedenza al fine di stimare, per ciascun soggetto, il parametro di centralità \mathcal{G}^m e il parametro di scala α . Per stimare i parametri d’interesse si è fatto riferimento al modello gerarchico espli-

citato nella Sezione 4.1 in cui, sotto ipotesi che il parametro di scala α sia costante, si sfrutta la proprietà di coniugatezza di cui gode la distribuzione a priori GCER definita sul parametro \mathcal{G}^m rispetto al modello CER. Questa proprietà offre un importante vantaggio computazionale poiché elimina la necessità di ricorrere a metodi algoritmici simulativi. Per ogni paziente, il parametro di scala α è stato stimato tramite l’approccio “Empirical Bayes”, descritto nel Capitolo 6 e i risultati ottenuti sono riportati in Tabella 7.1. Dalla sua analisi emerge che le osservazioni appaiono concentrate attorno al rispettivo parametro di locazione \mathcal{G}^m , in quanto i parametri α stimati per ogni paziente assumono valori prossimi al limite inferiore del suo dominio.

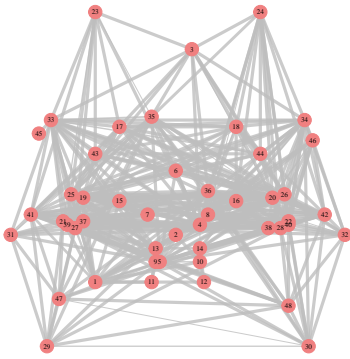
Paziente 1	Paziente 2	Paziente 3	Paziente 4	Paziente 5	Paziente 6
0.0621	0.0662	0.0542	0.0702	0.0596	0.0686
Paziente 7	Paziente 8	Paziente 9	Paziente 10	Paziente 11	Paziente 12
0.0819	0.0799	0.0726	0.0731	0.0605	0.0493
Paziente 13	Paziente 14	Paziente 15	Paziente 16	Paziente 17	Paziente 18
0.0769	0.0726	0.0745	0.0727	0.0723	0.0661
Paziente 19	Paziente 20	Paziente 21	Paziente 22	Paziente 23	Paziente 24
0.0712	0.0603	0.0563	0.0503	0.0502	0.0661
Paziente 25	Paziente 26	Paziente 27	Paziente 28	Paziente 29	Paziente 30
0.0969	0.0601	0.0664	0.0639	0.06280	0.0695

Tabella 7.1: Valori di α stimati per ogni paziente tramite approccio “Empirical Bayes” fissato $\nu_0 = 1$ modello CER prior GCER

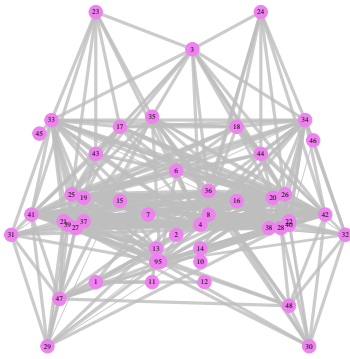
Per quanto riguarda la stima del parametro di locazione \mathcal{G}^m , essa è stata ottenuta campionando per ogni paziente dalla distribuzione a posteriori GCER (4.3).

La Figura 7.2 mostra degli esempi delle matrici di locazione \mathcal{G}^m stimate per sei pazienti in esame. Ogni nodo rappresenta una sezione differente del sistema cerebrale, mentre gli archi identificano la presenza di un legame tra le diverse regioni del cervello. In queste rappresentazioni lo spessore degli archi dipende dalla magnitudine della probabilità di osservare una connessione tra coppie di nodi. Maggiore è lo spesso degli archi, maggiore è la probabilità ad essi associata. Viceversa minore è il loro spessore, più bassa è la probabilità di osservare una connessione tra coppie distinte di nodi. L’incertezza associata ad ogni connessione consente di evidenziare eventuali differenze strutturali tra le configurazioni delle matrici \mathcal{G}^m e,

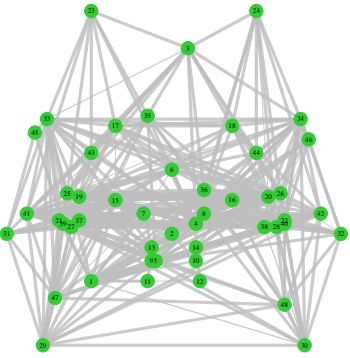
conseguentemente, tra i sistemi cerebrali dei pazienti.



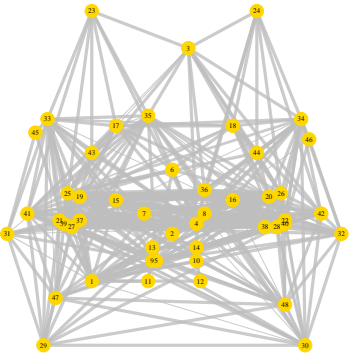
(a) Paziente 2



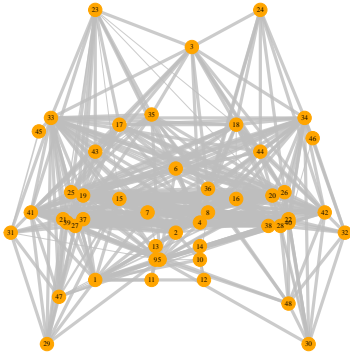
(b) Paziente 5



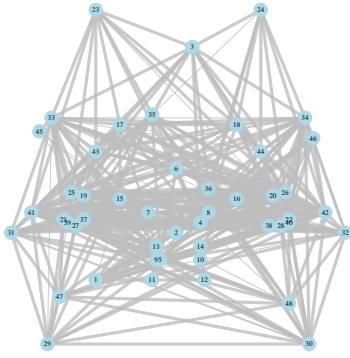
(c) Paziente 10



(d) Paziente 15



(e) Paziente 20



(f) Paziente 30

Figura 7.2: Stime \mathcal{G}^m per pazienti selezionati.

Risultati pressoché simili per le stime dei parametri α e \mathcal{G}^m sono stati ottenuti applicando le metodologie esposte nella Sezione 4.2, nel caso in cui α è considerata una variabile casuale. Dalle Tabelle 7.2 e 7.3 si osserva che le stime del parametro di scala ottenute tramite l’approccio “Empirical Bayes” sono coerenti con quelle ottenute tramite un approccio puramente bayesiano. Nello scenario in cui la funzione di verosimiglianza e la distribuzione a priori per il parametro di locazione \mathcal{G}^m condividono lo stesso parametro di scala α , la distribuzione congiunta dei parametri (\mathcal{G}^m, α) è stata stimata implementato un algoritmo Gibbs Sampling combinato con un Metropolis–Hastings Random Walk con un numero totale di iterazioni pari a 10000 di cui 1000 di burn-in.

Sotto ipotesi di indipendenza tra la funzione di verosimiglianza e la distribuzione a priori per \mathcal{G}^m , invece, è stato implementato un classico algoritmo Gibbs Sampling con lo stesso numero di iterazioni e di burn-in. In entrambi i casi, per la distribuzione a priori di α sono stati scelti $a_0 = 2$ e $b_0 = 3$ come iperparametri della distribuzione Beta.

Paziente	Valore α	Paziente	Valore α
1	0.0623	2	0.0663
3	0.0548	4	0.0706
5	0.0598	6	0.0692
7	0.0827	8	0.0806
9	0.0732	10	0.0736
11	0.0608	12	0.0496
13	0.0772	14	0.0732
15	0.0748	16	0.0733
17	0.0727	18	0.0664
19	0.0714	20	0.0606
21	0.0566	22	0.0505
23	0.0505	24	0.0663
25	0.0983	26	0.0603
27	0.0668	28	0.0641
29	0.0633	30	0.0700

Tabella 7.2: Valori stimati per α , scenario prior dipendenti fissato $\nu_0 = 1$.

Paziente	Valore α	Paziente	Valore α
1	0.055	2	0.057
3	0.0530	4	0.0710
5	0.0600	6	0.0703
7	0.0840	8	0.0800
9	0.0711	10	0.0730
11	0.0600	12	0.0501
13	0.0802	14	0.0740
15	0.0750	16	0.0740
17	0.0730	18	0.0670
19	0.0720	20	0.0610
21	0.0577	22	0.0510
23	0.0542	24	0.0702
25	0.0990	26	0.0600
27	0.0670	28	0.0645
29	0.0632	30	0.0731

Tabella 7.3: Valori stimati per α , scenario prior indipendenti fissato $\alpha_0 = 0.1$.

Nella Figura 7.3 sono riportate le diagnostiche relative alla stima di α_i , dove per α_i si intende la stima del parametro di scala per l' i -esimo soggetto ottenute applicando gli approcci simulativi descritti precedentemente. Analizzando i trace plot, si può osservare che entrambi gli algoritmi sembrerebbero aver raggiunto la convergenza in quanto le traiettorie delle stime oscillano stabilmente attorno ad un valore medio.

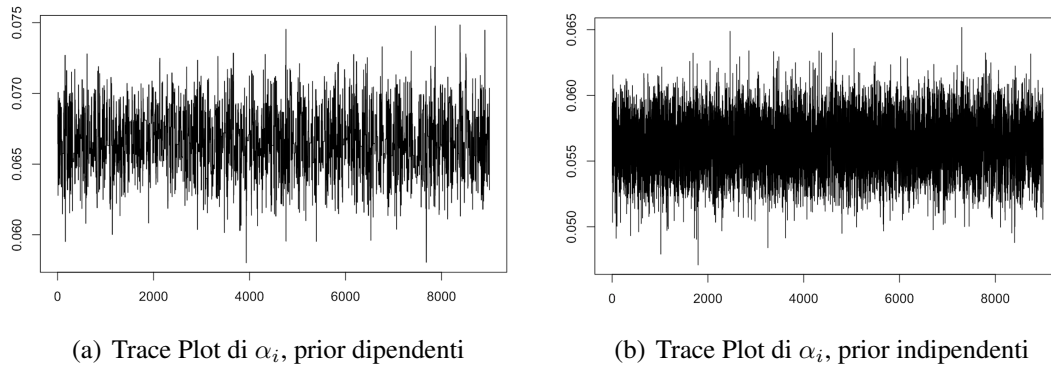


Figura 7.3: Trace plot

I risultati ottenuti dalla stima dei parametri (\mathcal{G}^m, α) , tramite l'applicazione del modello CER con distribuzione a priori per \mathcal{G}^m di tipo GCER, sono stati confrontati con quelli derivati dal modello CER con distribuzione a priori dello stesso tipo, al fine di valutare se i due approcci presentano eventuali differenze.

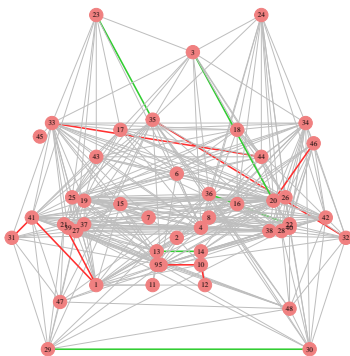
Anche in questo secondo scenario il parametro di scala α per ciascun paziente è stato stimato tramite l'approccio "Empirical Bayes". Per quanto riguarda il parametro di locazione \mathcal{G}^m , è stato implementato un algoritmo Gibbs Sampling con un numero di iterazioni pari a 10000 di cui 1000 di burn-in.

In Tabella 7.4 sono riportati i valori stimati di α per ogni paziente. Anche in questo caso si può osservare che i valori del parametro di scala sono molto simili tra loro. Confrontando i dati con quelli riportati in Tabella 7.4, si osserva che i parametri stimati utilizzando come distribuzione a priori per \mathcal{G}^m la distribuzione CER presentano una maggior magnitudine.

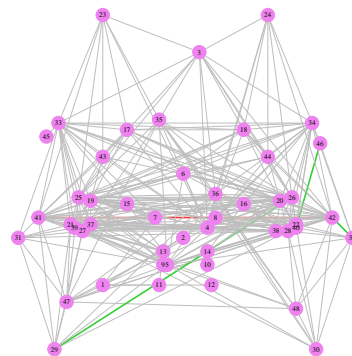
Paziente 1	Paziente 2	Paziente 3	Paziente 4	Paziente 5	Paziente 6
0.0998	0.0965	0.0936	0.1146	0.0962	0.1148
Paziente 7	Paziente 8	Paziente 9	Paziente 10	Paziente 11	Paziente 12
0.1122	0.1216	0.1058	0.1064	0.0946	0.1087
Paziente 13	Paziente 14	Paziente 15	Paziente 16	Paziente 17	Paziente 18
0.1065	0.1143	0.1083	0.1102	0.112	0.0.1065
Paziente 19	Paziente 20	Paziente 21	Paziente 22	Paziente 23	Paziente 24
0.1006	0.0927	0.1008	0.1183	0.0892	0.1050
Paziente 25	Paziente 26	Paziente 27	Paziente 28	Paziente 29	Paziente 30
0.1419	0.0948	0.0998	0.0975	0.01209	0.1142

Tabella 7.4: Valori di α stimati per ogni paziente tramite approccio “Empirical Bayes”, modello CER prior CER

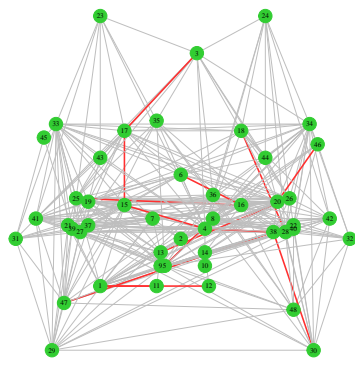
In Figura 7.4 sono rappresentati i parametri di locazione \mathcal{G}^m relativi a sei pazienti differenti. Ogni grafo evidenzia le connessioni stimate applicando i due scenari modellistici precedentemente citati. Le linee verdi rappresentano i legami stimati esclusivamente dal modello CER con distribuzione a priori CER, mentre le linee rosse le connessioni stimate dallo stesso modello CER ma con distribuzione a priori GCER. Le linee grigie, invece, identificano le connessioni comuni ai due approcci. In generale le matrici di locazioni stimate dai due metodi sono pressoché simili, non evidenziando particolari differenze tra i due metodi di stima.



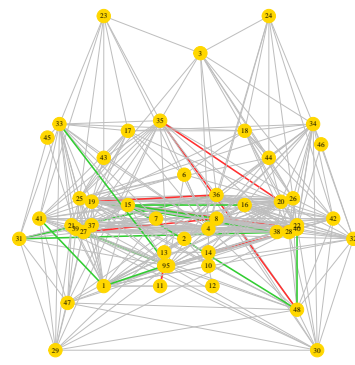
(a) Paziente 2



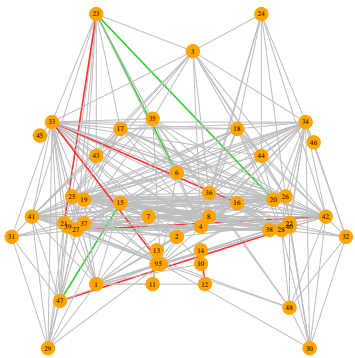
(b) Paziente 5



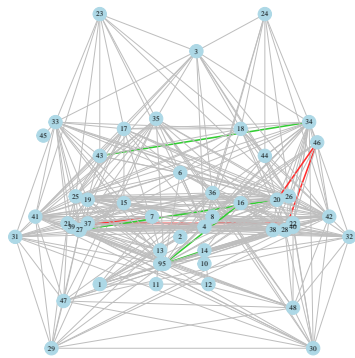
(c) Paziente 10



(d) Paziente 15



(e) Paziente 20



(f) Paziente 30

Figura 7.4: Stime di \mathcal{G}^m per sei pazienti selezionati.

La seconda parte dell'analisi si concentra sull'esposizione di un esempio di applicazione delle metodologie esposte nei capitoli precedenti, in particolare dell'analisi discriminante. L'obiettivo principale consiste nel valutare la capacità del modello CER con distribuzione a priori sul parametro \mathcal{G}^m di tipo GCER di classificare nuovi pazienti sulla base delle loro caratteristiche strutturali. Nel seguente studio sono presenti 30 soggetti distinti, ciascuno sottoposto a molteplici sessioni di screening. Lo scopo dell'analisi consiste nel determinare il gruppo di appartenenza di ciascun soggetto che si sottopone ad un nuovo screening sulla base delle sue caratteristiche cerebrali. Sotto ipotesi di eteroschedasticità delle osservazioni, per ogni paziente, una volta stimati i parametri di locazione \mathcal{G}^m e di scala α , è stata valutata la capacità discriminante del modello tramite l'applicazione della regola classificativa (5.2). A tal fine è stata eseguita una k -fold cross-validation con un numero di fold pari a 10 con

lo scopo di ottenere una stima robusta delle performance del modello. Per valutare la bontà classificativa del modello è stata utilizzata, come metrica principale, l'accuracy. Questa quantità misura la percentuale di classificazioni corrette rispetto al totale delle osservazioni e risulta utile perché fornisce una misura di sintesi della capacità del modello di distinguere le classi di appartenenza dei soggetti. Valori elevati dell'accuracy indicano una buona capacità classificativa del modello, al contrario valori inferiori segnalano una difficoltà del modello a discriminare in modo accurato le nuove osservazioni.

Il modello CER con distribuzione a priori GCER presenta un'ottima capacità discriminante, ottenendo un'accuratezza cross-validata pari al 99%. Dalle stime dei parametri di scala α riportati in Tabella 7.1 le osservazioni risultano essere molto concentrate attorno al rispettivo parametro di locazione identificando di conseguenza gruppi molto omogenei al loro interno con bassa variabilità e questo contribuisce a garantire una buona capacità classificativa del modello.

Nel Grafico 7.5 sono riportati i risultati ottenuti dall'analisi discriminante; ogni simbolo rappresenta una categoria distinta di pazienti mentre i quadratini colorati di blu rappresentano i soggetti che non sono stati correttamente classificati nel gruppo di appartenenza. Come si può osservare su 266 pazienti solamente due soggetti sono stati misclassificati, in particolare due soggetti appartenenti alla classe di pazienti con identificativo 0025433 e 0025451. Nel grafico i soggetti sono rappresentati in funzione di due quantità: 'Average Path Length' e 'Clustering Coefficient'. La prima metrica misura quanto distano mediamente le coppie di nodi che costituiscono una rete ed è definita nel seguente modo:

$$L = \frac{1}{N(N-1)} \sum_{i \neq j} d(i, j),$$

dove N è il numero di nodi che caratterizzano la rete e $d(i, j)$ rappresenta la distanza minima tra il nodo i e il nodo j . Questa quantità permette di definire l'efficienza con cui l'informazione si propaga all'interno della rete. Valori elevati di 'Average Path Length' stanno ad indicare che per passare da un nodo ad un altro occorre effettuare molti passi in termini di archi, al contrario valori bassi evidenziano che la rete presenta una buona connettività.

Il 'Clustering Coefficient', invece, permette di misurare la connettività locale della rete, ovvero se sono presenti dei sottogruppi di nodi strettamente connessi. Questa quantità è definita

nel seguente modo:

$$C = \frac{1}{N} \sum_{i=1}^N C_i,$$

dove

$$C_i = \frac{2E_i}{k_i(k_i - 1)}.$$

La quantità E_i corrisponde al numero di archi tra i vicini del nodo i , dove per vicini si intende tutti i nodi che condividono un arco con il nodo i . Il coefficiente k_i , invece, indica il grado del nodo i , ovvero il numero di nodi vicini. Valori elevati della misura di ‘Clustering Coefficient’ indicano che la rete presenta sottogruppi locali, al contrario valori bassi indicano che la rete presenta una struttura casuale.

Questa rappresentazione ha lo scopo di evidenziare se alcuni pazienti presentano comportamenti neurologici simili in modo da individuare eventuali anomalie cerebrali e conseguentemente eventuali patologie. Nel Grafico 7.5 si osserva che i soggetti tendono a distribuirsi in modo casuale all’interno dello spazio definito dalle due metriche indicando che i pazienti in studio presentano una configurazione neurologica pressoché simile.

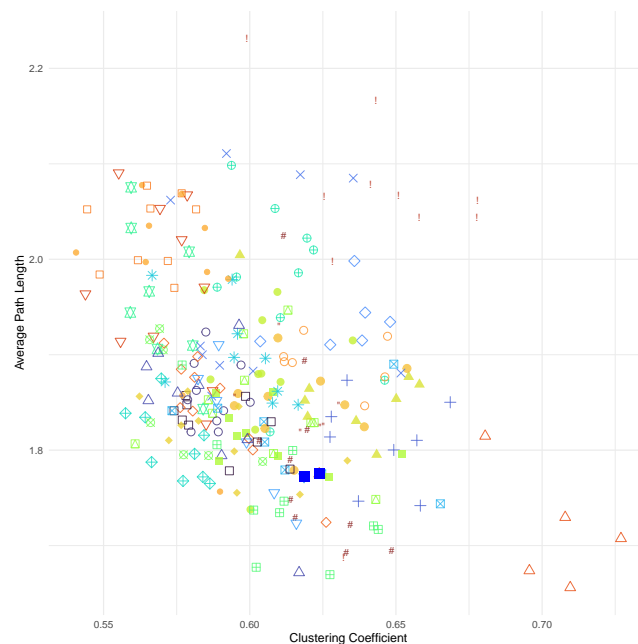


Figura 7.5: Risultati analisi discriminante cross validata modello CER con prior GCER

I risultati ottenuti dall'applicazione del modello CER con distribuzione a priori GCER sono stati confrontati con quelli ottenuti applicando lo stesso modello ma con distribuzioni a priori di tipo CER. Anche in questo scenario il modello presenta un'ottima capacità discriminante registrando un valore di accuracy pari al 98%.

Dal Grafico 7.6 si osserva che su 266 pazienti solamente quattro di essi sono stati erroneamente classificati.



Figura 7.6: Risultati analisi discriminante cross validata modello CER con prior CER

In generale entrambe le specificazioni presentano un'ottima capacità di classificare i pazienti sulla base delle loro caratteristiche strutturali. La scelta di una distribuzione a priori di tipo GCER permette di ottenere un miglioramento in termini di accuratezza, seppur minimo. Questo esempio applicativo non evidenzia per il modello CER una distinzione rilevante tra il caso in cui si utilizzi la prior CER o GCER, poiché si considerano delle osservazioni poco variabili e molto concentrate attorno alla propria misura di locazione garantendo ottime performance classificative in entrambi gli scenari. Come mostrato nell'analisi simulativa, l'introduzione di una distribuzione a priori più flessibile permette di migliorare l'adattamento del modello anche se in questo esempio applicativo tale miglioramento risulta poco evidente.

Conclusioni

In questa trattazione è stata introdotta una nuova estensione del modello CER, denominata GCER. In particolare è stato mostrato come questo modello goda di maggior flessibilità rispetto al modello base in quanto la misura di centralità che parametrizza la nuova distribuzione non è più un grafo binario ma un grafo pesato.

Oltre alla teorizzazione del nuovo modello probabilistico, sono state esposte le proprietà di cui esso gode e la relazione esistente tra il modello base CER e la sua generalizzazione.

Lo studio ha evidenziato un'importante proprietà del modello GCER per quanto riguarda la stima inferenziale del parametro di localizzazione, ossia la coniugatezza rispetto al modello CER. Questa proprietà garantisce un rilevante vantaggio computazionale nella stima del parametro di centralità sotto ipotesi che il parametro di scala sia costante.

Lo studio simulativo ha confermato che l'utilizzo di un modello GCER sia come modello da cui è stato generato il campione sia come distribuzione a priori per il parametro di centralità del modello CER, garantisce un miglioramento in termini di bontà di adattamento ai dati. Questo risultato evidenzia come modelli più flessibili possano meglio rappresentare i dati rispetto a modelli più rigidi, come il CER.

Nell'analisi applicativa si è ipotizzato che il modello generativo dei dati sia il modello CER, valutando l'impatto dell'adozione di una distribuzione a priori per il parametro di localizzazione più flessibile. Le tecniche inferenziali applicate al dataset reale 'Human Brain Networks', per la stima dei parametri di localizzazione e di scala, non hanno evidenziato un rilevante scostamento tra i risultati ottenuti utilizzando le specificazioni CER e GCER per definire la distribuzione a priori del parametro di concentrazione del modello CER.

Dall'analisi discriminante si è osservato che entrambe le specificazioni modellistiche presentano un'ottima capacità classificativa, anche se l'introduzione di una prior GCER permette di ottenere un leggero miglioramento in termini di bontà classificativa.

L'analisi applicativa si è limitata a confrontare il modello CER modellando il parametro di concentrazione tramite due specificazioni differenti di prior lasciando a successivi approfondimenti l'utilizzo del modello GCER come modello generativo dei dati.

Come dimostrato nel Capitolo 3, l'utilizzo del modello GCER per modellare i dati ha lo svantaggio di presentare un costo computazionale oneroso nel momento in cui il numero di nodi che caratterizzano il grafo risulta elevato. La forma funzionale del modello GCER, infatti,

è analiticamente complessa, per cui per fare inferenza sui parametri che parametrizzano la distribuzione risulta complesso ottenere una forma funzionale nota o facilmente trattabile per la distribuzione a posteriori.

Risulta inoltre necessario precisare un aspetto rilevante sulla struttura del modello probabilistico GCER, ossia la sua identificabilità. Un modello probabilistico è identificabile se combinazioni differenti dei parametri generano distribuzioni di probabilità distinte. Nel caso del modello GCER si possono avere problemi di identificabilità derivanti dal fatto che diverse combinazioni di parametri possono generare la stessa distribuzione di probabilità. Questo può rappresentare un problema in fase di stima dei parametri, in quanto non si è in grado di stimarli univocamente. Una soluzione per risolvere tale limitazione consiste nel fissare una componente del parametro di locazione W . In questo modo risulta possibile stimare in modo univoco le restanti componenti di W e il parametro α .

Bibliografia

- Barile, F., Lunagómez, S. & Nipoti, B. (2024). Bayesian nonparametric modeling of heterogeneous populations of networks. *ArXiv* .
- Christensen, R., Johnson, W., Branscum, A. & Hanson, T. (2010). *Bayesian Ideas and Data Analysis*. Chapman & Hall.
- Dubey, P. & Müller, H.-G. (2019). Fréchet analysis of variance for random objects. *Biometrika* .
- Geisser, S. & Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association* **74**, 153–160.
- Heard, N. (2021). *An Introduction to Bayesian Inference, Methods and Computation*. Springer.
- Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. Springer.
- Lunagómez, S., Olhede, S. C. & Wolfe, P. J. (2021). Modeling network populations via graph distances. *Journal of the American Statistical Association* **116**, 2023–2040.
- Marin, J.-M. & Robert, C. P. (2014). *Bayesian Essentials with R*. Springer.
- Roberts, G. O. & Rosenthal, J. S. (2001). Optimal scaling for various metropolis-hastings algorithms. *Statistical science* **16**, 351–367.