



UNIVERSITA' DEGLI STUDI DI MODENA E REGGIO EMILIA
Dipartimento di Scienze Fisiche, Informatiche e Matematiche

Corso di laurea in Matematica

IDENTIFICAZIONE DI GRUPPI COORDINATI IN SISTEMI
DINAMICI

Relatore:

Prof. Marco Villani

Tesi di Laurea di:

Camilla Rinaldi

Anno Accademico 2025-2026

Indice

Introduzione	1
1. Cell assemblies.....	2
1.1 Studio delle dinamiche corticali nel primate non umano	2
1.1.1 Obiettivo dell'esperimento	3
1.1.2 Risultati	4
1.2 Metodi per l'identificazione di cell assemblies.....	4
1.2.1 Tipologie di sincronizzazioni	5
1.2.2 Classificazione dei metodi di analisi: approcci puntuali e metodi di popolazione	7
1.3 Metodi di decomposizione e rilevamento di ensemble neuronali	11
1.3.1 Rilevamento di cell assemblies mediante riduzione della dimensionalità	12
1.3.2 Rilevamento di cell assemblies tramite analisi di rete.....	14
1.3.3 Costruzione agglomerativa di gruppi.....	15
2. Identificazione di gruppi dinamici.....	17
2.1 Nozioni di base	17
2.2 Cluster Index	19
2.3 Metodologia proposta	20
2.4 Versione standardizzata dell'integrazione	21
3. Un modello di rete neurale.....	24
3.1 Modellizzazione della Rete Neurale e Dinamiche del Segnale	24
3.2 Casi di Studio e Topologie di Rete.....	25
3.2.1 Topologia circolare.....	25
3.2.2 Topologia ad albero	27
3.3 Analisi dei dati ed osservazioni sperimentali.....	29
3.3.1 Finestra applicata alla Topologia circolare	30
3.3.2 Finestra applicata alla Topologia ad albero	36
4. Strutture che evolvono nel tempo.....	47
4.1 Validazione del framework Zi tramite tracciati dinamici con ground truth.....	47
4.1.1 Logica di Generazione dei Tracciati: il Modello Leader-Follower	48
4.2 L'Indice di Rand e la sua normalizzazione (Adjusted Rand Index)	49
4.3 Considerazioni Metodologiche sulla Dinamica dei Segnali e Criteri di Validazione	51
4.4 Cambio di Appartenenza di un Nodo	53

4.4.1 Analisi tramite Finestra Unica	53
4.4.2 Analisi Tramite Finestre Temporali	55
4.5 Nascita di un Nuovo Gruppo Indipendente	57
4.5.1 Analisi tramite Finestra Unica	57
4.5.2 Analisi Tramite Finestre Temporali	58
4.6 Rottura di un gruppo	61
4.6.1 Analisi tramite Finestra Unica	61
4.6.2 Analisi Tramite Finestre Temporali	62
4.7 Nascita di un Nuovo Gruppo (osservazione delle sole posizioni di firing)	65
4.7.1 Analisi tramite Finestra Unica	65
4.7.2 Analisi Tramite Finestre Temporali	66
4.7.3 Conclusioni	68
5. Analisi di dati biologici	70
5.1 Introduzione ai Dati Sperimentali	70
5.2 Analisi del Trend Temporale delle Attivazioni	71
5.3 Trattamento del Rumore di Fondo: Analisi "NoUni"	73
5.4 Analisi delle Frequenze di Sparo e Vincoli Biologici	74
5.5 Anomalie Sperimentali e Dinamiche di Spegnimento	76
5.6 Suddivisione in Aree Cerebrali e Analisi dei Tracciati	79
5.7 Discretizzazione dell'Attività: Approccio 1 vs Approccio 2	80
5.8 Studio della Stabilità dei Raggruppamenti tramite Indice ARI	82
Conclusione	85
Appendice A	86
Appendice B	93
Bibliografia	101

Introduzione

Si pensa che l'elaborazione delle informazioni nel cervello sia basata in larga misura sull'attività coordinata di gruppi di neuroni, noti come *cell assemblies*, che agiscono come vere e proprie unità funzionali attraverso una sincronizzazione temporale. Comprendere come queste reti neurali si riorganizzino dinamicamente per supportare funzioni cognitive complesse — come la trasformazione di uno scopo in un'azione motoria — rappresenta una delle sfide centrali della neuroscienza computazionale. Sebbene gli studi elettrofisiologici ad alta densità su topi e primati non umani forniscano dati ad altissima risoluzione sulle dinamiche corticali, l'identificazione e la caratterizzazione formale di queste sottostrutture funzionali all'interno di sistemi complessi solleva importanti sfide metodologiche.

Questa tesi si propone di utilizzare un framework analitico basato sulla teoria dell'informazione per individuare *cell assemblies* in tracciati neurologici reali. Il tool è basato sullo z-score dell'integrazione, unito a una procedura algoritmica di selezione (*sieving algorithm*) capace di tracciare l'organizzazione gerarchica del sistema.

Per verificare le opportunità ed i limiti dell'approccio si è reso necessario applicare il framework a tracciati di cui era noto il *ground truth*: allo scopo abbiamo quindi sviluppato due modelli, uno basato su strutture a grafo ed un altro di tipo più probabilistico. Inoltre è stato modificato l'approccio ai dati, in modo da (i) tenere conto della natura dei dati (che abbiamo trasformato da "singola emissione" a "misura su frequenze di emissione") e (ii) poter osservare anche gruppi dinamici di durata limitata.

Il lavoro è strutturato in cinque capitoli:

- Il **Capitolo 1** introduce il contesto neurobiologico, analizzando le dinamiche corticali in soggetti di *Macaca mulatta* durante compiti di discriminazione spaziale e temporale, evidenziando le proprietà uniche dell'attività "in squadra" (ASA).
- Il **Capitolo 2** delinea il quadro teorico, esaminando i concetti di entropia e integrazione, e presenta la metodologia originale basata sullo z-score dell'integrazione e sull'algoritmo di setaccio asintotico.
- Il **Capitolo 3** si sposta sulla modellizzazione computazionale, simulando la propagazione dei segnali in reti artificiali con topologie a rete: in particolare, topologie circolari e gerarchiche ad albero, analizzando l'impatto di ritardi, pesi sinaptici e rumore.
- Il **Capitolo 4** definisce un framework di validazione su modelli stocastici guidati da una logica *Leader-Follower*, dimostrando l'efficacia del tool nel decodificare riorganizzazioni strutturali profonde a fronte di una *ground truth* nota.
- Il **Capitolo 5** rappresenta la fase applicativa finale del lavoro, in cui il framework analitico viene testato su dati biologici reali provenienti da registrazioni elettrofisiologiche ad alta densità (al millisecondo) di modelli animali (topi) impegnati in test comportamentali di discriminazione. Il capitolo descrive le procedure di integrazione temporale in finestre discrete (bin di 0.1 secondi) e analizza la tendenza globale della rete, caratterizzata da una progressiva riduzione del numero dei nodi attivi lungo la sessione sperimentale.

1. Cell assemblies

1.1 Studio delle dinamiche corticali nel primate non umano

Presentiamo qui brevemente uno studio riguardante l'osservazione dell'attività cerebrale di un primate non umano, come esemplificazione delle metodologie applicate e degli interessi di ricerca degli scienziati del campo. L'elaborazione delle informazioni nel cervello (in particolare nella corteccia) si basa sull'attività coordinata di gruppi di neuroni chiamati *cell assemblies* (cell assemblies), che agiscono come vere e proprie unità funzionali.

Si ipotizza che i neuroni appartenenti allo stesso assemblamento esprimano una coordinazione temporale estremamente precisa nel generare impulsi elettrici (*spike*). Questa sincronizzazione temporale rende la trasmissione del segnale molto più efficace verso i neuroni successivi rispetto a impulsi non sincronizzati.

Per comprendere i meccanismi attraverso cui il cervello umano pianifica, decide e trasforma i pensieri in azioni, la neuroscienza computazionale si avvale del modello del primate non umano, in particolare del genere *Macaca mulatta*. I macachi possiedono un'organizzazione della corteccia frontale e prefrontale altamente omologa a quella umana, consentendo l'esecuzione di compiti cognitivi complessi che richiedono memoria di lavoro, attenzione selettiva e flessibilità comportamentale.

Le registrazioni elettrofisiologiche extracellulari ad alta densità effettuate su questi soggetti permettono di monitorare l'attività di centinaia di singoli neuroni simultaneamente mentre l'animale è vigile e attivamente impegnato nella risoluzione di un problema. Ciò fornisce una risoluzione temporale (al millisecondo) e spaziale (la singola cellula) che nessun metodo di neuroimaging non invasivo (come la fMRI o l'EEG) è in grado di eguagliare.

Nello studio dei macachi l'attenzione è focalizzata su compiti di discriminazione cognitiva (ad esempio, compiti basati sulla valutazione della durata o della distanza di stimoli visivi).

Un momento cruciale di questi esperimenti è la **fase di transizione dallo scopo all'azione** (*goal-action transformation*).

Durante un tipico trial:

1. L'animale riceve uno stimolo visivo che definisce la regola o la categoria (fase di codifica dello stimolo).
2. Segue un periodo di memoria (*delay period*), in cui l'informazione deve essere mantenuta attiva internamente. In questa fase lo "scopo" (*goal*) è astratto.
3. Al segnale di via, lo scopo astratto deve essere convertito in una pianificazione motoria specifica (l'azione motoria, l'azione motoria, come il rilascio di una leva o, nel caso di questo esperimento, il tocco di pulsanti).

La transizione tra la rappresentazione dello scopo e la selezione della risposta motoria non è immediata né banale. Essa richiede una riconfigurazione dinamica e rapidissima delle reti neurali nella corteccia frontale.

1.1.1 Obiettivo dell'esperimento

I ricercatori volevano capire come il cervello prende decisioni basate sul confronto tra due cose (il tempo e lo spazio). Per farlo, hanno addestrato due scimmie. A queste scimmie venivano presentati in sequenza due stimoli su uno schermo e dovevano stabilire quale dei due fosse più distante da un punto di riferimento o avesse una durata maggiore. Per consentire loro di interagire con i compiti, davanti alle scimmie erano stati posizionati tre interruttori a infrarossi.

Le prove in entrambi i compiti seguivano la stessa struttura temporale. Tuttavia, i due stimoli presentavano caratteristiche diverse.

Prima di iniziare, la scimmia doveva premere un pulsante centrale. Poi, sullo schermo comparivano due forme geometriche, **una alla volta** (prima lo "Stimolo 1" e poi lo "Stimolo 2").

Le forme erano sempre un cerchio blu e un quadrato rosso (se il primo era il cerchio, il secondo era il quadrato, e viceversa).

- **Compito della Durata (Tempo):** Le forme comparivano nello stesso identico posto, ma una restava sullo schermo più a lungo dell'altra. La scimmia doveva capire **quale delle due durava di più**.
- **Compito della Distanza (Spazio):** Le forme comparivano per lo stesso tempo, ma una era più vicina a un punto di riferimento e l'altra più lontana. La scimmia doveva capire **quale delle due compariva più lontano**.

Alla fine dei compiti, dopo una breve pausa in cui lo schermo restava vuoto, entrambe le forme ricomparivano contemporaneamente (una a destra e una a sinistra). Da quel momento la scimmia doveva toccare il pulsante corrispondente alla forma corretta (quella durata di più o quella più lontana). Le scimmie avevano a disposizione 6 secondi per selezionare lo stimolo che era più lontano dal punto di riferimento nel compito di distanza, o che era durato più a lungo nel compito di durata, toccando l'interruttore corrispondente. Prima del segnale di via, le scimmie non potevano pianificare alcuna risposta motoria.

Le risposte corrette venivano ricompensate con 0,1 ml di liquido, mentre le risposte errate erano seguite da un segnale acustico. Tutte le variabili del compito, il colore e la forma degli stimoli, sono state determinate in modo pseudocasuale.

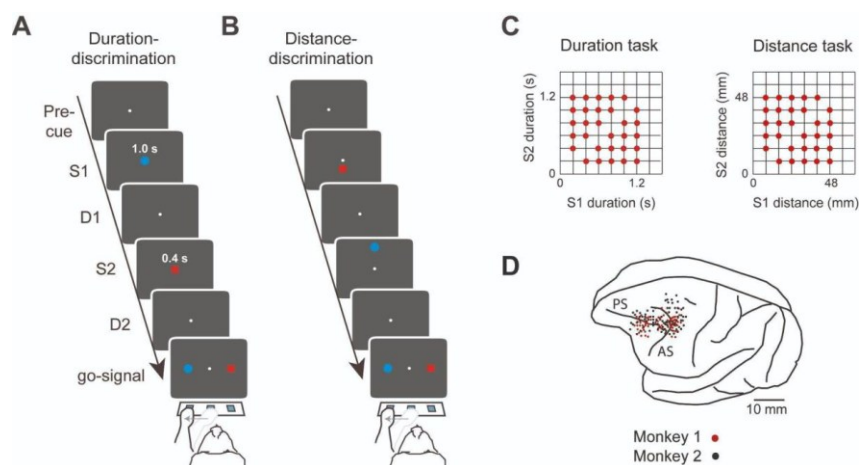


Figura 1.1 Compiti sperimentali, distribuzione degli stimoli e punti di registrazione

1.1.2 Risultati

Circa la metà dei neuroni registrati nella corteccia frontale dei macachi ha formato almeno un assemblaggio cellulare (una "squadra" coordinata) sia nel compito di distanza che in quello di durata. Nello specifico, il 25,1% e il 27,1% delle coppie di neuroni possibili si sono rivelate veri e propri assemblaggi coordinati.

Quando gli scienziati hanno confrontato le stesse squadre di neuroni attraverso i due diversi giochi (tempo vs spazio), hanno scoperto che nel 97% dei casi l'ordine sequenziale con cui i neuroni si attivavano l'uno dopo l'altro rimaneva lo stesso. Questo suggerisce che la sequenza temporale interna a un assemblaggio è fissa e indipendente dal tipo di compito eseguito.

Introducendo il metodo di filtraggio, i ricercatori hanno separato l'attività totale del neurone (**FSA - Full-Spikes Activity**) dall'attività registrata esclusivamente quando il neurone comunicava in sinergia con la sua squadra (**ASA - Assembly-Spikes Activity**).

Questo filtraggio ha rivelato che l'attività "in squadra" (ASA) possiede proprietà di codifica uniche. Un singolo neurone, partecipando in modo flessibile a più assemblaggi diversi, contribuisce a codificare variabili e informazioni del tutto nuove, che non sarebbero altrimenti visibili analizzando il neurone da solo (in isolamento).

Persino i neuroni inizialmente classificati come "non discriminativi" (ovvero che presi singolarmente sembravano non elaborare alcuna informazione utile per il compito) hanno acquisito una spiccata selettività e importanza nel momento in cui la loro attività veniva analizzata all'interno dell'azione collettiva dell'assemblaggio.

Lo studio dimostra formalmente che l'impatto reale di un neurone sull'elaborazione e sulla propagazione delle informazioni nel cervello dipende strettamente dal circuito e dal pattern di connessione sinaptica in cui è inserito in quel momento. Per questo motivo, l'unità fondamentale per l'elaborazione delle informazioni nel cervello è l'assemblaggio cellulare, non il singolo neurone.

I neuroni mostrano una straordinaria flessibilità d'azione. Comportandosi come un maestro di scacchi che gioca una partita simultanea su più scacchiere, lo stesso neurone partecipa contemporaneamente a diverse squadre (assemblaggi), elaborando informazioni distinte per scopi differenti.

Il cervello umano e animale cambia continuamente il modo di rappresentare le informazioni nel tempo (codifica dinamica). Gli autori concludono che questa dinamicità è dovuta proprio alla riconfigurazione della rete: quando la scimmia passa dal semplice *ricordo* di un obiettivo alla *trasformazione* di quel ricordo in un'azione motoria, i neuroni si riconfigurano rapidamente migrando da un assemblaggio all'altro. Al contrario, una riconfigurazione minima o assente è associata a fasi di memoria statiche.

1.2 Metodi per l'identificazione di cell assemblies

L'elaborazione delle informazioni nel cervello (in particolare nella corteccia) è organizzata attraverso **cell assemblies** (assemblamenti cellulari/gruppi neurali), concetto introdotto originariamente da Donald Hebb nel 1949. Si tratta di gruppi di neuroni che cooperano agendo come singole unità funzionali o di elaborazione.

Quando questi assemblamenti si attivano, i neuroni che ne fanno parte tendono a coordinare la loro attività elettrica (i loro impulsi, o *spike*) con una precisione temporale che può arrivare al millisecondo.

Grazie ai recenti progressi tecnologici nell'elettrofisiologia, i neuroscienziati sono oggi in grado di registrare l'attività di 100 o più neuroni contemporaneamente (i cosiddetti *massively parallel spike trains* o MPST). Se da un lato questo aumenta le probabilità di osservare l'attività di assemblamenti neuronali complessi, dall'altro crea un enorme problema di analisi dei dati:

- All'aumentare dei neuroni registrati, il numero di possibili combinazioni e pattern temporali cresce in modo esponenziale. Diventa matematicamente e computazionalmente impossibile contare o testare direttamente ogni singolo pattern;
- Effettuare test statistici classici su un numero così elevato di combinazioni porterebbe a tantissimi falsi positivi (rilevare pattern coordinati che in realtà avvengono solo per puro caso).

L'obiettivo fondamentale dei metodi statistici trattati è proprio quello di "isolare il segnale dal rumore", andando a mappare e quantificare le sincronizzazioni biologiche nascoste nel tracciato. È cruciale sottolineare che tale ricerca non si limita alle sole correlazioni a coppie (*pairwise*), ovvero all'analisi di quanti e quali neuroni sparino a due a due in modo coordinato. Le funzioni cognitive superiori si basano infatti su sincronizzazioni di ordine superiore (*higher-order correlations*), che coinvolgono pattern complessi e combinazioni multiple di numerosi neuroni (ad esempio, l'attivazione simultanea o sequenziale dei neuroni A, B, C e D).

1.2.1 Tipologie di sincronizzazioni

La codifica temporale (*temporal coding*) è stata associata a diverse forme (ma non necessariamente incompatibili) di correlazione degli impulsi su scala temporale ad alta risoluzione, ovvero con precisione al millisecondo. Queste possono spaziare dalla sincronizzazione di gruppi di cellule diversi, a sequenze di impulsi provenienti da neuroni specifici in un ordine temporale preciso, fino a sequenze di attività sincrona. Ciascun metodo considerato nella rassegna che qui vediamo è stato progettato per determinare la presenza di una di queste strutture di correlazione nei dati MPST (*massively parallel spike trains*). Di conseguenza, è innanzitutto necessario introdurre i rispettivi modelli di correlazione ed evidenziarne le somiglianze e le differenze.

Presentiamo cinque diversi tipi di attività di scarica (*spiking*) con correlazione temporale ad alta risoluzione che, nella letteratura, sono stati associati ai meccanismi di codifica temporale, sia in studi teorici sia sperimentali.

1. **Sincronizzazione di popolazione:** La sincronizzazione di popolazione si riferisce a un'attività di scarica (*spiking*) in cui una parte (o la totalità) dei neuroni osservati emette impulsi sincroni, in modo ripetuto nel tempo. I neuroni coinvolti non sono necessariamente sempre gli stessi, sebbene possano esserlo.
2. **Sincronizzazione a coppie:** Nella sincronizzazione a coppie (*pairwise synchronization*), le coppie di neuroni sincronizzano i propri impulsi in modo indipendente l'una dall'altra. Di conseguenza, le correlazioni di ordine superiore sono assenti. I pattern di dimensione pari o superiore a 3 sono ancora possibili, tuttavia solo come risultato della coincidenza casuale di emissioni simultanee di impulsi da parte di singoli neuroni o di coppie di neuroni. Questo tipo di attività di scarica è mostrato nella Fig. 2A.

Negli studi che si occupano dell'analisi delle correlazioni degli impulsi, c'è stato e c'è tuttora un forte focus sull'analisi a coppie. Il motivo non risiede nel fatto che la teoria predica esclusivamente correlazioni a coppie, bensì nella semplicità di tali analisi rispetto a quelle delle

correlazioni di ordine superiore. Ciò nonostante, lo studio di correlazioni puramente a coppie può rivelare, nei dati MPST, interessanti strutture di dipendenza che accennano a gruppi interagenti di neuroni più ampi, interazioni tra aree diverse o interazioni spaziali.

3. **Spike patterns (modelli a picco) sincroni:** è stato dimostrato sperimentalmente che un neurone che riceve input sinaptici sincroni ha maggiori probabilità di emettere un impulso rispetto a quando riceve input che arrivano in modo asincrono. Questa osservazione ha portato a ipotizzare che i neuroni sincronizzino le proprie attività anche al di là delle semplici coppie.

Per indagare questa ipotesi nei dati reali tramite test statistici, nonché per generare dati sintetici per la validazione dei metodi, sono stati formulati modelli probabilistici di treni di impulsi paralleli che includono una sincronizzazione di ordine superiore a quello a coppie.

Due esempi sono il modello a interazione singola (*single interaction model*) di Kuhn et al. (2002) e il modello a massima entropia (*maximum entropy model*) di Schneidman et al. (2003). Una realizzazione del processo a interazione singola, in cui un pattern di impulsi sincroni presenta occorrenze multiple, è mostrata nella Fig. 2B.

4. **Pattern spazio-temporali:** La sincronia degli impulsi può essere generalizzata aggiungendo una dimensione temporale alla correlazione: i neuroni coinvolti nell'attività coordinata non generano necessariamente impulsi in modo sincrono, ma secondo specifiche sequenze temporali con ritardi fissi (fino a una data precisione) tra gli impulsi consecutivi (vedi Fig. 1D). Questo tipo di attività viene generalmente definito pattern spazio-temporale (STP; Prut et al. 1998).
5. **Sequenze di eventi di impulsi sincroni:** Un tipo specifico di correlazione temporale che caratterizza sia la sincronizzazione degli impulsi sia la propagazione temporale è rappresentato dalle **sequenze di eventi sincroni** (SSE, *sequences of synchronous events*). Queste consistono in molteplici eventi sincroni, ciascuno dei quali coinvolge un gruppo specifico di neuroni, che si verificano uno dopo l'altro con un ritardo temporale fisso. Le registrazioni parallele da più strati di una catena synfire attiva mostrerebbero, ad esempio, tali pattern di impulsi (Schrader et al. 2008; Gerstein et al. 2012). Gli insiemi di neuroni coinvolti nei diversi eventi sincroni possono o meno sovrapporsi. Una realizzazione di una specifica SSE che si verifica due volte è mostrata nella Fig. 1E.

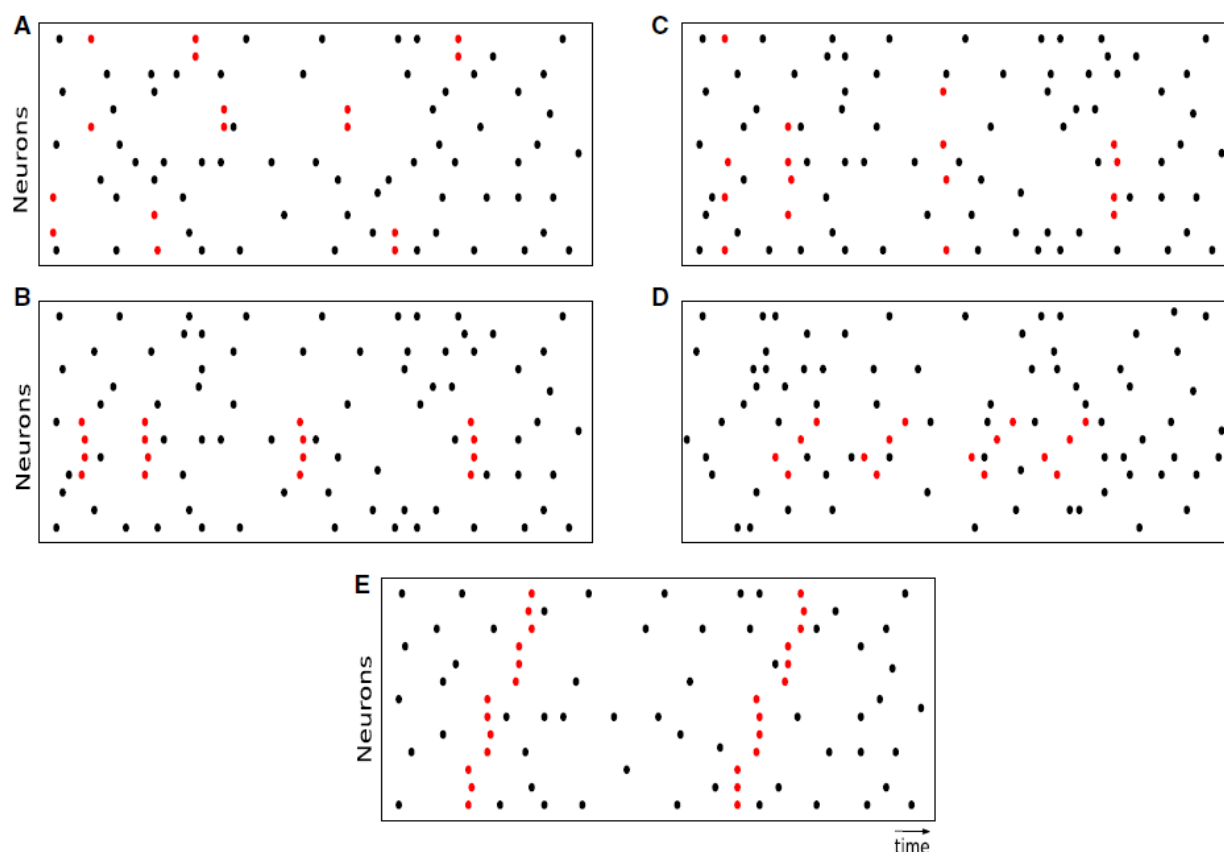


Figura 1.2 Raster plot di diversi tipi correlazioni. Ogni pannello mostra l'attività di scarica di treni di impulsi paralleli (un neurone per riga) nel corso del tempo (asse orizzontale). Ogni punto rappresenta un impulso (spike); i punti rossi, in particolare, rappresentano gli impulsi appartenenti a un pattern specifico. I diversi pannello si riferiscono a differenti forme di correlazione temporale degli impulsi.

a Modello di correlazione a coppie (Pairwise correlation). La popolazione contiene 6 coppie di neuroni sincronizzati (questi ultimi indicati dal basso verso l'alto): (1, 2), (1, 3), (2, 4), (8, 9), (8, 14), (13, 14). **b Pattern di impulsi sincroni (Synchronous spike patterns).** I neuroni 4, 5, 6, 7 sono coinvolti ripetutamente nel medesimo pattern. **c Sincronizzazione di popolazione (Population synchronization).** Diversamente dai pattern di impulsi del pannello b, i neuroni coinvolti in ciascun evento sincrono sono selezionati casualmente e cambiano da un evento all'altro. **d Pattern spazio-temporali (Spatio-temporal patterns).** Gli impulsi rossi corrispondono alle occorrenze di un STP. I neuroni coinvolti nei pattern sono i numeri 4, 5, 6, 7 (come nel pannello b), ma i loro impulsi si verificano ora in una successione temporale fissa con ritardi costanti. **e Sequenze di eventi di impulsi sincroni (Sequences of synchronous spike events).** Vengono mostrate due occorrenze della stessa SSE. In questo caso, tutti i neuroni osservati sono coinvolti, e gruppi di 4-4-4-3 neuroni che scaricano in modo sincrono si attivano in rapida successione.

1.2.2 Classificazione dei metodi di analisi: approcci puntuali e metodi di popolazione

In generale, si possono distinguere due gruppi di metodi che indicano diversi aspetti della correlazione degli impulsi:

- **Primo gruppo:** Mira a identificare l'ordine di correlazione (il numero di neuroni coinvolti) piuttosto che l'identità specifica dei neuroni stessi. Di conseguenza, ogni evento correlato può coinvolgere un sottoinsieme casuale di neuroni o essere composto da un gruppo specifico e sempre identico. Definiamo il tipo di correlazione alla base di questa classe di analisi come *sincronizzazione di popolazione*.

- **Secondo gruppo:** Presuppone un modello di correlazione in cui i neuroni correlati formano eventi di impulsi sincroni stereotipati o sequenze temporali di impulsi. Ci riferiamo a questi eventi come *spike patterns*.

L'obiettivo di questi metodi è recuperare la composizione neuronale e i tempi di occorrenza, ovvero l'istante temporale (espresso in millisecondi o secondi) in cui i neuroni del gruppo decidono di attivarsi insieme, degli *spike patterns*.

1) Metodi per rilevare la sincronizzazione di popolazione

Una delle sfide nella valutazione statistica degli eventi di impulsi sincroni nei dati MPST (*massively parallel spike trains*) è rappresentata dalla crescita esponenziale del numero di possibili pattern al crescere del numero di neuroni considerati. Tuttavia, questo problema può essere semplificato se l'interesse della ricerca risiede unicamente nel valutare la presenza e l'ordine di una sincronizzazione in eccesso (cioè superiore a quella casuale), senza determinare le identità specifiche dei neuroni coinvolti. Inoltre, i dati potrebbero contenere pattern di impulsi sincroni che cambiano la propria composizione neuronale ogni volta, facendo sì che la correlazione sia distribuita potenzialmente sull'intera popolazione osservata. Ci riferiamo alla sincronia che non coinvolge (o che si ipotizza non coinvolga) sottogruppi specifici di neuroni nella popolazione osservata come **sincronizzazione di popolazione**.

La maggior parte dei metodi per l'analisi della sincronizzazione di popolazione riduce i dati degli impulsi al numero di neuroni attivi (cioè agli impulsi) osservati in ogni intervallo di tempo (*time bin*). Uno spike train è descritto interamente dai suoi tempi di scarica e, data una discretizzazione temporale in piccoli intervalli, possiamo definire l'**istogramma di popolazione** come il conteggio degli impulsi che si sono verificati nel medesimo intervallo di tempo. Il conteggio massimo possibile dell'istogramma è quindi il numero N dei neuroni.

L'astrazione computazionale legata alla scelta di questa ampiezza di campionamento non è arbitraria, ma risponde a precisi vincoli di plausibilità biologica radicati nelle scale temporali intrinseche del sistema nervoso centrale. Nello specifico, la durata di un singolo passo temporale (*time-step* o *bin*) della simulazione è stata calibrata nell'ordine dei **10–15 millisecondi**. Questa impostazione trova un solido fondamento nella fisiologia neuronale e, in particolare, nelle dinamiche del **periodo refrattario**. I dati biologici evidenziano infatti che il tempo tipico necessario a un neurone per ripristinare il proprio potenziale di membrana a riposo — e poter quindi generare un nuovo potenziale d'azione, ovvero *ripartire* — si attesta intorno ai 15 millisecondi, se si considera l'estensione del periodo refrattario relativo (a fronte di un refrattario assoluto più breve, pari a circa 1–2 ms).

La definizione di questa finestra temporale discreta permette inoltre di sintetizzare computazionalmente i diversi ordini di grandezza delle latenze biologiche che regolano la comunicazione cellulare:

1. **Latenza sinaptica:** Il ritardo intrinseco tra l'arrivo dell'impulso elettrico al terminale presinaptico e la generazione dell'effetto sul neurone successivo (legato ai tempi biologici di rilascio, diffusione e legame dei neurotrasmettitori) presenta valori tipici stimati tra **1 e 5 millisecondi**.
2. **Latenza di integrazione:** Un neurone non risponde istantaneamente a un singolo stimolo, ma opera una sommazione spaziale e temporale degli input eccitatori e inibitori ricevuti lungo

l'albero dendritico all'interno di una determinata "finestra temporale" di eccitabilità. Questo tempo di elaborazione e decisione cellulare varia in base allo stato della membrana, richiedendo **da pochi millisecondi fino a decine di millisecondi**.

3. **Latenza di propagazione:** Il tempo impiegato dal segnale per viaggiare lungo l'assone presenta un'analogia variabilità (da **meno di 1 millisecondo a decine di millisecondi**), essendo strettamente dipendente dalla lunghezza della fibra nervosa e dalla presenza o meno della guaina mielinica.

Di conseguenza, impostare il *time-step* del modello in una finestra compresa tra i 10 e i 15 ms consente di racchiudere in un unico ciclo computazionale discreto la complessa cascata di micro-eventi biologici (latenza sinaptica, integrazione dei segnali e tempo di volo assonale), riflettendo in modo macroscopico il limite fisiologico imposto dalla frequenza massima di scarica dei singoli nodi.

I primi tre metodi qui presentati si basano su statistiche derivate dall'istogramma di popolazione. Sono stati sviluppati in successione, ciascuno per superare i limiti del precedente:

- **Analisi della Distribuzione di Complessità (CD - *Complexity Distribution*):** Sviluppata da Grün et al. (2008), propone un approccio statistico semplice basato puramente sulla distribuzione delle voci dell'istogramma di popolazione. Confronta tale distribuzione derivata empiricamente con quella attesa da neuroni che scaricano in modo indipendente per determinare la presenza di sincronizzazione in eccesso.
- **Inferenza della Correlazione Basata sui Cumulanti (CuBIC - *Cumulant-Based Inference of Correlation*):** Sviluppata da Staude et al. (2010a), deriva la distribuzione nulla analiticamente sotto assunzioni più specifiche riguardo ai dati, e inferisce l'ordine minimo di correlazione esistente nei dati stessi.
- **Analisi dell'Evento Unitario di Popolazione (PUE - *Population Unitary Event*):** Sviluppata da Rostami (2017), opera sotto le stesse assunzioni di CuBIC, ma utilizza una statistica di test differente che aumenta la potenza statistica del test. In questo modo, richiede campioni di dimensioni ridotte per una corretta identificazione della sincronia in eccesso, consentendo così anche un'analisi risolta nel tempo.

Il quarto metodo, denominato **indice di informazione della correlazione (CII - *Correlation Information Index*)**, è un approccio originariamente suggerito da Schneidman et al. (2006) come modo per condensare in un singolo scalare l'informazione fornita dai modelli a massima entropia costruiti su dati di treni di impulsi paralleli. Il metodo tiene conto dell'identità neuronale di ciascun impulso nei pattern sincroni osservati e costruisce un modello probabilistico completo di questi ultimi. Questo modello viene utilizzato per ottenere un singolo valore scalare, il CII, che quantifica la quantità di informazione in eccedenza contenuta nei dati che viene fornita dalle correlazioni di un determinato ordine.

2) Metodi per il rilevamento di pattern di spike

Il secondo gruppo di metodi è progettato per rilevare gruppi di neuroni coinvolti in pattern di impulsi (*spiking patterns*) precisi al millisecondo. Questi metodi raggiungono tale obiettivo individuando pattern di impulsi che si ripetono un numero di volte sufficientemente elevato da poter essere classifi-

cati come non casuali. I pattern non casuali sono considerati una firma dell'attivazione di un assemblaggio cellulare (Abeles 1991) e sono stati associati al comportamento in diversi studi sperimentali (ad es., Prut et al. 1998).

L'elevato numero di possibili pattern nelle registrazioni su vasta scala pone spesso problemi computazionali e statistici non indifferenti. Per avere un'idea di questo problema, si consideri una popolazione di N neuroni registrati in parallelo. Questi neuroni possono organizzare la propria attività in un numero massimo di pattern di impulsi sincroni pari a 2^N , che si avvicina a 10^{30} per $N = 100$, una quantità regolarmente disponibile nelle moderne registrazioni extracellulari. Questo numero aumenta di ordini di grandezza se si considerano gli STP (pattern spazio-temporali) arbitrari e non solo gli eventi sincroni.

Senza alcuna conoscenza preliminare sui neuroni potenzialmente coinvolti nella correlazione, è necessario eseguire una ricerca alla cieca dei pattern che si verificano più di quanto atteso sotto una certa ipotesi nulla, tenendo conto di tutte queste possibilità. L'onere computazionale può essere eccessivo (persino l'allocazione in memoria dei conteggi delle occorrenze di tutti i pattern possibili potrebbe risultare impossibile). Inoltre, testare singolarmente la significatività statistica di ogni pattern comporterebbe problemi insormontabili legati ai test multipli. Infine, la quantità di dati necessaria per raccogliere prove statistiche adeguate sarebbe immensa e, molto probabilmente, non disponibile.

I metodi qui considerati sono stati sviluppati proprio per affrontare tali problematiche.

All'inizio della neuroscienza computazionale (anni '60-'80), il numero di neuroni registrati contemporaneamente era molto basso (da 2 a 5 al massimo). I metodi sviluppati in questa fase erano pensati per scovare relazioni semplici:

- **Cross-Correlation Histograms (CCH):** Istogrammi di correlazione incrociata. Calcolano la probabilità che un neurone B spari dato lo spike di un neurone A, mostrando visivamente se c'è una sincronia o un ritardo fisso tra i due.
- **Joint Peristimulus Time Histogram (JPSTH):** Una griglia bidimensionale che permette di vedere come la dinamica di sincronizzazione tra due neuroni si evolve nel tempo in risposta a uno stimolo esterno.

Questi metodi sono strettamente **pairwise (a coppie)**. Se si provasse ad applicarli a 100 neuroni contemporaneamente, il numero di coppie da analizzare (e i relativi grafici) diventerebbe ingestibile, rendendo impossibile rilevare pattern in cui sparano 4 o 5 neuroni insieme.

Per superare il limite delle coppie e cercare sequenze temporali precise (il concetto di *"io emetto perché tu hai emesso prima"*), sono nati algoritmi capaci di scansionare tracciati paralleli alla ricerca di ritardi fissi tra molteplici neuroni:

- **L'approccio di Abeles (Synfire Chains):** Algoritmi pionieristici per cercare onde di attività che si propagano nel circuito con ritardi millimetrici costanti.
- **Template Matching (Rilevazione a stampo):** Questo metodo definisce un "pattern ideale" (un modello di sequenza di sparo) e scansiona l'intero tracciato per vedere quante volte quel modello esatto si ripete nel tempo.

Il limite è che richiede che il ricercatore sappia già, a grandi linee, che tipo di pattern sta cercando, oppure richiede un tempo di calcolo immenso per testare ogni possibile combinazione di ritardi.

Frequent Itemset Mining (FIM) è la famiglia di metodi più recente e potente descritta nel paragrafo, introdotta per risolvere il problema delle registrazioni massive (100+ neuroni). Invece di testare le combinazioni alla cieca, questi metodi usano algoritmi mutuati dall'informatica aziendale (quelli usati, ad esempio, per capire quali prodotti vengono comprati più spesso insieme nei supermercati). Il tracciato viene discretizzato in finestre temporali (*bin*). L'algoritmo (come il celebre *FP-growth*) setaccia i dati in modo estremamente efficiente, isolando istantaneamente solo i gruppi di neuroni che compaiono insieme con una frequenza minima (chiamata *support*). Tutto ciò che compare raramente viene potato via subito dall'albero di calcolo, salvando risorse computazionali.

Il metodo SPADE (Spike Pattern Detection and Evaluation): unisce il *Frequent Itemset Mining* (per trovare i pattern candidati in modo rapido) a un severo controllo statistico tramite dati surrogate (per eliminare i falsi positivi dovuti al rumore stocastico). SPADE è in grado di estrarre dal caos sia le sincronie pure (allineate verticalmente) sia i pattern spazio-temporali con ritardi (STP), dicendo esattamente quali neuroni li compongono.

1.3 Metodi di decomposizione e rilevamento di ensemble neuronali

Negli ultimi anni, i progressi tecnologici (come l'introduzione di sonde ad altissima densità come i *Neuropixels*) hanno permesso di registrare l'attività neurale su larghissima scala. Questo genera una mole immensa di dati elettrofisiologici extracellulari che richiedono algoritmi specifici per essere interpretati.

Vi sono due approcci sperimentali attualmente utilizzati in neuroscienze per catturare l'attivazione simultanea dei neuroni a diverse scale temporali:

- **Registrazioni elettrofisiologiche in-vivo e in-vitro:** Effettuate tramite microelettrodi multipli ad alta densità, offrono una risoluzione temporale finissima, nell'ordine dei millisecondi;
- **Calcium Imaging (Imaging del Calcio):** Permette di registrare ampi gruppi di neuroni identificandone con precisione la posizione spaziale sul lungo periodo. Ha però una risoluzione temporale più lenta, nell'ordine delle centinaia di millisecondi, a causa della durata dei transienti del calcio.

Sono stati ideati vari metodi per rilevare gli cell assemblies dalle registrazioni con microelettrodi e dai dati di *calcium imaging*, basati su una serie di quadri teorici e metodologici. Tuttavia, diverse tecniche esistenti sono state testate solo su registrazioni simulate o non sono state utilizzate al di fuori del laboratorio in cui sono state sviluppate.

Ora ci concentreremo sui metodi che sono stati utilizzati per rilevare gli cell assemblies nelle registrazioni elettrofisiologiche extracellulari *in-vivo* da microelettrodi multipli.

I primi metodi di rilevamento degli assemblaggi si concentravano sull'attività basata sugli eventi (*event-based*), identificando gruppi di neuroni con un'attività correlata a specifiche variabili comportamentali, come l'erogazione di una ricompensa, la posizione dell'animale nello spazio e stimoli sen-

soriali, introducendo potenzialmente dei bias. Più di recente, si è registrato un crescente interesse per i metodi in grado di rilevare gli *cell assemblies* indipendentemente dagli eventi ambientali, il che può essere particolarmente utile per esaminare l'attività cerebrale priva di chiari correlati comportamentali. Nessuno degli approcci che descriveremo richiede informazioni sui correlati comportamentali dei neuroni.

La maggior parte delle tecniche inizia con il rilevamento della co-attivazione di singole coppie di neuroni. Questa può poi essere espressa sotto forma di matrici di co-attivazione (come matrici di correlazione, di similarità o di connettività funzionale), che vengono successivamente utilizzate per estrarre relazioni funzionali multicellulari più complesse.

Ora discuteremo di tre approcci principali: la riduzione della dimensionalità (analisi delle componenti principali [PCA], analisi delle componenti indipendenti [ICA], fattorizzazione a matrice non negativa [NMF]), l'analisi di rete (rilevamento di comunità [community detection], processi di Markov) e la costruzione agglomerativa.

1.3.1 Rilevamento di cell assemblies mediante riduzione della dimensionalità

La riduzione della dimensionalità mappa dati ad alta dimensionalità in uno spazio a dimensionalità inferiore, ed è comunemente impiegata per l'analisi di dati su larga scala. In neuroscienze, un esempio di spazio ad alta dimensionalità è quello formato dall'attività di una popolazione neurale. Ogni neurone rappresenta una dimensione all'interno di questo spazio. Di conseguenza, una registrazione di 100 neuroni nel tempo rappresenterebbe uno spazio a 100 dimensioni e può essere ridotta a uno spazio più piccolo di, ad esempio, 10 dimensioni (o fattori). I 10 fattori risultanti mirano a preservare la maggior parte delle informazioni sulla co-attività neuronale presente nella registrazione originale. In sostanza, ogni dimensione nel nuovo spazio ridotto rappresenta un pattern ricorrente di attività della popolazione nella matrice originale (ad alta dimensionalità) degli *spike*.

Per le tecniche di riduzione lineare della dimensionalità, ogni nuova dimensione nello spazio ridotto corrisponde a una combinazione lineare di tutti i neuroni, in cui ciascun neurone ha un peso (*weight*) che quantifica il suo contributo a quel determinato fattore. L'interpretazione del pattern di popolazione ottenuto dipende dalla tecnica di riduzione della dimensionalità impiegata. Ad esempio, se una tecnica viene applicata a una matrice di correlazione calcolata con le correlazioni a coppie di Pearson, la presenza di più cellule che condividono pesi elevati all'interno di un dato fattore indicherebbe che esse possiedono un punteggio di correlazione di Pearson elevato.

1. PCA

La PCA riduce il dataset originale proiettandolo in un insieme di dimensioni definite da componenti principali (autovettori) tra loro ortogonali. La prima componente viene calcolata per spiegare la massima quota di varianza (co-attività neurale) dei dati, mentre le successive spiegano il resto della varianza sotto il vincolo della perpendicolarità matematica.

- **Vantaggi:** elevata velocità computazionale, scalabilità per dataset massivi e oggettività statistica (l'estrazione delle componenti significative è regolata da soglie matematiche precise, come la distribuzione di Marčenko-Pastur).
- **Limiti:** la rigidità geometrica dell'ortogonalità impedisce di mappare fedelmente la realtà biologica dei circuiti neurali, in cui un singolo elemento polivalente può pren-

dere parte a più assemblaggi simultanei (*overlapping assemblies*). La PCA tende infatti a escludere che lo stesso neurone abbia pesi (*weights*) elevati in componenti diverse.

2. ICA

L'ICA supera il vincolo geometrico della PCA cercando l'**indipendenza statistica** delle sorgenti attraverso la massimizzazione della loro *non-gaussianità* (secondo il principio del Teorema del Limite Centrale). Nella pipeline standard di rilevamento, la PCA viene usata come passo preliminare per eliminare il rumore e ridurre lo spazio, mentre l'ICA interviene successivamente per "ruotare" gli assi in modo non ortogonale.

- **Vantaggi:** è l'algoritmo ideale per identificare cell assemblies biologicamente sovrapposti, assegnando correttamente un singolo neurone a più network funzionali.
- **Limiti:** è un processo iterativo stocastico, il che lo rende sensibile alle condizioni di inizializzazione e suscettibile di convergere in minimi locali (restituendo assemblaggi leggermente diversi a ogni esecuzione). Inoltre, al pari della PCA, soffre dell'arbitrarietà del segno matematico dei pesi, che complica l'interpretazione dei neuroni con segno opposto all'interno della stessa componente.

3. NMF (Non-negative Matrix Factorization)

L'NMF affronta il problema da una prospettiva puramente fisiologica, decomponendo la matrice originale dei dati (ad esempio, l'attività di scarica dei neuroni nel tempo) nel prodotto di due matrici a dimensionalità ridotta, con il vincolo fondamentale che tutti gli elementi siano esclusivamente positivi o nulli.

- **L'idea dell'algoritmo e le due matrici:** Matematicamente, data una matrice iniziale V , l'algoritmo cerca di approssimarla tramite il prodotto $V \approx W \times H$. Vengono generate **due matrici** perché ciascuna mappa una componente essenziale del fenomeno biologico:
 - 1) La matrice dei pesi o delle basi (W):** definisce il profilo spaziale dell'ensemble, ovvero descrive l'identità dei neuroni che formano lo specifico modulo e "quanto" ciascun neurone contribuisca ad esso.
 - 2) La matrice dei coefficienti o di attivazione (H):** definisce il profilo temporale, ovvero indica "quando" e con quale intensità quel determinato modulo si attiva nel corso del tempo.
- **Come vengono create:** Le matrici vengono inizializzate con valori positivi casuali e poi aggiornate iterativamente attraverso algoritmi numerici (come le regole di aggiornamento moltiplicativo o i minimi quadrati alternati non negativi). L'obiettivo è minimizzare una funzione di costo (come la distanza di Frobenius o la divergenza di Kullback-Leibler) tra la matrice originale V e il prodotto $W \times H$, costringendo ad ogni passo i valori a rimanere non negativi.
- **Vantaggi:** Risolve l'ambiguità interpretativa di PCA e ICA. Da un punto di vista biologico, un neurone può aumentare la sua frequenza di scarica (valore positivo) o rimanere silente (valore zero), ma non può mostrare una scarica "negativa". L'assenza di pesi negativi permette una rappresentazione basata sulle parti (*parts-based representation*): i pattern estratti dall'NMF sono immediatamente interpretabili come moduli di attivazione diretta o parti isolate di circuiti neurali.
- **Limiti:** Non avendo vincoli di ortogonalità o di indipendenza statistica, l'NMF richiede algoritmi di ottimizzazione complessi e criteri di convergenza stringenti, risultando

talvolta meno immediato nella standardizzazione delle soglie statistiche per definire l'appartenenza di un neurone all'assemblaggio.

1.3.2 Rilevamento di cell assemblies tramite analisi di rete

L'analisi di rete rappresenta il secondo grande paradigma metodologico per l'isolamento degli cell assemblies. A differenza delle tecniche di riduzione della dimensionalità (PCA, ICA), questo approccio mappa l'attività neurale sfruttando la teoria dei grafi: i neuroni vengono trattati come **nodi** e le loro interazioni temporali (calcolate tramite matrici di similarità coseno o connettività funzionale) costituiscono gli **archi** (o collegamenti). In questo contesto, un assemblaggio cellulare corrisponde a una **comunità**, ovvero a un sottogruppo di nodi densamente interconnessi tra loro e debolmente collegati al resto del grafo.

Esistono principalmente tre famiglie di algoritmi basati su questo principio:

1. Rilevamento di Comunità tramite Somiglianza (Metodo STC)

Il metodo *Spike Train Communities* (STC), introdotto da Humphries (2011), converte i treni di *spike* in una matrice di somiglianza coseno, la quale viene scomposta in autovettori per raggruppare i neuroni tramite l'algoritmo *k-means*. Il numero ottimale di assemblaggi viene definito massimizzando le connessioni interne ai cluster e minimizzando quelle esterne.

- **Vantaggi e Limiti:** Non richiede parametri arbitrari oltre alla scala temporale e possiede varianti prive di bin (*binless*). Tuttavia, risente dell'instabilità stocastica del *k-means* (sensibilità ai minimi locali) e, soprattutto, impone un vincolo di esclusività: un singolo neurone può appartenere a **una sola comunità alla volta**, escludendo la sovrapposizione biologica delle cellule.

2. Stabilità di Markov (MSCD)

L'algoritmo *Markov Stability Community Detection* (MSCD), proposto da Billeh et al., modella matematicamente la connettività funzionale tenendo conto della direzionalità (asimmetria) del segnale e distinguendo l'impatto biologico dei neuroni eccitatori da quelli inibitori. Le comunità vengono individuate simulando un cammino casuale (*random walk*) sul grafo: le particelle tendono a rimanere intrappolate più a lungo all'interno dei compartimenti densamente connessi (gli assemblaggi).

- **Vantaggi e Limiti:** Offre un'eccezionale flessibilità grazie al concetto di "Tempo di Markov", che permette di variare la granularità dell'analisi (tempi brevi isolano molti piccoli assemblaggi, tempi lunghi macro-strutture). Di contro, presenta un costo computazionale molto elevato e soffre anch'esso del limite di esclusività del neurone per singolo tempo di scarica.

3. Modelli di Markov Nascosti (HMM)

Gli HMM estendono i processi stocastici introducendo delle variabili non misurate (nodi nascosti) che influenzano i dati osservabili (gli *spike*). L'algoritmo stima quali configurazioni nascoste (gli assemblaggi) spieghino meglio le co-attivazioni di popolazione (come i *population burst events*).

- **Vantaggi e Limiti:** Rappresenta lo strumento più versatile di questa sezione: **consente la partecipazione di un neurone a più assemblaggi simultanei**, può rilevare sequenze temporali di attivazione e non pone limiti matematici al numero di assemblaggi scoperti. Il limite principale risiede nella necessità, per l'utente, di stimare a priori il numero di nodi nascosti (un errore di sovrastima genera artefatti integrando il rumore), oltre a richiedere tempi di calcolo estremamente lunghi.

1.3.3 Costruzione agglomerativa di gruppi

A differenza dei metodi di riduzione della dimensionalità (che comprimono l'intero spazio dei dati) e di quelli di rete (che segmentano un grafo preesistente), l'approccio agglomerativo opera con una logica *bottom-up* (dal basso verso l'alto). L'algoritmo parte dall'elemento costitutivo più piccolo — una coppia di neuroni significativamente coattivi — e progressivamente espande il gruppo. Attraverso un processo iterativo, cerca nello spazio dei dati altri singoli neuroni che mostrino una co-attivazione statistica significativa con il nucleo originario (n cellule), inglobandoli per formare un modello più ampio ($n+1$ cellule). Il processo si arresta solo quando nessuna nuova cellula soddisfa i criteri di inclusione.

L'Algoritmo CAD (Cell Assembly Detection): Sviluppato da Russo e Durstewitz, il metodo CAD applica questo principio analizzando i conteggi degli *spike* a diversi ritardi temporali (*time lags*). Per neutralizzare il problema delle stazionarietà — ovvero le fluttuazioni globali della frequenza di scarica dovute a stati comportamentali o oscillazioni cerebrali che potrebbero falsare le correlazioni — il CAD introduce un innovativo **test di asimmetria temporale**. L'algoritmo confronta la sequenza di scarica nel tempo reale (*forward pass*) con quella riprodotta all'indietro (*reverse pass*): se due neuroni sono indipendenti, la probabilità che A spari dopo B è identica a quella che B spari dopo A. Uno sbilanciamento statistico in una direzione indica una relazione funzionale autentica.

- **Vantaggi:** Consente di scoprire un numero virtualmente illimitato di assemblaggi (pari a tutte le combinazioni probabilistiche tra i neuroni registrati) e permette la **sovrapposizione biologica** (*overlapping*), poiché lo stesso neurone può far parte di più gruppi. Inoltre, identifica l'identità esatta di ciascun membro, mappa le attivazioni sequenziali ed è estremamente robusto contro i bias derivanti dai tassi di scarica globali.
- **Limiti:** La sensibilità del metodo è strettamente legata alla frequenza di attivazione dell'assemblaggio: se un gruppo esteso si attiva raramente durante la registrazione, l'algoritmo perde potere statistico e rischia di frammentarlo, rilevando solo piccoli sotto-gruppi isolati. Infine, soffre del rischio di inclusione impropria: un neurone potrebbe essere aggregato a un intero assemblaggio solo perché fortemente correlato a un singolo sotto-insieme di esso, e non alla struttura nel suo complesso.

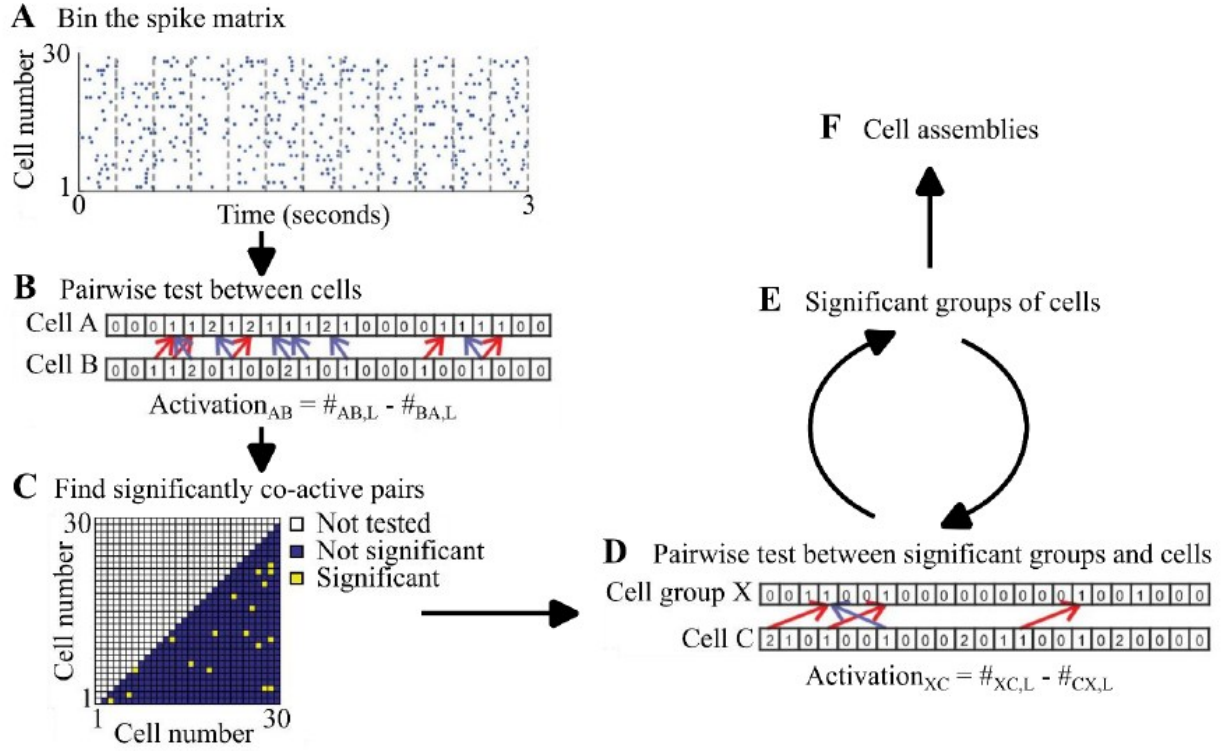


Figura 1.2

Il rilevamento agglomerativo degli assemblaggi di Russo e Durstewitz [adattato da 6].

A. Il dataset composto da N cellule viene suddiviso in finestre temporali (binned) in base a un intervallo dt impostato dall'utente. **B.** Viene valutata la co-attività di tutte le coppie di cellule presenti nel dataset. $\#_{AB,L}$ rappresenta il numero di volte in cui la cellula B è attiva dopo la cellula A al ritardo temporale (time lag) L ; $\#_{BA,L}$ rappresenta il numero di volte in cui la cellula A è attiva dopo la cellula B al ritardo temporale L . Activation_{AB} costituisce la differenza tra i due valori. **C.** Vengono determinate le coppie di neuroni significativamente coattive. **D-E.** I gruppi cellulari vengono progressivamente espansi. Il test a coppie viene eseguito tra i gruppi cellulari significativi X individuati nell'iterazione precedente e ciascuna cellula rimanente. Le cellule vengono aggiunte al gruppo cellulare se risultano significativamente coattive con il gruppo stesso. **F.** Gli assemblaggi cellulari vengono ottenuti quando non è più possibile aggiungere nuove cellule ai gruppi cellulari.

2. Identificazione di gruppi dinamici

In questo capitolo viene presentato il quadro teorico e metodologico sviluppato precedentemente a questa tesi per identificare e caratterizzare sottoinsiemi di variabili coordinate all'interno di sistemi complessi. L'obiettivo principale è superare i limiti computazionali e interpretativi delle metriche tradizionali, introducendo un approccio innovativo basato sulla teoria dell'informazione e su procedure algoritmiche di selezione.

Il percorso si articola attraverso quattro sezioni fondamentali:

- **Nozioni di base (2.1):** Vengono definiti i concetti cardine di entropia e integrazione per un vettore casuale, ponendo le basi formali e descrivendo il comportamento asintotico e i bias delle loro controparti empiriche.
- **Cluster Index (2.2):** Si esamina l'Indice Cluster (CI), una misura storicamente introdotta nelle neuroscienze per individuare gruppi di variabili funzionalmente integrati ma isolati dal resto del sistema, evidenziandone i punti di forza e le complessità applicative.
- **Metodologia proposta (2.3):** Viene introdotto l'approccio, che impiega lo *z-score* della sola integrazione unito a un algoritmo iterativo di setaccio (*sieving algorithm*) per mappare l'organizzazione gerarchica del sistema in macro-strutture non sovrapposte.
- **Versione standardizzata dell'integrazione (2.4):** Si approfondisce la derivazione statistica della metrica proposta, dimostrando come, grazie al teorema di Wilks, sia possibile approssimare asintoticamente la distribuzione dell'integrazione empirica tramite una variabile chi-quadro, azzerando di fatto la necessità di lunghe simulazioni numeriche.

2.1 Nozioni di base

Sia $X_S = \{X_u\}_{u \in S}$ un vettore casuale indicizzato da un insieme finito di elementi $S = \{u_1, \dots, u_k\}$ in un qualche sistema di interesse.

Assumiamo che $\emptyset \subset S \subset U$, dove $U = \{1, \dots, q\}$ è l'insieme di tutti gli elementi del sistema e $\emptyset$ denota l'insieme vuoto.

Come caso particolare, se $S = \{u_1\}$ si ha che $X_{\{u_1\}} = X_{u_1}$ è una variabile casuale corrispondente a un singolo elemento del sistema. L'idea è che il vettore $X_U = \{X_u\}_{u \in U}$ modelli una lettura affetta da rumore dello stato del sistema e che, considerando X_S ci si concentri sul sottosistema composto dagli elementi S .

L'entropia di S è definita come:

$$H(S) = - \sum_{x_1 \in \mathcal{X}_1} \dots \sum_{x_k \in \mathcal{X}_k} p_S(x_1, \dots, x_k) \log p_S(x_1, \dots, x_k) \quad (1)$$

dove $\mathcal{X}_j$ è l'alfabeto della variabile X_{u_j} , per $j = 1, \dots, k$, e

$$p_S(x_1, \dots, x_k) = P\{X_{u_1} = x_1, \dots, X_{u_k} = x_k\}, \quad x_1 \in \mathcal{X}_1, \dots, x_k \in \mathcal{X}_k \quad (2)$$

è la funzione di massa di probabilità (*probability mass function*) di X_S . La base del logaritmo utilizzato nella (1) determina l'unità di informazione: un bit, se si utilizzano logaritmi binari, o un nat, se si utilizzano logaritmi naturali. Se H_b è l'entropia con logaritmo in base b , allora $H_b = \log_b(e) H_e$. Nel seguito utilizzeremo i nat (salvo diversa indicazione).

La formula dell'entropia può essere riscritta attraverso la formulazione di Shannon:

$$H(S) = - \sum_{x_k \in \mathcal{X}_k} p(x_k) \log p(x_k)$$

Quando $S = \{u_1, \dots, u_k\}$ con $k \geq 2$, possiamo definire l'integrazione di X_S come

$$I(S) = -H(S) + \sum_{j=1}^k H(u_j) \quad (3)$$

Dove $H(u_j)$ è l'entropia della variabile X_{u_j} .

Sia $X_S^{(1)}, \dots, X_S^{(n)}$ un campione casuale di osservazioni provenienti dalla distribuzione sconosciuta di X_S . Stimiamo la distribuzione di X_S per mezzo della distribuzione empirica di $X_S^{(1)}, \dots, X_S^{(n)}$, la cui funzione di massa di probabilità $\hat{p}_S(x_1, \dots, x_k)$, $x_1 \in \mathcal{X}_1, \dots, x_k \in \mathcal{X}_k$, è data dalle frequenze relative dei valori possibili di X_S nel campione. Utilizzando questa distribuzione nella (1), otteniamo l'entropia empirica di $X_S^{(1)}, \dots, X_S^{(n)}$, che denotiamo con $H_n(S)$.

Analogamente, utilizzando le entropie empiriche nella (3), definiamo l'integrazione empirica $I_n(S)$ di $X_S^{(1)}, \dots, X_S^{(n)}$.

È noto fin dai tempi di Miller e Madow [16] che, in media, l'entropia empirica $H_n(S)$ sottostima la sua controparte teorica $H(S)$:

$$E[H_n(S)] = H(S) - \frac{c(S) - 1}{2n} + O\left(\frac{1}{n^2}\right) \quad (4)$$

dove $c(S) = \prod_{j=1}^k |\mathcal{X}_j| > 1$ è il numero di celle nella tabella indicizzata da S e si sottintende che $p_S(x_1, \dots, x_k) > 0$ per tutti gli $x_1 \in \mathcal{X}_1, \dots, x_k \in \mathcal{X}_k$; si noti che $|\mathcal{X}_j|$ denota la cardinalità di $\mathcal{X}_j$.

Ne consegue che l'integrazione empirica sovrastima la sua controparte teorica:

$$E[I_n(S)] = I(S) + \frac{d_k}{2n} + O\left(\frac{1}{n^2}\right) \quad (5)$$

dove

$$d_k = c(S) - 1 + k - \sum_{j=1}^k |\mathcal{X}_j| \quad (6)$$

è un intero strettamente positivo. Nel caso speciale $k = 2$, la quantità nella (6) può essere scritta come $d_2 = (|\mathcal{X}_1| - 1)(|\mathcal{X}_2| - 1)$.

2.2 Cluster Index

L'Indice Cluster (*Cluster Index* - *CI*) è una misura proposta negli anni '90 da Tononi e Edelman [9,10], all'interno delle ricerche sui processi cerebrali umani, il cui scopo è l'identificazione di sottoinsiemi di variabili in un sistema dinamico che si comportano in modo coordinato, pur avendo uno scambio relativamente limitato di informazioni con il resto del sistema; questi sottoinsiemi possono poi essere utilizzati per descrivere l'organizzazione dell'intero sistema.

Dato un sottoinsieme $S = \{u_1, \dots, u_k\}$ di elementi in un sistema indicizzato da $U = \{1, \dots, q\}$, dove $1 < k < q$, Tononi et al. [10] hanno definito l'integrazione di S come nella (3) e l'hanno considerata come un'approssimazione del grado di coordinazione tra le k variabili indicizzate da S . Successivamente, Tononi et al. [9] hanno considerato la mutua informazione tra S e $U \setminus S$

$$M(S) = H(S) + H(U \setminus S) - H(S) \quad (7)$$

come misura della dipendenza reciproca tra il sottoinsieme S e il resto del sistema $U \setminus S$, e hanno definito il $C(I)$ come il rapporto tra l'integrazione di S e la mutua informazione tra S e $U \setminus S$

$$C(I) = \frac{I(S)}{M(S)} \quad (8)$$

In pratica, le analisi ricorrono alla versione empirica $CI_n(S) = I_n(S)/M_n(S)$, basata sulle frequenze relative osservate in un campione di stati del sistema.

Valori elevati della (8) corrispondono a sottoinsiemi di U in cui la coordinazione interna supera lo scambio di informazioni con il resto del sistema, consentendo in tal modo l'identificazione di gruppi interessanti di variabili. Poiché il valore del $C(I)$ dipende sia dalle dimensioni del gruppo in esame sia dalle dimensioni del sistema, questo indice deve essere normalizzato prima che gruppi di dimensioni diverse possano essere confrontati.

Tononi et al. [10] hanno proposto come costanti di normalizzazione per il numeratore e il denominatore della (8) le medie dell'integrazione e della mutua informazione calcolate su gruppi di eguale dimensione integrati in un sistema, chiamato *sistema omogeneo*, avente lo stesso numero di variabili e nessuna organizzazione dinamica:

$$CI_n^*(S) = \frac{I_n(S) - \langle I_n(k) \rangle_0}{M_n(S) - \langle M_n(k) \rangle_0} \quad (9)$$

Infine, lo z-score del $C(I)$ normalizzato (identico numericamente allo z-score del $C(I)$) può essere utilizzato come indice di evidenza (*evidence index*):

$$zCI_n(S) = \frac{CI_n^*(S) - \langle CI_n^*(k) \rangle_0}{sd_0(CI_n^*(k))} = \frac{CI_n(S) - \langle CI_n(k) \rangle_0}{sd_0(CI_n(k))} \quad (10)$$

Infine, ed è interessante notare, l'indice $zC(I)$ è correlato all'identificazione della criticità dinamica nei sistemi complessi.

L'implementazione dell'indice $C(I)$ però presenta due sfide principali:

1. La cardinalità dell'insieme di tutti i possibili sottoinsiemi di un insieme è enorme. Tuttavia, ancor prima di considerare lo sforzo computazionale necessario per gestire questa mole di dati, è fondamentale notare che questo ampio insieme contiene molti gruppi inclusi l'uno

nell'altro e un numero molto grande di gruppi parzialmente sovrapposti. Tutte queste situazioni richiedono ulteriori analisi per valutarne l'effettiva rilevanza o indipendenza. Infatti, un valore elevato dell'indice non è condizione sufficiente per caratterizzare un sottoinsieme come rilevante, poiché tale valore potrebbe derivare dalla presenza di un sottoinsieme più piccolo caratterizzato da una coordinazione interna tra le variabili ancora più elevata. contrariamente, un insieme con un valore dell'indice elevato potrebbe raggiungere un valore ancora maggiore se ad esso venissero aggiunte altre variabili rilevanti.

2. È complesso calcolare le medie dell'integrazione e della mutua informazione su un sistema omogeneo adeguato. Sebbene le simulazioni numeriche a partire dal sistema omogeneo siano concettualmente semplici, esse devono essere ripetute per tutti i sottoinsiemi di interesse, il che si traduce in tempi di calcolo molto lunghi. Inoltre, per le simulazioni è necessario selezionare un sistema omogeneo specifico, il che introduce un grado indesiderato di arbitrarietà nell'analisi.

2.3 Metodologia proposta

Per ovviare a questi problemi, si utilizza l'integrazione da sola (attraverso il suo z-score, indicato nel seguito come zI).

L'uso della sola integrazione ci consente di superare la sfida (1) per mezzo di un'**approssimazione asintotica** che è valida per tutti i sistemi con variabili indipendenti e non richiede di effettuare simulazioni da nessuno di essi: per n grande (ampiezza del campione), l'integrazione moltiplicata per il doppio del numero di osservazioni segue approssimativamente una distribuzione chi-quadrato. I gradi di libertà di tale distribuzione dipendono dal numero di variabili appartenenti al sottogruppo analizzato e dalla cardinalità dei loro alfabeti; tale approssimazione può essere utilizzata per ottenere gli z-score con uno sforzo computazionale trascurabile.

L'impiego della sola integrazione risolve inoltre un aspetto problematico legato alla divisione per la mutua informazione nell'indice $C(I)$: valori bassi di $M(S)$ possono derivare sia da un basso scambio di informazioni tra il sottogruppo S e *ogni* elemento del resto del sistema, sia da un elevato scambio di informazioni tra S e una *piccola parte* del resto del sistema, mentre le altre parti non sono coinvolte nello scambio. Lo z-score del $C(I)$ non è in grado di distinguere tra queste due situazioni configurazionali.

La sfida (1) richiede una procedura per confrontare diversi sottogruppi. Una possibile soluzione è utilizzare una procedura di confronto (chiamata nel seguito *vaglio*, o algoritmo di setaccio, *sieving algorithm*) basata sulla seguente considerazione: se un insieme A è un sottoinsieme proprio di un insieme B e presenta un rango (valore dell'indice) più elevato rispetto a B , allora A deve essere considerato più rilevante di B . Pertanto, il setaccio trattiene solo quegli insiemi che non sono inclusi in, e non includono, nessun altro insieme con un valore di zI superiore. La procedura di confronto si compone quindi di due fasi fondamentali:

- i. Il rilevamento di insiemi di variabili rilevanti basato sul calcolo della metrica zI ;
- ii. L'applicazione dell'algoritmo di setaccio, che raffina e seleziona i risultati.

Questo approccio consente di identificare una plausibile organizzazione del sistema in termini di gruppi di variabili non sovrapposti (*non-overlapping*).

Al fine di analizzare l'organizzazione gerarchica del sistema in esame, si propone una versione iterativa del metodo di setaccio che raggruppa uno o più insiemi in un'unica entità per derivare una gerarchia. Il modo più semplice, ma efficace, per farlo consiste nell'eseguire iterativamente l'algoritmo di setaccio sui medesimi dati, utilizzando ogni volta una nuova rappresentazione in cui il sottoinsieme rilevante che ha ottenuto il punteggio più alto nell'iterazione precedente (in termini di valori di zI) viene considerato come atomico e sostituito da una singola variabile macroscopica (variabile di gruppo).

Ogni iterazione produce quindi un nuovo gruppo atomico di variabili: le iterazioni terminano quando il zI del sottoinsieme rilevante più alto in classifica scende al di sotto di una soglia prefissata, solitamente pari a 3.0 (ovvero, tre deviazioni standard dalla condizione di riferimento di indipendenza delle variabili).

L'approccio del setaccio iterativo evidenzia l'organizzazione di un sistema dinamico partizionandolo in insiemi di variabili rilevati a diverse iterazioni del vaglio. Al contempo, il processo di aggregazione, mediante l'aggiunta di nuovi elementi ai gruppi già esistenti, consente alla procedura di identificare le variabili del "resto del sistema" che scambiano informazioni con il sottoinsieme originariamente in esame. In questo modo, l'algoritmo riesce a discriminare tra le due situazioni problematiche di bassa mutua informazione descritte in precedenza. L'iterazione del processo di "setaccio e successiva aggregazione delle variabili" permette così di identificare le parti del sistema che possono essere aggregate in macro-strutture.

Infine, è importante notare che non si è interessati ad analizzare l'interezza dell'insieme delle parti del sistema in questione, bensì si vuole identificare esclusivamente i sottoinsiemi che presentano i valori massimi dell'indice scelto. È pertanto possibile utilizzare procedure di ottimizzazione numerica, il cui scopo è trovare i minimi o massimi globali senza dover procedere a un'enumerazione completa. A tale scopo, in passato sono state utilizzate diverse euristiche, tra cui varianti specifiche degli algoritmi genetici. Infine, è possibile affrontare (almeno parzialmente) il problema della maledizione della dimensionalità (*curse of dimensionality*) del sistema adottando strategie di calcolo parallelo.

2.4 Versione standardizzata dell'integrazione

Come anticipato, siamo interessati al calcolo della distribuzione dell'integrazione empirica $I_n(S)$ sotto l'ipotesi che le variabili indicizzate da S siano stocasticamente indipendenti (distribuzione nulla, o distribuzione del sistema omogeneo). Si determinerà un'approssimazione asintotica (per grandi campioni) che non dipende dalle distribuzioni marginali di $X_{u_1}, \dots, X_{u_k}$, ovvero dallo specifico sistema omogeneo scelto come benchmark per l'assenza di coordinazione tra le variabili.

L'osservazione chiave risiede nel fatto che i conteggi $n_s(x_1, \dots, x_k) = n\hat{p}_s(x_1, \dots, x_k)$ formano un vettore casuale multinomiale di dimensione n e probabilità di classe $p_s(x_1, \dots, x_k)$.

Ciò dà origine alla funzione di verosimiglianza:

$$L_n(p_s) \propto \prod_{x_1 \in \mathcal{X}_1} \dots \prod_{x_k \in \mathcal{X}_k} p_s(x_1, \dots, x_k)^{n_s(x_1, \dots, x_k)} \quad (11)$$

dove la costante di proporzionalità omessa è il coefficiente multinomiale che conta i modi di raggruppare n osservazioni in $c(S)n$ celle, con $n_s(x_1, \dots, x_k)$ osservazioni nella cella $x_1, \dots, x_k$.

Se le probabilità $p_s(x_1, \dots, x_k)$ sono libere di variare nel semplice standard, il punto di massimo della (11) (stimatore di massima verosimiglianza) è costituito dalle frequenze relative $\hat{p}_s(x_1, \dots, x_k)$. Ne consegue che l'entropia empirica $H_n(S)$ è uguale, a meno di una costante additiva, all'opposto della log-verosimiglianza media negativa massimizzata $-\bar{\ell}_n(\hat{p}_s) = -n^{-1} \log L_n(p_s)$:

$$\bar{\ell}_n(\hat{p}_s) = \sum_{x_1 \in \mathcal{X}_1} \dots \sum_{x_k \in \mathcal{X}_k} \hat{p}_s(x_1, \dots, x_k) \log p_s(x_1, \dots, x_k) \quad (12)$$

D'altra parte, se si introduce il vincolo di indipendenza stocastica $p_s(x_1, \dots, x_k) = \prod_{j=1}^k p_{u_j}(x_j)$, dove $p_{u_j}(x_j) = P\{X_{u_j} = x_j\}$ (assumendo cioè che le variabili indicizzate da S siano stocasticamente indipendenti), la funzione di verosimiglianza (11) può essere riscritta come:

$$L_n^0(p_s) \propto \prod_{x_1 \in \mathcal{X}_1} p_{u_1}(x_1)^{n_{u_1}(x_1)} \dots \prod_{x_k \in \mathcal{X}_k} p_{u_k}(x_k)^{n_{u_k}(x_k)} \quad (13)$$

dove $n_{u_j}(x_j)$ è il conteggio per il valore x_j nel campione marginale $X_{u_j}^{(1)}, \dots, X_{u_j}^{(n)}$, per $j = 1, \dots, k$. Il punto di massimo della (10) è chiaramente costituito dalle frequenze relative marginali $\hat{p}_{u_j}(x_j) = n_{u_j}(x_j)/n$, ed è evidente che la sommatoria $\sum_{j=1}^k H_n(u_j)$ è uguale a $-\bar{\ell}_n^0(\hat{p}_{u_1}, \dots, \hat{p}_{u_k}) = -n^{-1} \log L_n(\hat{p}_{u_1}, \dots, \hat{p}_{u_k})$ a meno della medesima costante additiva precedente.

In virtù di un teorema classico della statistica matematica dovuto a Wilks, e assumendo l'uso di logaritmi naturali, la *likelihood-ratio test statistic*

$$\Lambda_n = 2n \left(\bar{\ell}_n(\hat{p}_s) - \bar{\ell}_n^0(\hat{p}_{u_1}, \dots, \hat{p}_{u_k}) \right),$$

per n grande converge in distribuzione a una variabile casuale chi-quadro.

Poiché vale l'uguaglianza $I_n(S) = \bar{\ell}_n(\hat{p}_s) - \bar{\ell}_n^0(\hat{p}_{u_1}, \dots, \hat{p}_{u_k})$ la *likelihood-ratio test statistic* può essere riscritta $2nI_n(S)$ fornendo così una distribuzione asintotica per l'integrazione empirica.

Si noti che, per costruzione, la statistica del test del rapporto di verosimiglianza è superiormente limitata dallo zero (è non negativa) e si annulla quando $\hat{p}_s(x_1, \dots, x_k) = \prod_{j=1}^k \hat{p}_{u_j}(x_j)$, il che conferma le medesime proprietà analitiche per $I_n(S)$.

Nel caso particolare della mutua informazione tra due variabili, in cui i gradi di libertà possono essere scritti come $d_2 = (|\mathcal{X}_1| - 1)(|\mathcal{X}_2| - 1)$, il risultato sopra presentato risale a Luce [18], con una giustificazione diretta basata sul lavoro pionieristico di Wilks [30]. Se si utilizzano logaritmi binari, la distribuzione asintotica chi-quadro si applica alla statistica $2n \ln(2) I_n(S) \approx 1.3863n I_n(S)$; Wilks [18] utilizza il bit come unità di informazione ed esprime il risultato in questa forma. In generale, se b è la base utilizzata per i logaritmi, possiamo scrivere:

$$2n \ln(b) I_n(S) \xrightarrow{d} \chi^2(d_k)$$

dove $\xrightarrow{d}$ indica che la quantità a sinistra ammette il termine a destra come distribuzione asintotica (distribuzione limite per $n \rightarrow \infty$).

Poiché una distribuzione chi-quadrato con d_k gradi di libertà ha media pari a d_k e varianza pari a $2d_k$, la versione standardizzata dell'integrazione empirica è definita da:

$$zI_n(S) = \frac{2n \ln(b) I_n(S) - d_k}{\sqrt{2d_k}} \quad (14)$$

dove $b = e$ per i nat e $b = 2$ per i bit. Si noti che lo z-score approssimato nella (14) è coerente con la stima del bias analizzata nella (5), ma tiene conto anche della variabilità campionaria; esso riflette l'ampiezza del campione n e, tramite il termine d_k , sia il numero di variabili indicizzate da s sia la cardinalità dei loro alfabeti.

Alternativamente, è possibile confrontare sottoinsiemi di dimensioni diverse utilizzando l'indice normalizzato semplice $2nI/d = 2nI_n(S)/d_k$, oppure il complemento a uno del p-value chi-quadro di $2nI_n(S)$, denotato con $ChSq$.

L'indice zI è dunque una misura della distanza del comportamento del gruppo osservato dal comportamento che esso avrebbe avuto in caso di assenza di organizzazione fra le sue componenti, utilizzando come unità di misura la deviazione standard. Il campo della Fisica è abituato a "misurare" l'evidenza utilizzando proprio le deviazioni standard. Ad esempio, una distanza maggiore di 3 deviazioni standard indica che l'evento osservato avrebbe una probabilità di avvenire inferiore allo 0.3%, mentre 5 deviazioni standard corrispondono ad eventi con probabilità di accadimento inferiori a 1 su 1.744 milioni. Nel seguito del lavoro utilizzeremo spesso una soglia pari a 3 deviazioni standard - a volte 5 deviazioni standard.

3. Un modello di rete neurale

3.1 Modellizzazione della Rete Neurale e Dinamiche del Segnale

In analogia con le strutture biologiche descritte nei capitoli precedenti, il modello computazionale sviluppato in questo lavoro si propone di simulare le dinamiche di interazione e sincronizzazione all'interno di una rete neurale semplificata. L'obiettivo è riprodurre alcuni dei meccanismi fondamentali di trasmissione sinaptica e di integrazione dei segnali tipici del sistema nervoso centrale.

Il nucleo matematico del modello si basa sul principio biologico dell'integrazione dendritica. Ogni singolo neurone della rete riceve stimoli da molteplici nodi afferenti; l'attività del neurone ricevente viene quindi modellata attraverso una **sommatoria pesata degli ingressi** (input).

Formalmente, lo stato di attivazione di un nodo è determinato non solo dalla presenza del segnale in ingresso, ma dall'efficacia della connessione sinaptica stessa, espressa attraverso una matrice di pesi. Questo approccio consente di simulare la plasticità e la forza differenziale delle connessioni interneuronali, dove pesi maggiori rappresentano sinapsi più efficienti nel propagare l'eccitazione o l'inibizione.

A differenza dei modelli statici, l'evoluzione temporale della rete è gestita attraverso un processo di **iterazione discreta dei segnali**. Ad ogni passo temporale (time-step), il sistema aggiorna lo stato di attivazione di tutti i nodi in base agli input calcolati nel ciclo precedente.

Per approssimare in modo realistico la fisiologia cerebrale, in cui l'informazione non si propaga istantaneamente a causa dei tempi di conduzione assonale e del ritardo sinaptico, nel codice è stata introdotta una gestione esplicita dei **ritardi temporali**. Nello specifico, le connessioni sono caratterizzate da un tempo di volo del segnale, impostando un ritardo massimo pari a $\text{ritmax} = 1$. Questa limitazione impone che l'influenza di un neurone su un altro possa manifestarsi al passo temporale successivo, introducendo nella rete una memoria storica a breve termine e permettendo l'emergenza di sfasamenti temporali complessi.

Il codice implementato permette di valutare l'impatto di diverse topologie di rete (come strutture circolari a *closed-loop* o strutture gerarchiche ad albero) sulla propagazione del segnale.

Infine, per analizzare l'output di questa simulazione e identificare la formazione di *cell assemblies* (gruppi di neuroni che si attivano in modo coordinato), il modello prevede l'applicazione di una **finestra temporale mobile** (con avanzamenti discreti, ad esempio di 64 passi). Questo espediente matematico permette di passare da una logica di pura simultaneità a una di frequenza e ritmo, isolando le reali sincronizzazioni funzionali dal rumore di fondo causato dai ritardi intrinseci del sistema.

3.2 Casi di Studio e Topologie di Rete

Per valutare il comportamento del modello e analizzare la propagazione dei segnali binari, sono state implementate e testate differenti configurazioni strutturali (topologie di rete). L'obiettivo è osservare come la geometria delle connessioni influenzi la dinamica temporale del sistema e l'emergenza di pattern di attivazione coordinati.

Nello specifico, l'analisi si è concentrata su due casi di studio principali:

1. **Topologia Circolare (*Closed-Loop*):** Una struttura in cui i neuroni sono connessi in un anello chiuso. In questa configurazione, il segnale tende a ricircolare all'interno del sistema, facilitando la creazione di dinamiche cicliche o retroazioni stabili.
2. **Topologia ad Albero:** Una struttura gerarchica in cui l'informazione si propaga da nodi "padre" (centrali) verso nodi "figli" (periferici), ramificandosi progressivamente. Questa configurazione consente di studiare la trasmissione unidirezionale del segnale e i ritardi di cascata lungo i diversi livelli della gerarchia.

3.2.1 Topologia circolare

Il primo scenario analizzato prevede una configurazione ad anello chiuso composta da 5 nodi (indicati nella simulazione con indici da 0 a 4).

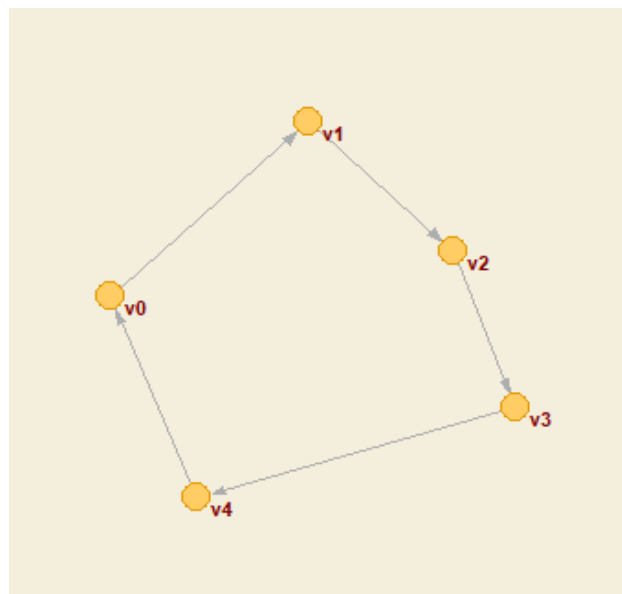


Figura 3.1 Topologia circolare

Per comprendere il funzionamento del simulatore, l'analisi si è concentrata sulla struttura generata dallo script `gen_rete0.py` e mappata nel file di configurazione `input_rete.txt`.

5			
0	1	0	0
1	4		
1	-0.31460097		
1	1		
1			
1	1	0	0
1	0		
1	-0.89110585		
1	1		
2	0	0	0
1	1		
1	-0.31308524		
1	1		
3	1	0	0
1	2		
1	0.82736186		
1	1		
4	0	0	0
1	3		
1	-0.17915401		
1	1		

Figura 3.2 Rete topologia circolare.

La prima riga indica il numero di nodi del sistema. Ogni gruppetto di righe successive elenca i dati necessari alla simulazione di un nodo del sistema. In particolare, in ogni gruppetto: ID del nodo; Stato (0 oppure 1); Soglia; Probabilità di deviazione spontanea dal comportamento definito dalla regola di attivazione

La seconda riga: Numero nodi in ingresso; Elenco degli identificatori dei nodi in ingresso.

La terza riga: Ancora il numero di nodi in ingresso. Per ogni nodo in ingresso, segue il peso del link associato.

La quarta riga: Ancora il numero di nodi in ingresso. Per ogni nodo in ingresso, segue il peso il ritardo di trasmissione del link associato.

Il file di input definisce una rete a **5 nodi** (da 0 a 4). Analizzando i blocchi di connessione, emerge la struttura ad anello (*closed-loop*), dove ogni nodo riceve un input da un singolo nodo precedente, ma con la presenza di **connessioni inibitorie** (pesi negativi) e **connessioni eccitatorie** (pesi positivi) come mostrato in *figura 5*.

Sono state fatte delle simulazioni, si veda Appendice B.

Da queste simulazioni si evince che il sistema ha notevoli difficoltà ad identificare come unica coppia due nodi con attivazioni separate da tempi di ritardo.

3.2.2 Topologia ad albero

In questa seconda parte della sperimentazione, abbiamo abbandonato l'architettura ad anello per studiare una **topologia ad albero composta da 15 nodi** (da N000 a N014).

Guardando la struttura dei collegamenti che abbiamo generato, si nota subito che non siamo più davanti a un percorso lineare. Ciascun nodo fa da "padre" a uno o più nodi "figli". Ad esempio:

- Il **Nodo 0** è la radice e si collega ai nodi 1, 2.
- Il **Nodo 3** proietta sui nodi 7 e 8.
- Il **Nodo 4** proietta sui nodi 9 e 10.
- I nodi successivi continuano a diramarsi (es. il Nodo 5 si collega a 11 e 12; il Nodo 6 a 13 e 14).

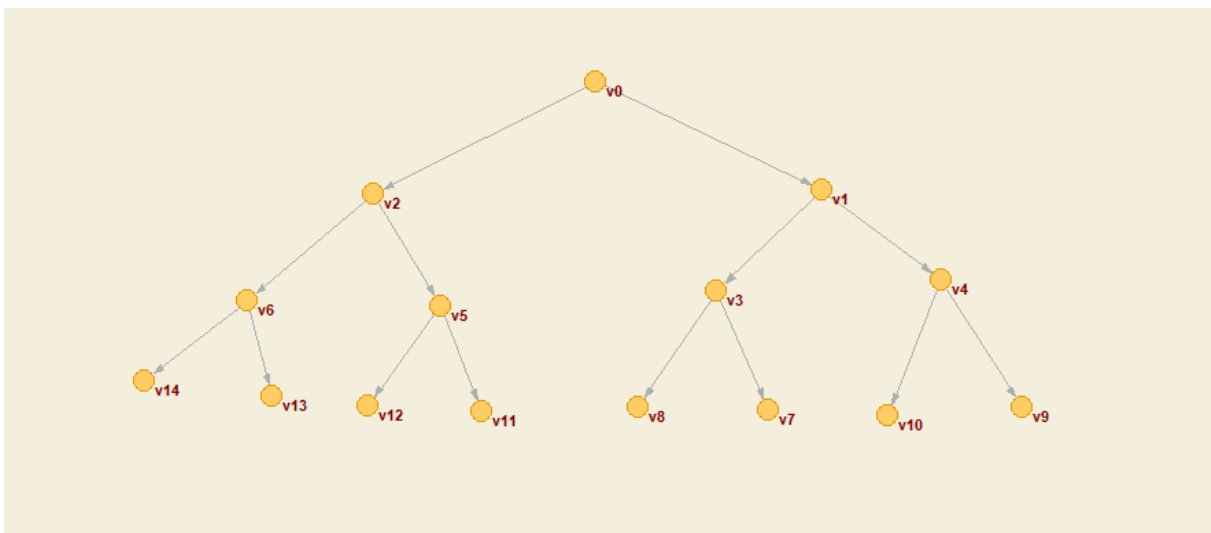


Figura 3.3 Topologia ad albero

Osservando la traiettoria dei dati ottenuta, notiamo un fenomeno molto netto. Al passo iniziale ($t = 0$), la rete si accende in un punto ben preciso: risultano attivi contemporaneamente i nodi **N002, N008, N009, N010 e N012**.

0	0	0	0	5	0	0	0	10	1	0	0
0				1	2			1	4		
0				1	-0.933			1	-0.5199		
0				1	1			1	1		
1	0	0	0	6	0	0	0	11	0	0	0
1	0			1	2			1	5		
1	0.2704			1	-0.9434			1	0.431		
1	1			1	1			1	1		
2	1	0	0	7	0	0	0	12	1	0	0
1	0			1	3			1	5		
1	0.0399			1	0.2244			1	-0.9432		
1	1			1	1			1	1		
3	0	0	0	8	1	0	0	13	0	0	0
1	1			1	3			1	6		
1	0.9464			1	-0.4346			1	-0.5719		
1	1			1	1			1	1		
4	0	0	0	9	1	0	0	14	0	0	0
1	1			1	4			1	6		
1	-0.9551			1	-0.2743			1	-0.0325		
1	1			1	1			1	1		

Figura 3.4 Rete topologia ad albero. Rete visualizzata su tre colonne contigue per ragioni di spazio. L'ordine di lettura si sviluppa verticalmente in ciascuna colonna, muovendosi sequenzialmente da sinistra a destra.

Sono state fatte delle simulazioni, si veda Appendice B.

Anche da queste simulazioni si evince che ha notevoli difficoltà ad identificare come unica coppia due nodi con attivazioni separate da tempi di ritardo. Al tempo stesso, possono essere identificati come un unico gruppo nodi che hanno in realtà in comune nodi a monte "a pari distanza".

Per evitare di moltiplicare il numero di analisi da fare - facendo scorrere le colonne le une rispetto alle altre di un numero di passi pari al sospetto tempo di ritardo, in un pesante crescendo combinatorio - sembra possa invece essere utile passare ad un'analisi basata sull'integrazione dei segnali all'interno di una finestra temporale. In tal modo, se la dimensione della finestra è maggiore dei tempi di ritardo tipici, i tempi di ritardo stessi sono mediati via, ed è possibile utilizzare il framework da noi proposto.

3.3 Analisi dei dati ed osservazioni sperimentali

L'analisi dei dati generati dalla simulazione ha permesso di evidenziare dinamiche complesse nella rete, identificabili solo attraverso una corretta metodologia di elaborazione del segnale. Le osservazioni principali possono essere riassunte in tre aspetti metodologici e fenomenologici fondamentali.

I dati grezzi generati dal modello sono essenzialmente stringhe binarie di 0 e 1 (neurone spento o acceso a ogni singolo istante). Per estrarre informazioni significative, l'approccio classico basato sulla sincronizzazione millimetrica (simultaneità assoluta) si rivela spesso insufficiente o troppo rigido, specialmente in presenza di ritardi di conduzione.

Per ovviare a questo problema, si è scelto di applicare una **finestra di integrazione**. Lo scopo della finestra è contare quante volte, all'interno di quel determinato intervallo di tempo, una coppia di nodi risulta attiva. Questo espediente matematico permette di **trasformare i dati biologici discreti (acceso/spento) in una misura continua di frequenza di sparo** (frequenza di scarica nella finestra).

Questo approccio alleggerisce notevolmente il problema dei ritardi intrinseci del segnale. Mentre i metodi attuali tentano di effettuare complessi *shift* (slittamenti) temporali dei dati per riallineare i segnali, l'integrazione in finestra assume che la sincronizzazione esatta al millisecondo potrebbe non essere il fattore cruciale. Al contrario, ciò che rileva è che due nodi abbiano "sparato molto o poco insieme" all'interno di un intervallo temporale biologicamente coerente (ad esempio, nell'arco di un secondo).

Per studiare l'evoluzione finestre ampie 10 passi, facendole scorrere 1 passo alla volta (come poi è indicato in seguito) e finestre ampie 64 passi, facendole scorrere 64 passi alla volta. Questo campionamento a blocchi permette di analizzare finestre temporali adiacenti ma distinte dello stesso fenomeno.

Questo ultimo approccio, che divide la sequenza di osservazioni in diverse parti, ha permesso di osservare che molte delle associazioni e delle sincronizzazioni funzionali tra neuroni chiaramente visibili all'interno della prima finestra temporale tendono a svanire o a modificarsi nelle due finestre successive. Questo comportamento dimostra l'instabilità o la natura transitoria di determinati *cell assemblies*, evidenziando come la rete sperimenti stati di attivazione differenti in punti temporali diversi.

Un'importante evidenza emersa durante le simulazioni riguarda la necessità del **rumore di fondo** nel sistema. In assenza di perturbazioni (caso deterministico ideale), i sistemi tendono a cadere rapidamente all'interno dei loro **attrattori**, esaurendo la propria dinamica o rimanendo bloccati in uno stato fisso di totale immobilità. Il rumore è quindi biologicamente e computazionalmente fondamentale per "scuotere" il sistema, mantenerlo attivo e permettere l'esplorazione di nuovi stati computazionali.

Nella nostra sperimentazione, il rumore è stato concentrato strategicamente su nodi specifici. Analizzando i risultati di questa configurazione, si è osservato un fenomeno controintuitivo legato alla topologia ad albero: spesso, gli impulsi e le attivazioni **si manifestano prima nei nodi figli** rispetto ai nodi padre. Dal punto di vista del segnale, sui nodi principali (i padri) si riescono a registrare e visualizzare solo le "code" di attivazione provenienti dal nodo in cui è concentrato il rumore, alterando la percezione della classica gerarchia di trasmissione dall'alto verso il basso.

3.3.1 Finestra applicata alla Topologia circolare

1. Applichiamo la finestra 10 al nodo 2 perturbato con probabilità 0.2

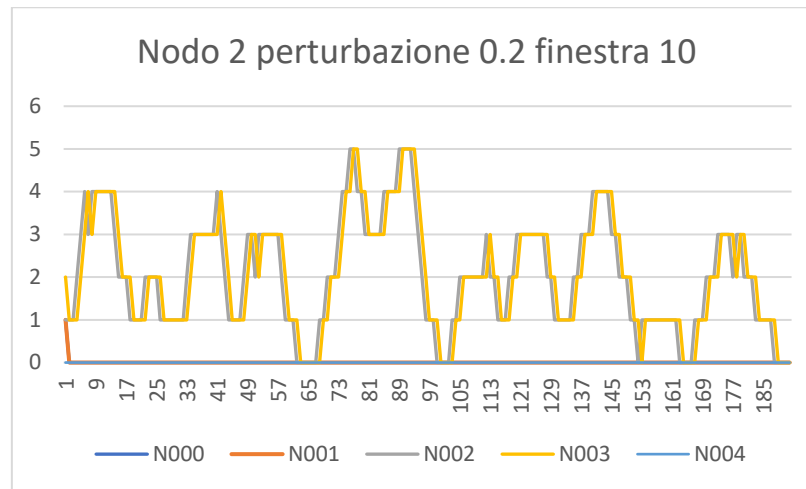


Figura 3.5

I nodi attivi sono il Nodo 2 e il Nodo 3. Il Nodo 0 e il Nodo 1 sono attivi al primo passo per poi spegnersi e non accendersi più, mentre il Nodo 4 rimane sempre spento.

Per quanto riguarda lo Zi:

8.13863878	3	_N000_N001_	N003	N004
8.13863878	2	_N000_N001__N003_	N004	

Figura 3.6

Ora andiamo a studiare la nostra traiettoria dividendola in tre tracciati.

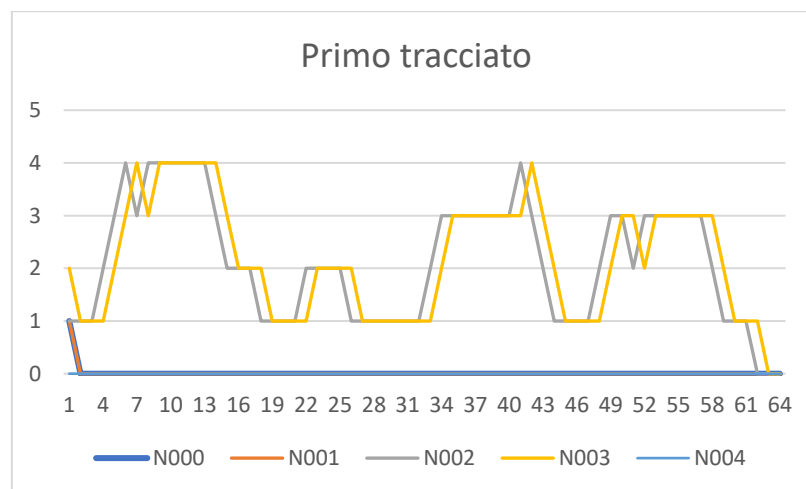


Figura 3.7

Anche nel primo tracciato I nodi attivi sono il Nodo 2 e il Nodo 3. Il Nodo 0 e il Nodo 1 sono attivi al primo passo per poi spegnersi e non accendersi più, mentre il Nodo 4 rimane sempre spento.

Per quanto riguarda lo Zi:

6.577549	3	_N000_N001_	N003	N004
6.577549	2	_N000_N001_N003_	N004	

Figura 3.8

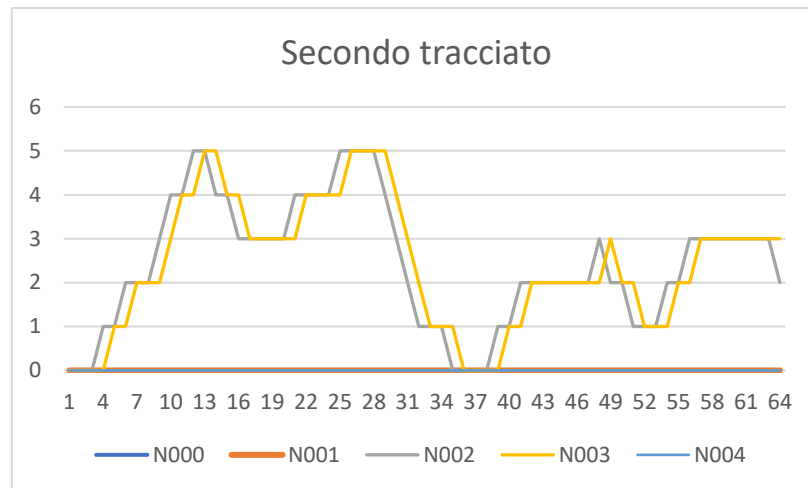


Figura 3.9

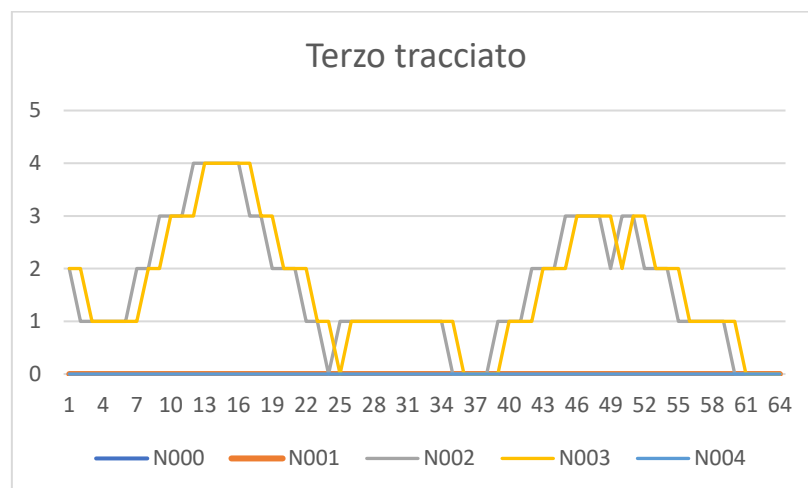


Figura 3.10

Nel secondo e terzo tracciato, gli unici nodi attivi sono il Nodo 2 e il Nodo 3, mentre gli altri non ci accendono mai, quindi il massimo che potrebbe fare sono gruppi da due. Per questo motivo non si riescono a ottenere valori per lo Zi.

2. Applichiamo la finestra 10 al Nodo 4 perturbato con probabilità 0.2:

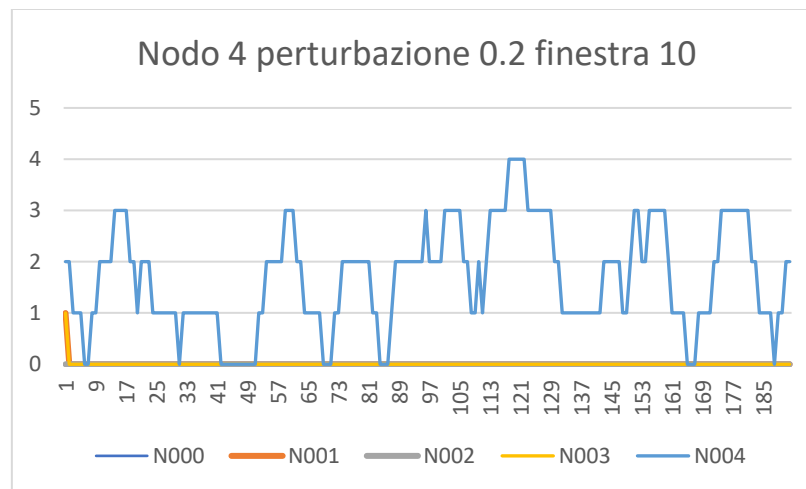


Figura 3.11

L'unico nodo attivo è il Nodo 4. I Nodi 0, 1 e 3 sono attivi al primo passo per poi spegnersi e non accendersi più, mentre il Nodo 2 non si accende mai.

Per quanto riguarda lo Zi:

8.13863878	3	_N000_N001_	N003	N004
8.13863878	2	__N000_N001__N003_	N004	

Figura 3.12

Andiamo a studiare i tre tracciati

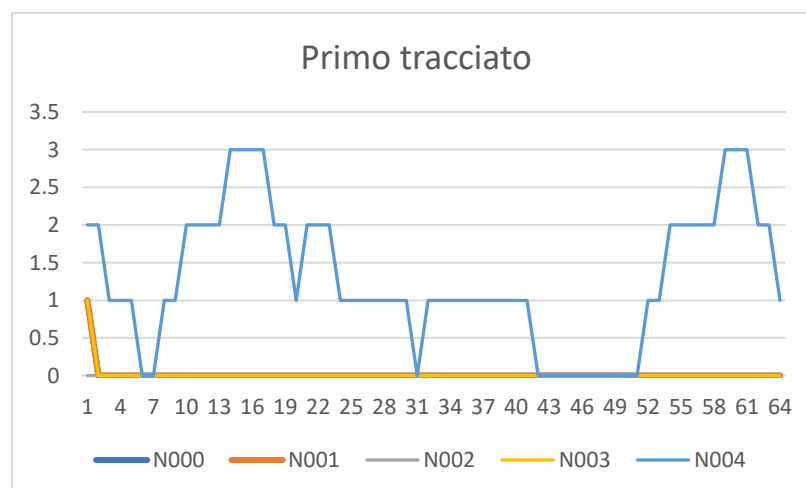


Figura 3.13

L'unico nodo attivo è il Nodo 4. I Nodi 0, 1 e 3 sono attivi al primo passo per poi spegnersi e non accendersi più, mentre il Nodo 2 non si accende mai.

Per quanto riguarda lo Zi:

6.577549	3	_N000_N001_	N003	N004
6.577549	2	_N000_N001_N003_	N004	

Figura 3.14

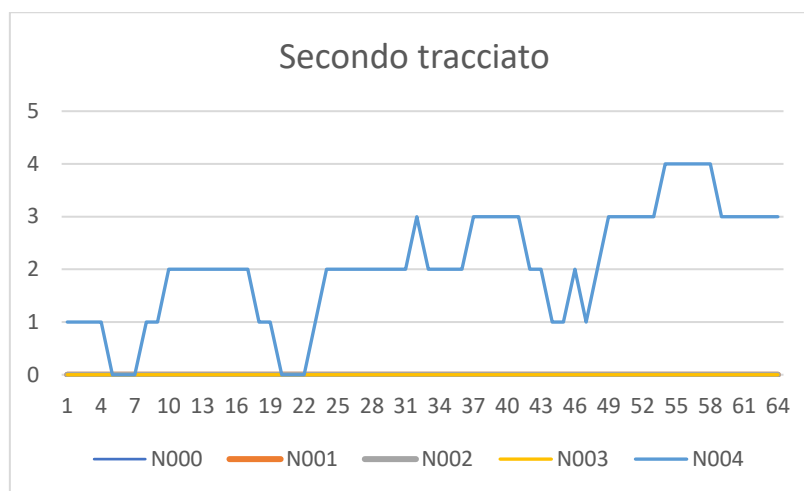


Figura 3.15

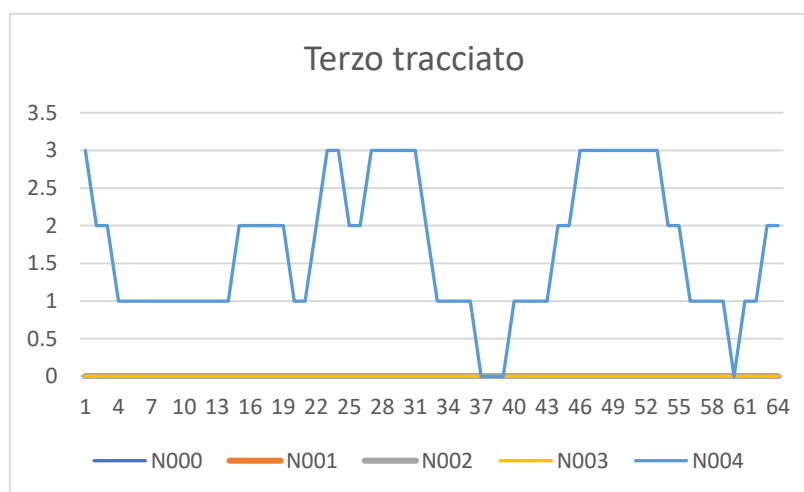


Figura 3.16

Nel secondo e nel terzo tracciato l'unico nodo attivo è il Nodo 4, mentre il resto dei nodi non si accende mai. Per questo motivo non si riescono a ottenere risultati per lo Zi.

3. Applichiamo la finestra 10 al Nodo 2 e al Nodo 4 perturbati contemporaneamente con probabilità 0.2

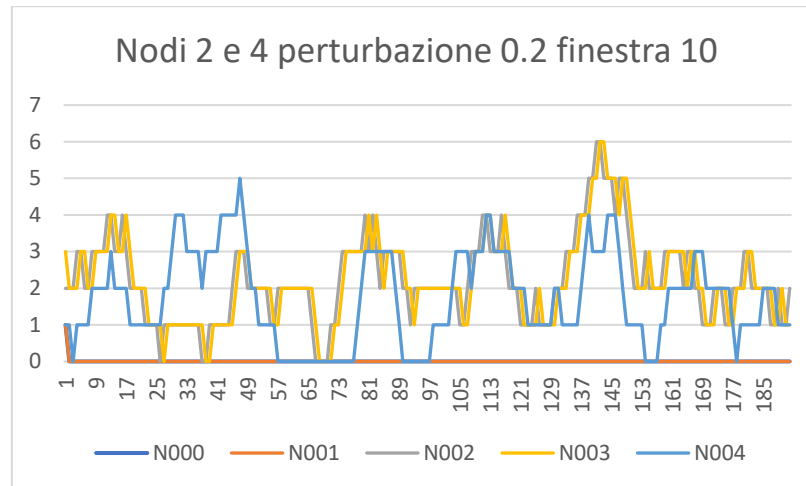


Figura 3.17

I nodi attivi sono il Nodo 2, 3 e 4. Il Nodo 0 e il Nodo 1 sono attivi al primo passo per poi spegnersi e non accendersi più.

Per quanto riguarda lo Zi:

25.9023921	4	_N002_N003_	N000	N001	N004
8.13863878	3	_N000_N001_	_N002_N003_	N004	
1.75574126	2	_N002_N003_N004_	_N000_N001_		

Figura 3.18

Andiamo a studiare i tre tracciati

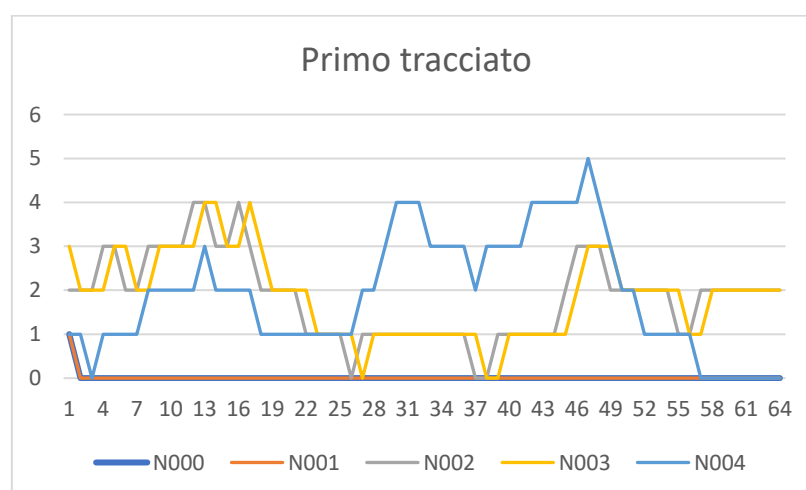


Figura 3.19

I nodi attivi sono il Nodo 2, 3 e 4. Il Nodo 0 e il Nodo 1 sono attivi al primo passo per poi spegnersi e non accendersi più.

Per quanto riguarda lo Zi:

11.405617	4	_N002_N003_	N000	N001	N004
6.5775491	3	_N000_N001_	_N002_N003_	N004	
2.9432795	2	_N002_N003_N004_	_N000_N001_		

Figura 3.20

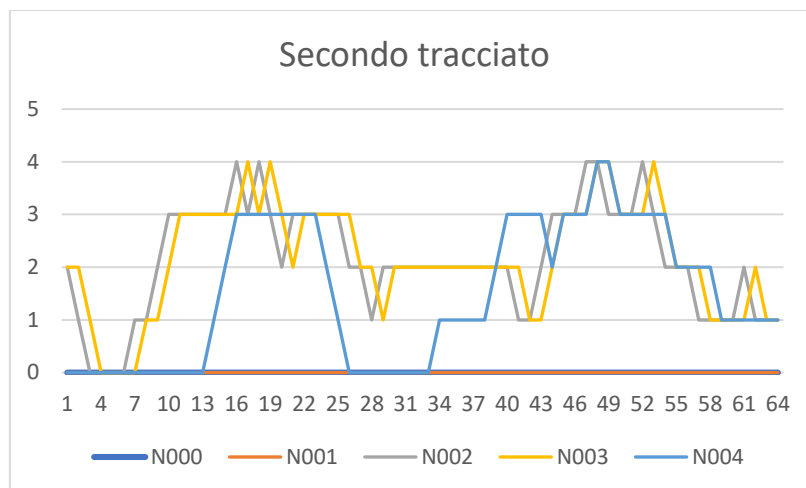


Figura 3.21

I nodi attivi sono il Nodo 2, 3 e 4, mentre il Nodo 0 e il Nodo 1 non si accendono mai.

Per quanto riguarda lo Zi:

10.6167375	2	_N002_N003_	N004
------------	---	-------------	------

Figura 3.22

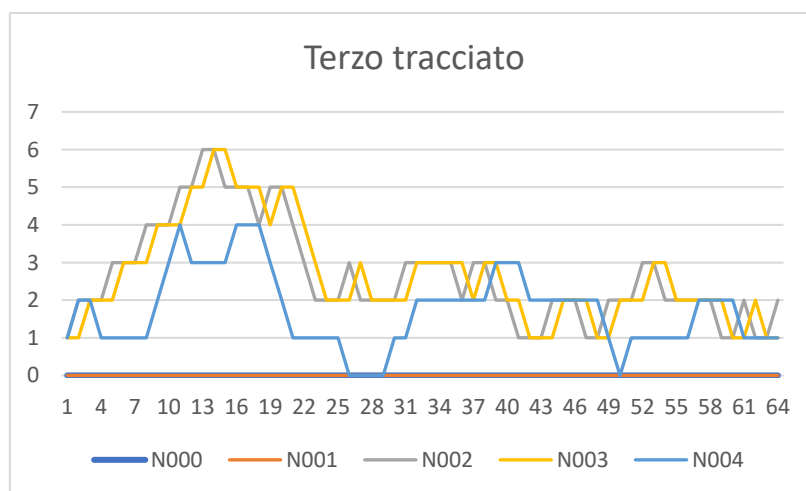


Figura 3.23

I nodi attivi sono il Nodo 2, 3 e 4, mentre il Nodo 0 e il Nodo 1 non si accendono mai.

Per quanto riguarda lo Zi:

8.5470579	2	_N002_N003_	N004
-----------	---	-------------	------

Figura 3.24

3.3.1.1 Osservazioni

Da queste simulazioni si evince che:

1. la topologia statica è solo un supporto alla dinamica, che dipende anche dalle regole di attivazione dei singoli nodi
2. collegamenti con pesi positivi (attivatori) permettono al rumore id propagarsi (è sufficiente che il segnale del nodo in ingresso permetta il superamento della soglia del neurone target)
3. collegamenti con pesi negativi (inibitori) raramente permettono al rumore di propagarsi, perché l'azione di inibizione è realmente efficace solo se il nodo target è in procinto di emettere il proprio segnale

3.3.2 Finestra applicata alla Topologia ad albero

Ora l'algoritmo non sta più guardando se un neurone è acceso o spento in un singolo istante, ma sta sommando quante volte quel neurone ha sparato all'interno di una finestra mobile di 10 passi.

1. Applichiamo la finestra 10 al Nodo 0 perturbato con probabilità 0.1:

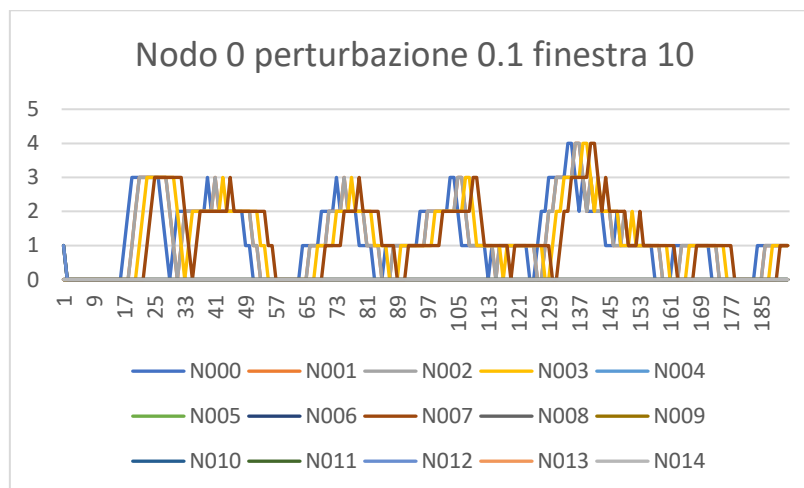


Figura 3.25

I nodi attivi sono il Nodo 0, il Nodo 1, il Nodo 2, il Nodo 3 e il Nodo 7.

I nodi 8, 9, 10 e 12 sono attivi al primo passo per poi spegnersi e non accendersi più.

Per quanto riguarda lo Zi:

86.5984397	8	_N001_N002_	N000	N003	N007	N008	N009	N010	N012
31.1609838	7	_N003_N007_	_N001_N002_	N000	N008	N009	N010	N012	
26.9695606	6	_N001_N002_N000_	_N003_N007_	N008	N009	N010	N012		
8.13863878	5	_N008_N010_	_N001_N002_N000_	_N003_N007_	N009	N012			
8.13863878	4	_N008_N010_N012_	_N001_N002_N000_	_N003_N007_	N009				
8.13863878	3	_N008_N010_N012_N009_	_N001_N002_N000_	_N003_N007_					
2.03624689	2	_N001_N002_N000_N003_N007_	_N008_N010_N012_N009_						

Figura 3.26

Nei test precedenti ad alta frequenza, avevamo notato che lo Z-score della via dinamica tendeva a crollare vicino allo zero (saturazione da rumore). Con l'introduzione della finestra a 10, lo Z-score della catena principale schizza a **86.5984**. L'integrazione nel tempo permette all'algoritmo di ignorare i micro-ritardi sinaptici casuali e di focalizzarsi sull'evento globale: i nodi stanno sparando insieme all'interno dello stesso intervallo temporale significativo.

Ora andiamo a studiare la nostra traiettoria dividendola in tre tracciati (diversi periodi di osservazione dello stesso fenomeno)

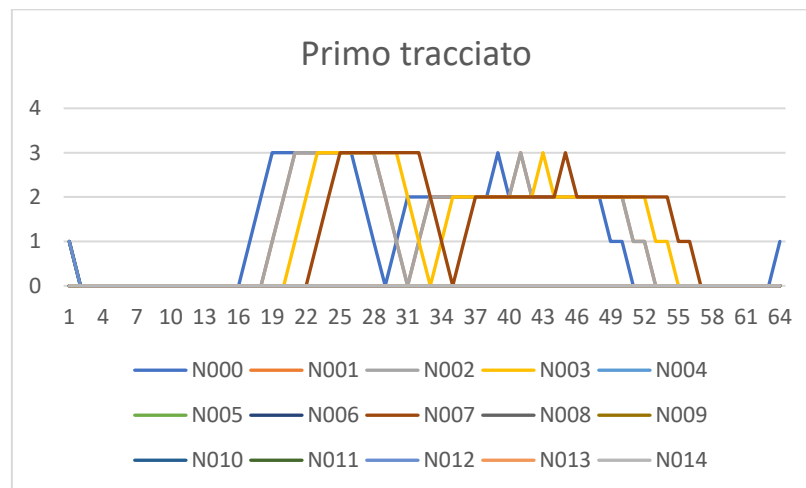


Figura 3.27

Anche nel primo tracciato i nodi attivi sono il Nodo 0, il Nodo 1, il Nodo 2, il Nodo 3 e il Nodo 7. I Nodi 8, 9, 10 e 12 sono attivi al primo passo per poi spegnersi e non accendersi più.

Per quanto riguarda lo Zi:

32.406702	8	_N001_N002_	N000	N003	N007	N008	N009	N010	N012
13.745574	7	_N003_N007_	_N001_N002_	N000	N008	N009	N010	N012	
10.853624	6	_N001_N002_N000_	_N003_N007_	N008	N009	N010	N012		
6.5775491	5	_N008_N010_	_N001_N002_N000_	_N003_N007_	N009	N012			
6.5775491	4	_N008_N010_N012_	_N001_N002_N000_	_N003_N007_	N009				
6.5775491	3	_N008_N010_N012_N009_	_N001_N002_N000_	_N003_N007_					
-0.6988451	2	_N008_N010_N012_N009_N001_N002_N000_	_N003_N007_						

Figura 3.28

Non vi sono grandi differenze tra questo primo tracciato e la traiettoria completa.

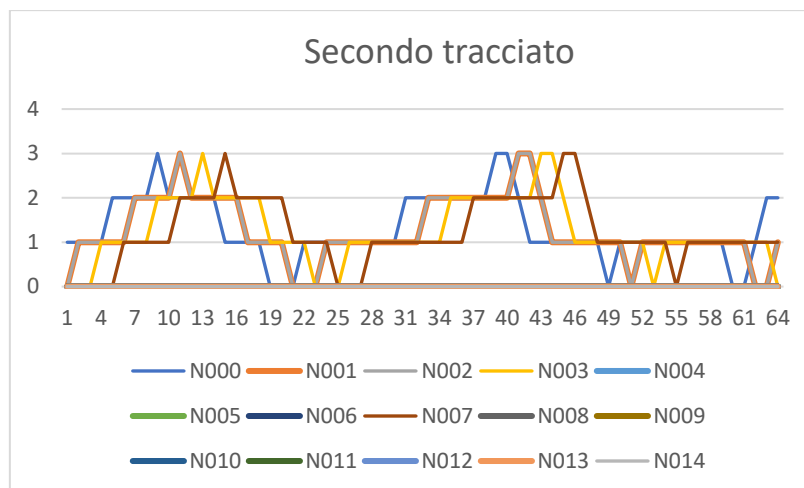


Figura 3.29

Nel secondo tracciato i nodi attivi sono il Nodo 0, il Nodo 1, il Nodo 2, il Nodo 3 e il Nodo 7, mentre il resto dei nodi resta sempre spento.

Per quanto riguarda lo Zi:

30.0369231	4	_N001_N002_	N000	N003	N007
8.23522323	3	_N003_N007_	_N001_N002_	N000	
5.6420716	2	_N001_N002_N000_	_N003_N007_		

Figura 3.30

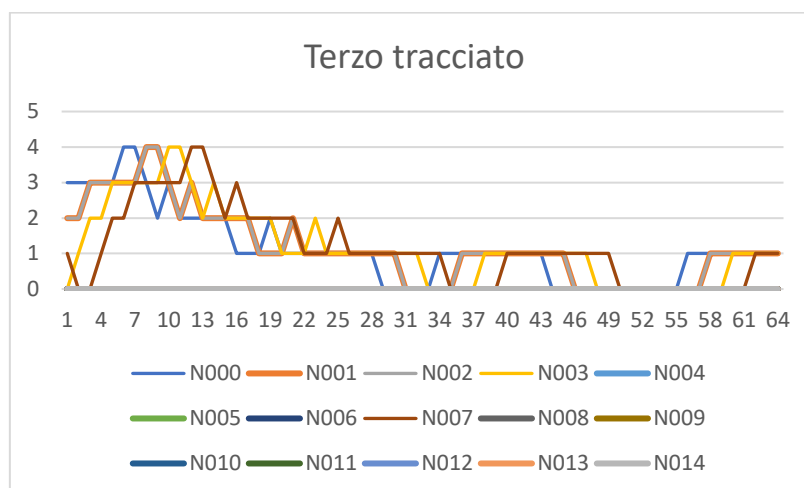


Figura 3.31

Nel secondo tracciato i nodi attivi sono il Nodo 0, il Nodo 1, il Nodo 2, il Nodo 3 e il Nodo 7, mentre il resto dei nodi resta sempre spento.

Per quanto riguarda lo Zi:

27.4250039	4	_N001_N002_	N000	N003	N007
11.380575	3	_N001_N002_N000_	N003	N007	
9.2776763	2	_N003_N007_	_N001_N002_N000_		

Figura 3.32

2. Applichiamo la finestra 10 al Nodo 0 perturbato con probabilità 0.2:

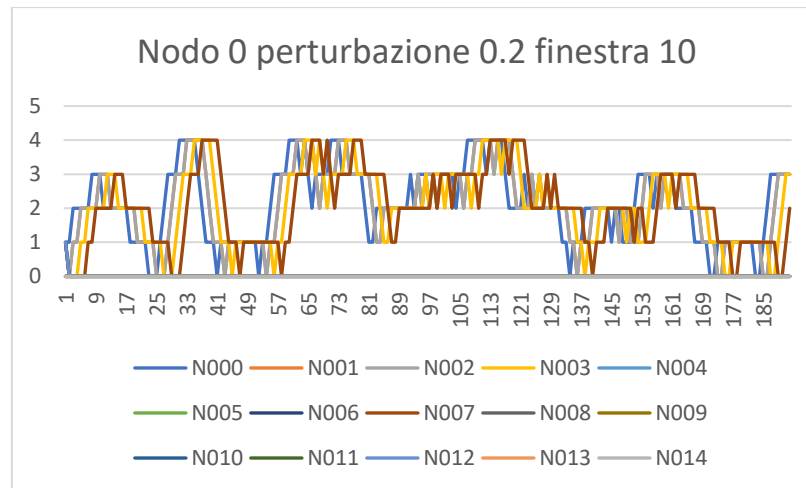


Figura 3.33

I nodi attivi sono il Nodo 0, il Nodo 1, il Nodo 2, il Nodo 3 e il Nodo 7. I Nodi 8, 9, 10 e 12 sono attivi al primo passo per poi spegnersi e non accendersi più.

Per quanto riguarda lo Zi:

95.8531828	8	_N001_N002_	N000	N003	N007	N008	N009	N010	N012
27.7119294	7	_N003_N007_	_N001_N002_	N000	N008	N009	N010	N012	
23.7534526	6	_N001_N002_N000_	_N003_N007_	N008	N009	N010	N012		
8.13863878	5	_N008_N010_	_N001_N002_N000_	_N003_N007_	N009	N012			
8.13863878	4	_N008_N010_N012_	_N001_N002_N000_	_N003_N007_	N009				
8.13863878	3	_N008_N010_N012_N009_	_N001_N002_N000_	_N003_N007_					
2.51743099	2	_N001_N002_N000_N003_N007_	_N008_N010_N012_N009_						

Figura 3.34

Andiamo a studiare i tre tracciati

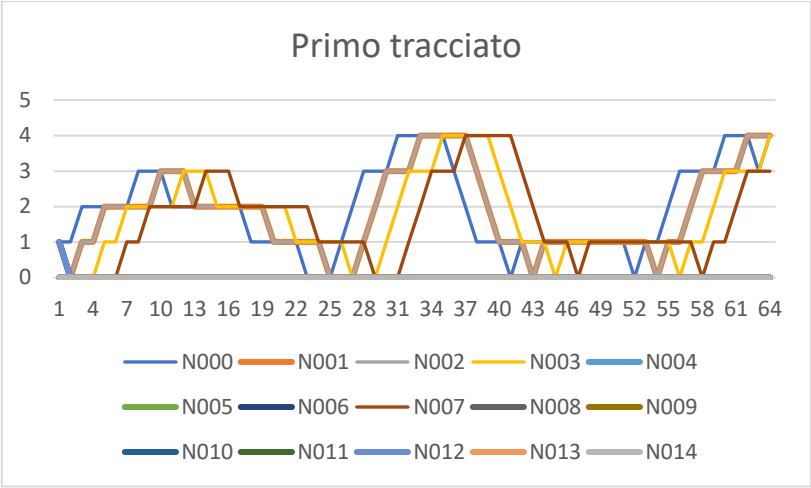


Figura 3.35

Anche nel primo tracciato i nodi attivi sono il Nodo 0, il Nodo 1, il Nodo 2, il Nodo 3 e il Nodo 7. I Nodi 8, 9, 10 e 12 sono attivi al primo passo per poi spegnersi e non accendersi più

Per quanto riguarda lo Zi:

29.9250139	8	_N001_N002_	N000	N003	N007	N008	N009	N010	N012
7.286786807	7	_N003_N007_	_N001_N002_	N000	N008	N009	N010	N012	
6.5775491	6	_N008_N009_	_N003_N007_	_N001_N002_	N000	N010	N012		
6.5775491	5	_N008_N009_N010_	_N003_N007_	_N001_N002_	N000	N012			
6.5775491	4	_N008_N009_N010_N012_	_N003_N007_	_N001_N002_	N000				
6.133622754	3	_N001_N002_N000_	_N008_N009_N010_N012_	_N003_N007_					
-1.14868738	2	_N001_N002_N000_N008_N009_N010_N012_	_N003_N007_						

Figura 3.36

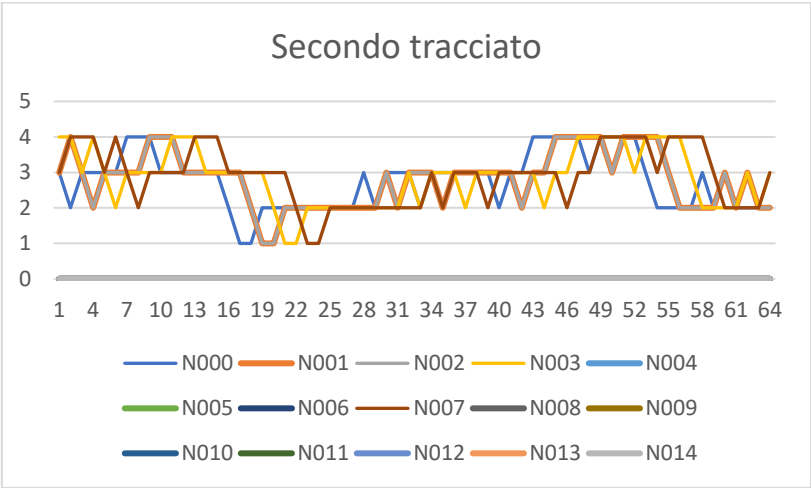


Figura 3.37

Nel secondo tracciato i nodi attivi sono il Nodo 0, il Nodo 1, il Nodo 2, il Nodo 3 e il Nodo 7, mentre il resto dei nodi resta sempre spento.

Per quanto riguarda lo Zi:

32.8574532	4	_N001_N002_	N000	N003	N007
5.43144588	3	_N003_N007_	_N001_N002_	N000	
5.32100208	2	_N001_N002_N000_	_N003_N007_		

Figura 3.38

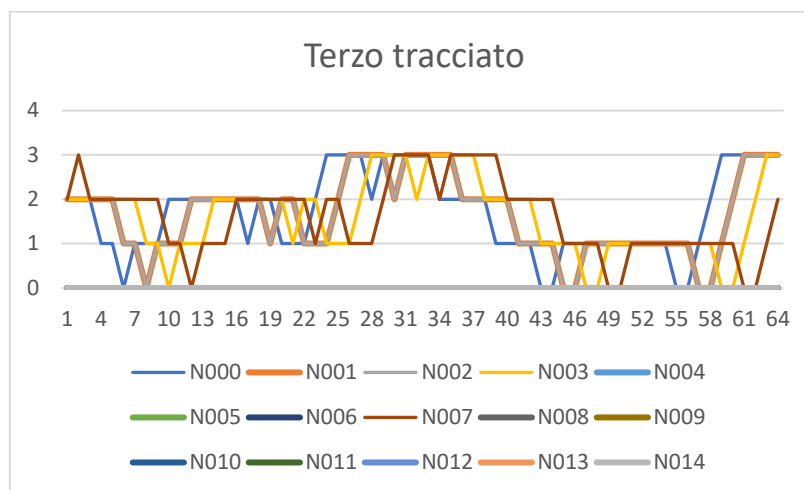


Figura 3.39

Nel terzo tracciato i nodi attivi sono il Nodo 0, il Nodo 1, il Nodo 2, il Nodo 3 e il Nodo 7, mentre il resto dei nodi resta sempre spento.

Per quanto riguarda lo Zi:

35.827102	4	_N001_N002_	N000	N003	N007
8.51867388	3	_N001_N002_N000_	N003	N007	
6.87380553	2	_N003_N007_	_N001_N002_N000_		

Figura 3.40

3. Applichiamo la finestra 10 al nodo 3 perturbato con probabilità 0.1:

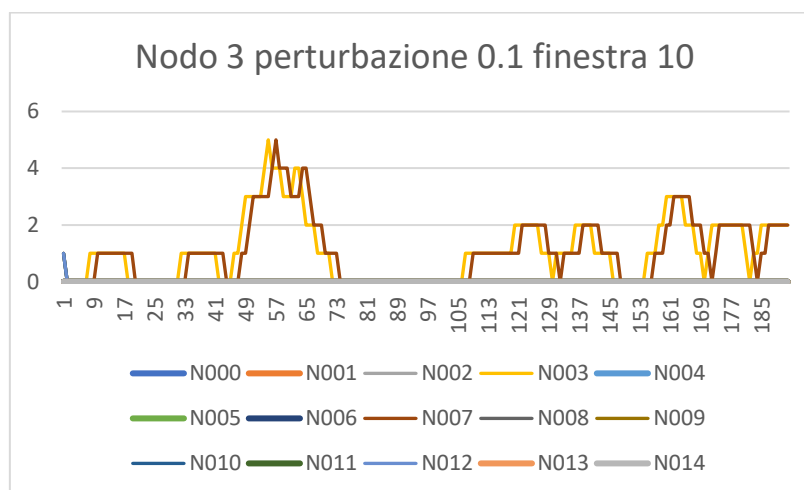


Figura 3.41

Gli unici nodi attivi sono il Nodo 3 e il Nodo 7. I Nodi 2, 8, 9, 10 e 12 sono attivi al primo passo per poi spegnersi e non accendersi più.

Per quanto riguarda lo Zi:

28.4634562	6	_N003_N007_	N002	N008	N009	N010	N012
8.13863878	5	_N008_N010_	_N003_N007_	N002	N009	N012	
8.13863878	4	_N002_N012_	_N008_N010_	_N003_N007_	N009		
8.13863878	3	_N002_N012_ N008_N010_	_N003_N007_	N009			
8.13863878	2	_N002_N012_ N008_N010_ N009_	_N003_N007_				

Figura 3.42

Andiamo a studiare i 3 tracciati

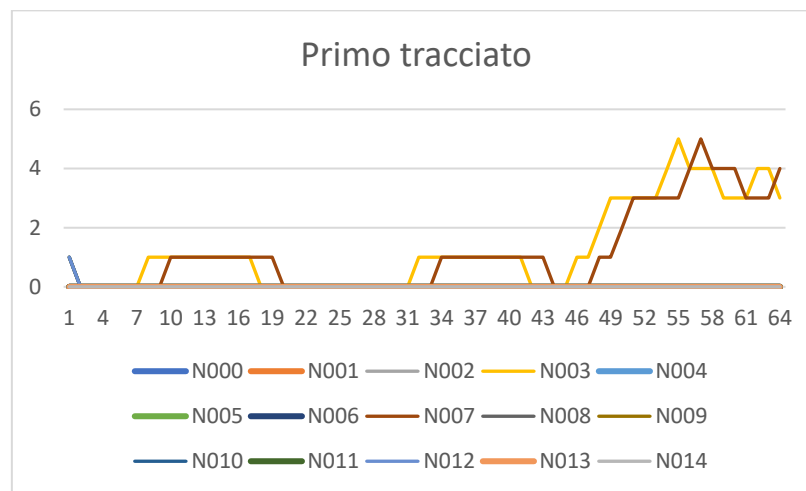


Figura 3.43

Anche nel primo tracciato gli unici nodi attivi sono il Nodo 3 e il Nodo 7. I Nodi 2, 8, 9, 10 e 12 sono attivi al primo passo per poi spegnersi e non accendersi più.

Per quanto riguarda lo Zi:

8.79927328	6	_N003_N007_	N002	N008	N009	N010	N012
6.5775491	5	_N008_N010_	_N003_N007_	N002	N009	N012	
6.5775491	4	_N002_N012_	_N008_N010_	_N003_N007_	N009		
6.5775491	3	_N002_N012_ N008_N010_	_N003_N007_	N009			
6.5775491	2	_N002_N012_ N008_N010_ N009_	_N003_N007_				

Figura 3.44

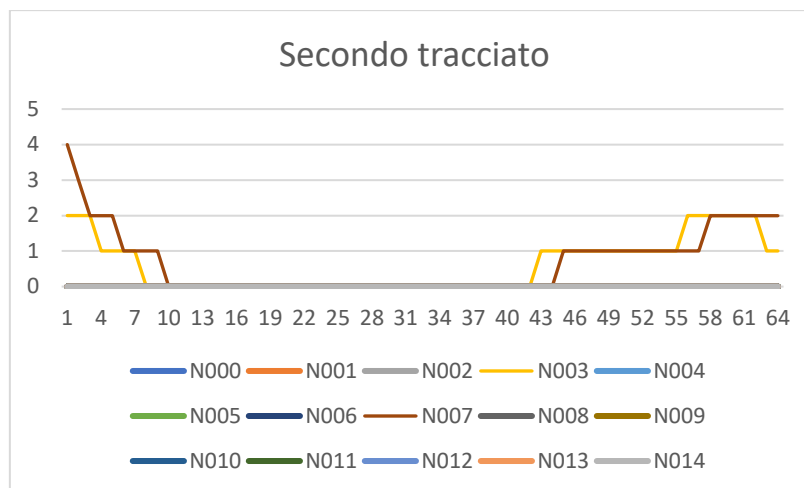


Figura 3.45

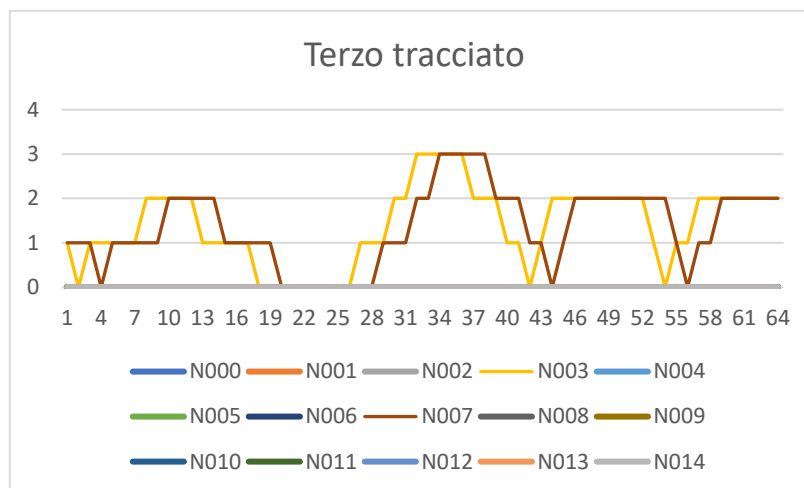


Figura 3.46

Nel secondo e terzo tracciato, gli unici nodi attivi sono il Nodo 3 e il Nodo 7, mentre gli altri non ci accendono mai, quindi il massimo che potrebbe fare sono gruppi da due. Per questo motivo non si riescono a ottenere valori per lo Zi (per ottenere risultati sullo Zi servono almeno 3 variabili).

4. Applichiamo la finestra 10 al nodo 3 perturbato con probabilità 0.2

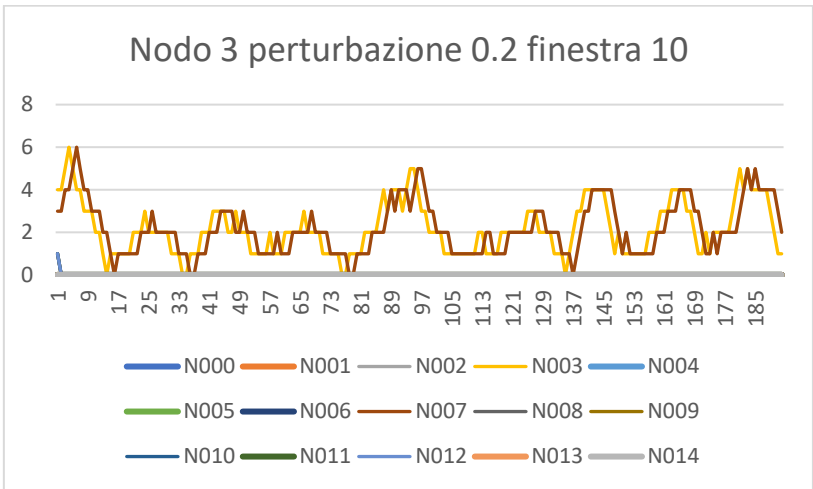


Figura 3.47

Gli unici nodi attivi sono il Nodo 3 e il Nodo 7.
 I Nodi 2, 8, 9, 10 e 12 sono attivi al primo passo per poi spegnersi e non accendersi più.

Per quanto riguarda lo Zi:

14.2600279	6	_N003_N007_	N002	N008	N009	N010	N012
8.13863878	5	_N008_N010_	_N003_N007_	N002	N009	N012	
8.13863878	4	_N002_N012_	_N008_N010_	_N003_N007_	N009		
8.13863878	3	_N002_N012_ _N008_N010_	_N003_N007_	N009			
8.13863878	2	_N002_N012_ _N008_N010_ _N009_	_N003_N007_				

Figura 3.48

Andiamo a studiare i tre tracciati

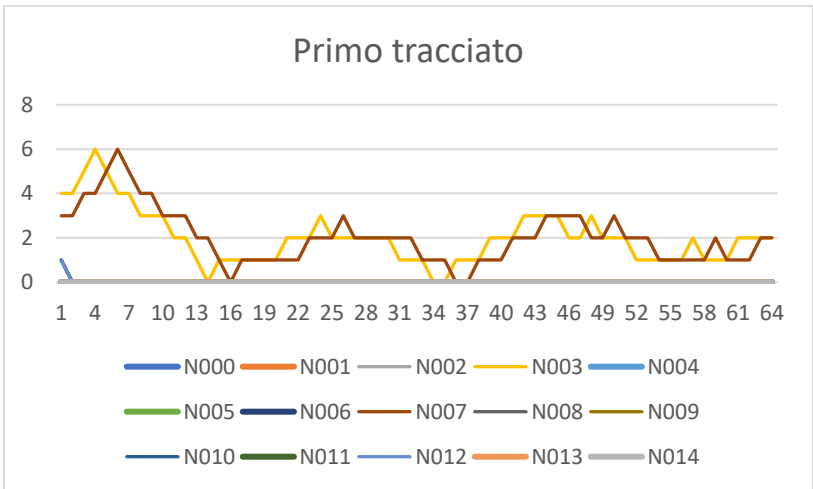


Figura 3.49

Gli unici nodi attivi sono il Nodo 3 e il Nodo 7.

I Nodi 2, 8, 9, 10 e 12 sono attivi al primo passo per poi spegnersi e non accendersi più.

Per quanto riguarda lo Zi:

6.577549	6	_N008_N009_	N002	N003	N007	N010	N012
6.577549	5	_N008_N009_N002_	N003	N007	N010	N012	
6.577549	4	_N008_N009_N002_N012_	N003	N007	N010		
6.577549	3	_N008_N009_N002_N012_N010_	N003	N007			
4.013965	2	_N003_N007_	_N008_N009_N002_N012_N010_				

Figura 3.50

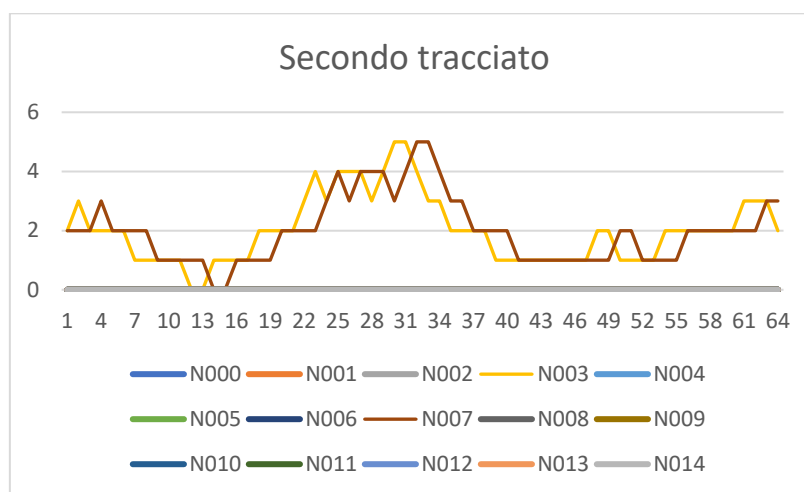


Figura 3.51

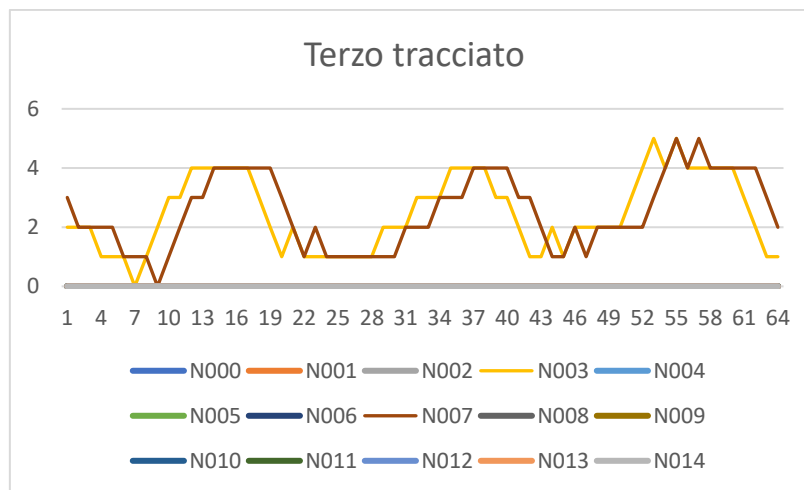


Figura 8

Gli unici nodi attivi sono il Nodo 3 e il Nodo 7, quindi anche in questo caso non si riescono a ottenere risultati sullo Zi.

3.3.2.1 Osservazioni

Da queste simulazioni si evince che:

1. la topologia statica è solo un supporto alla dinamica, che dipende anche dalle regole di attivazione dei singoli nodi
2. collegamenti con pesi positivi (attivatori) permettono al rumore di propagarsi (è sufficiente che il segnale del nodo in ingresso permetta il superamento della soglia del neurone target)
3. collegamenti con pesi negativi (inibitori) raramente permettono al rumore di propagarsi, perché l'azione di inibizione è realmente efficace solo se il nodo target è in procinto di emettere il proprio segnale
4. se la situazione cambia nel tempo è opportuno prendere delle finestre di osservazione più piccole, che non tutta la sequenza.

4. Strutture che evolvono nel tempo

4.1 Validazione del framework Zi tramite tracciati dinamici con ground truth

Nel Capitolo 3 è stato esplorato un modello di rete di tipo strutturale, in cui il comportamento del sistema è definito a livello microscopico attraverso un grafo di connessioni, topologie specifiche (ad albero e circolare) e leggi di somma pesata degli input. Sebbene tale approccio sia fondamentale per simulare l'attività biologica, esso presenta un limite intrinseco dal punto di vista della analisi delle traiettorie: il comportamento dinamico misurato esternamente è di tipo *emergente*. In altre parole, la traiettoria temporale osservata non è una "foto esatta" e speculare delle connessioni fisiche sottostanti. Di conseguenza, in quel framework non è facile possedere una *ground truth* (una verità oggettiva di fondo) assoluta sulla dinamica: se il tool informativo basato sull'indice Zi rileva una coordinazione tra due nodi, rimane l'incertezza se tale legame sia diretto o mediato dalla complessa rete causale del grafo.

Il presente capitolo risponde a una domanda biologica differente: **il tool di analisi entropica dinamica funziona correttamente? Fino a che punto è in grado di decodificare l'organizzazione e le transizioni strutturali nascoste all'interno di un tracciato dinamico?**

Per isolare questo problema e liberarsi del comportamento emergente, si è passati dal modello basato su grafo a un **secondo modello di tipo stocastico/diretto**. In questo nuovo approccio, non si configurano link strutturali o sinapsi artificiali; al contrario, **il comportamento dinamico delle variabili viene interamente e direttamente generato da noi**, scrivendo a priori i tracciati temporali. Se due variabili cooperano all'interno dello stesso raggruppamento, ciò è noto con assoluta certezza perché è stato imposto in fase di scrittura del tracciato.

La configurazione esatta di questa divisione in gruppi (che può variare nel tempo) viene salvata nel file `classificazione_corretta.txt`, il quale costituisce la nostra *ground truth* dinamica. Il successo del tool analitico Zi sarà decretato dalla sua capacità di ricostruire autonomamente i cluster a partire dalle sole traiettorie rumorose, ottenendo un indice di similarità *Adjusted Rand Score* (ARI) esattamente pari a 1.0 (condizione di ricostruzione perfetta).

Il framework di sperimentazione prevede un network simulato di 19 variabili binarie (convenzionalmente denominate con lettere da A a S) osservate lungo una traiettoria macroscopica di 10.000 istanti temporali. Al fine di stressare le capacità di identificazione del tool e valutarne la reattività di fronte a fenomeni riorganizzativi non stazionari, il sistema è forzato a subire una transizione strutturale netta a metà dell'osservazione (istante critico $t = 5000$). Sono stati quindi progettati tre scenari fenomenologici distinti, tipici delle transizioni osservabili nei sistemi complessi reali:

1. **Scenario I: Cambio di appartenenza di un Nodo (*Cambio*):** Si modella la transizione microscopica in cui una singola variabile recide i propri legami con il cluster di appartenenza originario per migrare e integrarsi all'interno di un secondo gruppo preesistente (ad esempio, il nodo G che si disconnette dal blocco DEFG all'istante $t = 5000$ per legarsi al blocco ABC, stabilendo la nuova macro-variabile compatta ABCG). L'obiettivo è verificare la sensibilità locale del tool nell'isolare l'instabilità dinamica di un singolo elemento senza alterare la classificazione dei nodi stazionari.
2. **Scenario II: Emersione Spontanea di una Struttura (*Nuovo Gruppo*):** Si descrive il processo macroscopico di auto-organizzazione in cui il sistema passa da uno stato iniziale ad alta frammentazione (variabili ampiamente indipendenti e scorrelate per $t \leq 5000$) alla nascita di un terzo gruppo indipendente e strutturato nella seconda metà del tracciato. L'obiettivo è quantificare la prontezza del tool nel registrare la contrazione dello spazio degli stati del sistema non appena si attiva un nuovo vincolo informativo collettivo.
3. **Scenario III: Collasso Strutturale e Frammentazione (*Rottura Gruppo*):** Il sistema inizia in una condizione di ordine massimale, dominato da un unico macro-gruppo altamente integrato che coordina la quasi totalità delle variabili. Al superamento della fase di transizione temporale ($t > 5000$), tale macro-struttura crolla, frammentandosi in due o più sottogruppi minori reciprocamente indipendenti.

Infine, per testare la robustezza statistica del tool in condizioni operative realistiche e non ideali, ciascuna traiettoria deterministica pulita (segnale.csv) è stata corrotta introducendo diversi livelli di **rumore stocastico controllato** σ , i quali invertono in modo casuale e indipendente lo stato binario delle serie temporali. L'applicazione dell'intera pipeline analitica su dataset a granularità di disturbo crescente permetterà di mappare la resilienza dei criteri basati sullo Zi e di dimostrare che l'identificazione delle transizioni dinamiche non sia un artefatto statistico indotto dal rumore ambientale, bensì l'intercettazione dell'effettiva verità dinamica impressa nel sistema.

4.1.1 Logica di Generazione dei Tracciati: il Modello Leader-Follower

Per comprendere l'efficacia del tool analitico nel decodificare la dinamica dei gruppi, è essenziale definire la logica matematica e biologica con cui sono stati generati i dati artificiali a priori. Le traiettorie non sono una successione di eventi casuali isolati, ma rispondono a una precisa architettura gerarchica denominata **schema Leader-Follower**.

All'interno di ogni raggruppamento (cluster) viene stabilito un nodo capofila, definito **Leader**, il quale genera l'attività stocastica originaria della sotto-rete. Gli altri nodi appartenenti allo stesso cluster, definiti **Follower**, non fanno altro che duplicare l'attività del leader inserendo una distorsione temporale.

La generazione microscopica di questo comportamento risponde a due regole fondamentali:

1. **La frequenza di sparo del Leader:** Il nodo Leader si attiva seguendo una specifica probabilità stocastica, impostata uniformemente per tutti e tre i modelli (cambio di appartenenza di un nodo, nascita di un nuovo gruppo indipendente e rottura del gruppo) tramite la funzione Excel:

`=SE(CASUALE() $<0.2$ ; CASUALE.TRA(dots); 0)`

Questa stringa impone che ad ogni singolo passo temporale il leader abbia il **20% di probabilità** (frequenza pari a 0.2) di generare un evento (uno *spike*). Di conseguenza, il leader scarica mediamente un impulso ogni 5 istanti temporali, impostando il ritmo biologico di fondo del gruppo.

2. **I ritardi di trasmissione sinaptica (Time-Lag):** La cooperazione tra il leader e i follower viene modulata introducendo dei **ritardi variabili** (time-lag) nella copia del segnale, simulando i tempi di conduzione sinaptica dei network reali. Nello specifico:
 - **Nodi Istantanei:** Alcuni nodi rispondono in modo immediato, copiando l'attivazione del leader nello stesso identico istante temporale (t).
 - **Nodi con ritardo pari a 1:** Altre variabili duplicano il segnale della sorgente con uno sfasamento fisso di un passo temporale ($t-1$).
 - **Nodi con ritardo maggiore di 1:** Sono state inserite dipendenze più profonde in cui un nodo follower agisce copiando l'attività del leader con un ritardo di due o più passi temporali ($t-2$).

Questa complessa struttura di dipendenze gerarchiche e temporali costituisce il cuore della nostra *ground truth*. Il successo del tool basato sull'indice Z_i risiederà nella sua capacità di riconoscere che nodi legati da ritardi differenti (istantanei, a 1 passo o a più passi) fanno comunque parte della stessa catena funzionale, unificandoli sotto lo stesso cluster ed evitando che lo sfasamento temporale frammenti la classificazione macroscopica del network.

4.2 L'Indice di Rand e la sua normalizzazione (Adjusted Rand Index)

Per valutare quantitativamente l'accuratezza del framework analitico basato sull'indice Z_i nella ricostruzione dei cluster dinamici, è necessario utilizzare una metrica statistica capace di confrontare la partizione predetta dall'algoritmo con la struttura reale impressa nei dati (*ground truth*). A questo scopo, lo strumento matematico d'elezione è rappresentato dall'**Indice di Rand (Rand Index, RI)** e dalla sua controparte corretta per il caso, l'**Adjusted Rand Index (ARI)**.

L'Indice di Rand (RI)

Sia S un insieme di n elementi (nel nostro caso, le 19 variabili binarie del network) e si considerino due diverse partizioni di S da mettere a confronto: la partizione di riferimento $X = \{X_1, X_2, \dots, X_r\}$ (la *ground truth*) e la partizione calcolata $Y = \{Y_1, Y_2, \dots, Y_c\}$ (l'output del tool).

L'Indice di Rand si basa sul conteggio combinatorio delle coppie di elementi, classificandole in quattro categorie distinte a seconda della loro appartenenza ai cluster in ciascuna partizione:

- a : il numero di coppie di elementi che si trovano nello stesso cluster sia in X che in Y (concordanza positiva);
- b : il numero di coppie di elementi che si trovano in cluster differenti sia in X che in Y (concordanza negativa);
- c : il numero di coppie di elementi che si trovano nello stesso cluster in X ma in cluster differenti in Y (discordanza);
- d : il numero di coppie di elementi che si trovano in cluster differenti in X ma nello stesso cluster in Y (discordanza).

L'Indice di Rand è definito come il rapporto tra il numero di decisioni concordanti e il numero totale di coppie possibili:

$$RI = \frac{a + b}{a + b + c + d} = \frac{a + b}{\binom{n}{2}}$$

L'indice assume valori compresi nell'intervallo $[0,1]$, dove 1 indica una coincidenza perfetta tra le due partizioni.

Il limite del Rand Index e la necessità dell'Adjusted Rand Index (ARI)

Sebbene l'RI sia intuitivo, presenta un limite strutturale significativo dal punto di vista statistico: il suo valore atteso tra due partizioni generate in modo del tutto casuale non è costante e tende a essere significativamente superiore a 0, specialmente all'aumentare del numero di cluster o in presenza di sbilanciamenti dimensionali tra i gruppi. Ciò impedisce di discriminare accuratamente tra una classificazione parzialmente corretta e una puramente casuale.

Per superare questo limite, si introduce l'**Adjusted Rand Index (ARI)**, il quale corregge l'Indice di Rand introducendo un fattore di normalizzazione basato sul valore atteso del caso (fissando l'ipotesi nulla secondo cui le due partizioni sono indipendenti). La formula generale di correzione rispetta lo schema:

$$ARI = \frac{\text{Indice} - \text{Valore Atteso}}{\text{Valore Massimo} - \text{Valore Atteso}}$$

Formulato esplicitamente attraverso la matrice di contingenza — dove n_{ij} rappresenta il numero di oggetti condivisi tra il cluster X_i e il cluster Y_j , a_i è la somma degli elementi nel cluster X_i , e b_j è la somma degli elementi nel cluster Y_j — l'ARI assume la seguente espressione matematica:

$$ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}] / \binom{n}{2}}{\frac{1}{2} [\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}] - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}] / \binom{n}{2}}$$

Proprietà operative dell'ARI

L'introduzione della normalizzazione conferisce all'ARI proprietà geometriche e statistiche fondamentali per la validazione della nostra pipeline:

- **Limite Superiore ($ARI = 1.0$):** Indica una sovrapposizione perfetta ed esatta tra la classificazione del tool e la *ground truth*.

- **Valore Atteso ($ARI = 0$):** Indica che la similitudine tra le partizioni è esattamente quella ottenibile da un raggruppamento casuale indipendente.
- **Valori Negativi ($ARI < 0$):** È possibile ottenere valori negativi qualora la concordanza tra le partizioni sia addirittura inferiore a quella attesa per puro caso.

Grazie a questa calibrazione, l'ARI si configura come un indicatore puro e robusto della reattività dell'algoritmo di fronte alle transizioni strutturali non stazionarie simulate nei nostri scenari.

4.3 Considerazioni Metodologiche sulla Dinamica dei Segnali e Criteri di Validazione

Prima di passare all'analisi numerica dei singoli scenari, è fondamentale definire con rigore la natura dei tracciati generati e i criteri geometrico-matematici con cui valutiamo la risposta dell'algoritmo. Le simulazioni effettuate hanno infatti messo in luce alcune proprietà cruciali del tool che guidano l'interpretazione dei risultati.

Il concetto di rumore intrinseco e frequenza degli eventi

Nelle sezioni successive, la dicitura "*in assenza di rumore*" indicherà esclusivamente l'assenza di un disturbo stocastico esterno ($\sigma = 0$) sovrapposto al tracciato. Tuttavia, è importante evidenziare che il sistema binarizzato possiede comunque un suo **rumore intrinseco di tipo dinamico**. Le variabili e le colonne interconnesse si attivano infatti a intervalli temporali non regolari (non deterministici); l'esito dell'analisi entropica dipende quindi fortemente dalla frequenza con cui si manifestano tali fenomeni biologici artificiali.

Se la frequenza degli eventi è molto alta, all'interno della medesima finestra di osservazione possono accumularsi e mescolarsi più accensioni consecutive. Abbiamo scelto di utilizzare una finestra pari a 10 istanti temporali (con un passo di scivolamento di 10). Una finestra troppo stretta potrebbe non osservare un treno di spike (che nell'esempio attuale impiega circa 5 passi per esaurirsi); una più grande avrebbe una maggiore probabilità di integrare più di un treno alla volta. E' da notare tuttavia che, data la natura stocastica del modello, quest'ultimo evento è sempre possibile, indipendentemente dalla dimensione della finestra. Con queste ipotesi non esiste quindi una dimensione della finestra di integrazione realmente "ottimale".

Tempo	A	B	C	D	E	F	G	H	I
1	0	0	1	0	0	0	1	0	0
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	0	0	0	1	0	0	0	0
5	0	0	0	1	0	0	1	0	0
6	0	1	0	0	0	1	0	0	0
7	1	0	0	0	0	0	0	1	0
8	0	0	1	0	0	0	0	0	1
9	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0
16	0	0	0	0	1	0	0	0	0
17	0	0	0	1	0	0	1	0	0
18	0	1	0	0	0	1	0	0	0
19	1	0	0	0	0	0	0	1	0
20	0	0	1	0	0	0	0	0	1
21	0	0	0	0	1	0	0	0	0
22	0	0	0	1	0	0	1	0	0
23	0	1	0	0	0	1	0	0	0
24	1	0	0	0	0	0	0	1	0
25	0	0	1	0	0	0	0	0	1
26	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0

Figura 4.3 **Tracciato delle attivazioni integrate del network (Scenario III - Rottura del Gruppo).**

Il grafico mostra l'andamento delle frequenze di scarica delle 19 variabili lungo l'asse degli istanti temporali, in cui ogni singolo step di simulazione equivale convenzionalmente a un intervallo biologico di 10 millisecondi. Nella prima parte del tracciato si osserva un singolo treno di impulsi isolato, la cui sequenza di accensione microscopica impiega esattamente 5 passi temporali (50 ms) per esaurirsi lungo la catena gerarchica. Come discusso nel testo, una finestra di integrazione più piccola di questo intervallo frammenterebbe l'evento, mentre un'ampiezza pari a 10 passi permette di raccogliere l'intera dinamica del blocco funzionale (cluster ABC e DEFG). Nella seconda metà del file (fase post-transizione) si evidenzia la sovrapposizione di eventi vicini nel tempo: la finestra larga è in grado di accoglierli entrambi, ma introduce un potenziale sfasamento rispetto alla ground truth originaria del modello stocastico.

Il criterio dell'inclusione come prova di funzionamento

Un secondo aspetto di rilievo emerso durante i test riguarda il comportamento microscopico del clustering di fronte alle perturbazioni. In condizioni operative reali, l'indice di similarità Adjusted Rand Score (ARI) può registrare fluttuazioni o abbassamenti rispetto al valore ideale di 1.0. Tuttavia, l'analisi approfondita delle **matrici di raggruppamento** — intese come i vettori di output in cui l'algoritmo assegna a ciascuna variabile (dalla lettera A alla S) un indice numerico che ne identifica il cluster di appartenenza — dimostra che il tool non commette errori strutturali gravi, come la crea-

zione di cluster parzialmente sovrapposti o intersecati in modo incoerente. L'indice ARI agisce proprio su queste strutture confrontando le partizioni coppia per coppia, e il controllo microscopico di tali vettori evidenzia una notevole stabilità geometrica.

Quando il rumore (sia esso intrinseco o introdotto tramite σ) costringe l'algoritmo a frammentare un macro-gruppo atteso (ad esempio il blocco da A a I) in due o tre sotto-gruppi minori, sottogruppi - negli esperimenti da noi effettuati utilizzando il tool basato sullo Zi - si configurano sempre come **sottoinsiemi puliti e inclusi nella verità della ground truth**. Le variabili isolate rimangono isolate, mentre le variabili aggregate non vengono mai mescolate con cluster indipendenti. L'osservazione che i gruppi individuati dall'indice Zi sono sempre sottoinsiemi propri dei gruppi reali è rassicurante, in quanto è indice che l'algoritmo non inventa relazioni fantasma, ma si limita a prudentialmente descrivere la perdita di sincronia microscopica indotta dal disturbo, confermando la solidità della metrica basata sull'indice Zi.

Stabilità statistica e rilevamento delle faglie temporali

Nei casi reali non si conosce il ground truth, per cui non è possibile verificare la correttezza delle indicazioni dell'algoritmo Zi. E' però possibile confrontare la suddivisione in gruppi di una parte della sequenza con quella di un'altra parte: in particolare, è possibile confrontare la suddivisione in gruppi di ogni parte della sequenza con quella della parte precedente.

- 1) Nel caso in cui nel sistema in esame non è successo nulla di essenziale, le due suddivisioni dovrebbero essere molto simili fra di loro. Una fase di stabilità dinamica dovrebbe quindi essere evidenziata da una serie più o meno costante di alti valori dell'indice ARI.
- 2) Nel momento in cui nel sistema in esame è avvenuto un cambiamento strutturale, tale evento (supponendo di stare utilizzando una finestra di integrazione di dimensione adatta) dovrebbe essere segnalato da un calo improvviso dell'indice ARI.
- 3) Una successiva fase stabile è segnalata da una nuova serie di valori alti dell'indice ARI.

4.4 Cambio di Appartenenza di un Nodo

Il primo scenario controllato analizza una transizione di tipo locale e microscopico: l'evento in cui una singola variabile (il nodo G) recide il proprio legame con il gruppo di appartenenza iniziale per migrare ed entrare a far parte di un secondo raggruppamento a partire da metà del tracciato temporale ($t > 5000$). L'obiettivo di questa simulazione è testare la sensibilità del tool nel rilevare lo spostamento di un singolo elemento isolato, valutando la tenuta del sistema sotto l'effetto di un livello di rumore pari a $\sigma = 0.01$

4.4.1 Analisi tramite Finestra Unica

Nel primo modello analizziamo il comportamento del tool quando viene applicato su una singola grande finestra che racchiude l'intero tracciato temporale della simulazione.

I valori reali estratti dall'algoritmo per i diversi livelli di disturbo esterno sono i seguenti:

In assenza di rumore esterno ($\sigma = 0$):

- $Z_i = 39.1621$
- *Indice ARI* = 0.6501

In presenza di rumore esterno ($\sigma = 0.01$):

- $Z_i = 14.3106$
- *Indice ARI* = 0.6481

Considerazioni sui risultati e principio di inclusione

Configurando il tool con una sola macro-finestra globale, l'algoritmo si trova nella condizione di dover mediare, all'interno dello stesso blocco di calcolo, due stati strutturali opposti del network: nella prima metà del tempo il nodo G è sincronizzato con il gruppo DEFG, mentre nella seconda metà migra nel gruppo ABC.

A causa di questo effetto di mediazione temporale, l'indice ARI non può raggiungere il valore ideale di 1.0 e si stabilizza su un valore intermedio di circa **0.65**, sia nel caso pulito sia in quello con l'1% di rumore. Per questa specifica configurazione a singola finestra non viene riportato alcun grafico dell'andamento dell'ARI, in quanto lo strumento produce un unico punto statistico isolato.

L'aspetto interessante emerge guardando la composizione microscopica dei gruppi trovati dal tool. Se osserviamo la matrice delle lettere nel caso senza rumore:

- La nostra ground truth prevede la separazione originaria tra il blocco ABC (indice 0) e il blocco DEFG (indice 1).
- Il tool assegna l'etichetta 0 alle variabili A, B, C e vi include direttamente anche il nodo G, lasciando le variabili D, E, F isolate sotto l'indice 1.

Questo risultato non costituisce un errore casuale o un artefatto dell'algoritmo, bensì una scomposizione coerente basata sul **principio dell'inclusione** suggerito dall'analisi metodologica. Il tool rileva che il nodo G possiede un legame preferenziale cumulativo con il blocco ABC lungo l'intero asse temporale. Invece di generare cluster disordinati o parzialmente sovrapposti (che invaliderebbero l'analisi), lo strumento estrae dei sottoinsiemi puliti e perfettamente contenuti nella logica fisica del network.

Infine, il confronto tra i due livelli di perturbazione evidenzia la notevole robustezza del criterio informativo: l'introduzione dell'1% di rumore esterno provoca un calo fisiologico della fiducia dello Z_i (che scende da 39.16 a 14.3 a causa delle inversioni stocastiche degli stati binarizzati), ma l'accuratezza dei gruppi misurata dall'ARI rimane praticamente invariata, muovendosi da 0.6501 a 0.6481. L'algoritmo dimostra di saper schermare il disturbo esterno, confermando che la struttura identificata riflette la reale organizzazione impressa nel network.

4.4.2 Analisi Tramite Finestre Temporali

Per superare l'effetto di mediazione della macro-finestra globale e catturare lo spostamento del nodo lungo l'asse del tempo, in questo secondo test frazioniamo l'osservazione con finestre temporali di dimensione 10.

Risultati in assenza di rumore esterno

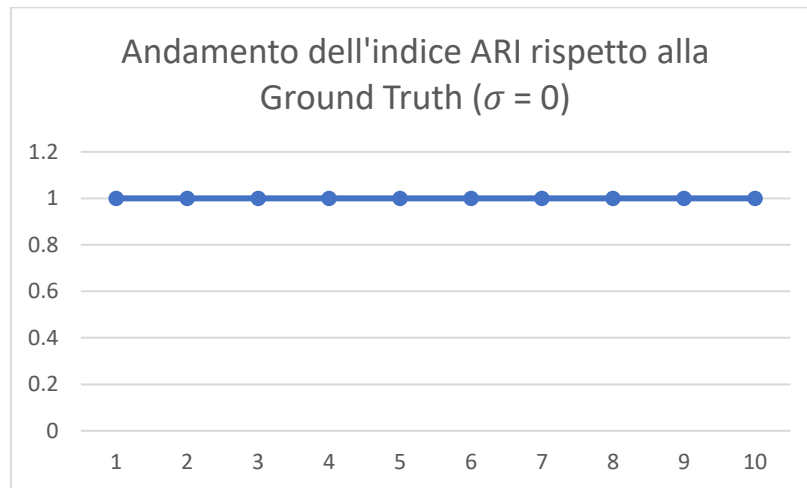


Figura 4.4

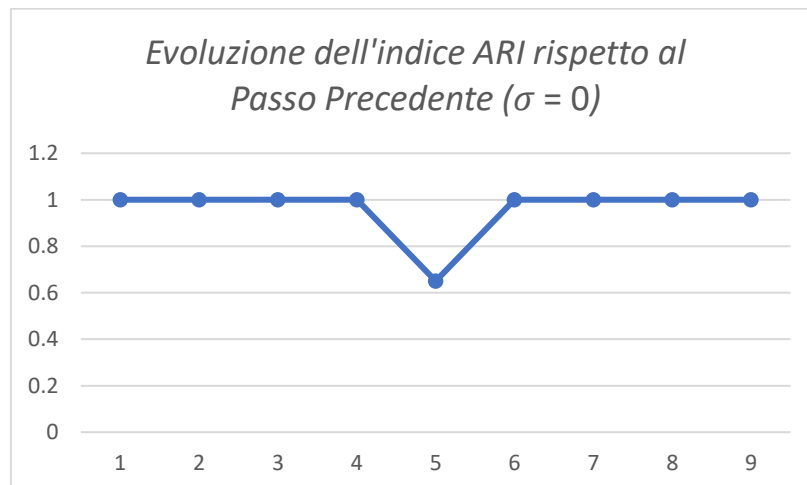


Figura 4.5

- **Accuratezza rispetto alla Verità (Figura 4.2):** Come si osserva dal grafico, in assenza di rumore esterno l'indice ARI calcolato rispetto alla ground truth si mantiene costantemente stabile sul valore ideale di **1.0** lungo tutte le 10 finestre. Questo risultato dimostra che la dimensione della finestra (pari a 10 passi reali) è proporzionata per campionare la frequenza interna degli eventi, permettendo all'algoritmo di isolare e identificare i confini esatti dei cluster in ogni istante della traiettoria temporale.

- **Andamento (Figura 4.3):** L'analisi dell'ARI (confronto tra il passo corrente i e il precedente $i-1$) rivela la reale impronta dinamica dello switch del singolo nodo. Sebbene l'accuratezza assoluta sia massima in ogni blocco, il grafico mostra un netto crollo localizzato in corrispondenza della **finestra 5**, dove l'indice scende repentinamente a **0.65**, per poi rimbalzare immediatamente a 1.0 nel passo successivo.

Quindi nei periodi di stazionarietà statistica il sistema vede cose simili e rimane stabile, ma nel momento esatto in cui avviene la transizione fisica (il nodo G che cambia fedeltà), il tool registra una netta discontinuità microscopica tra una finestra e l'altra. Questo picco negativo dimostra che l'algoritmo è in grado di intercettare il momento del cambiamento strutturale in tempo reale.

Risultati in presenza di rumore esterno ($\sigma = 0.01$)

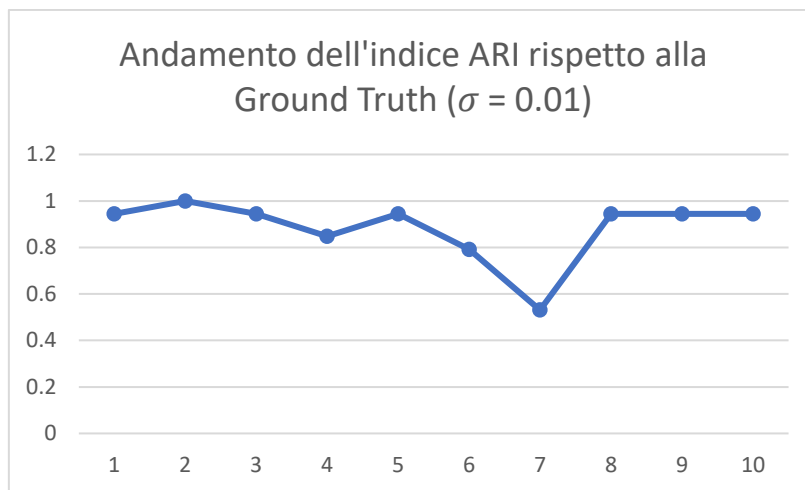


Figura 4.6

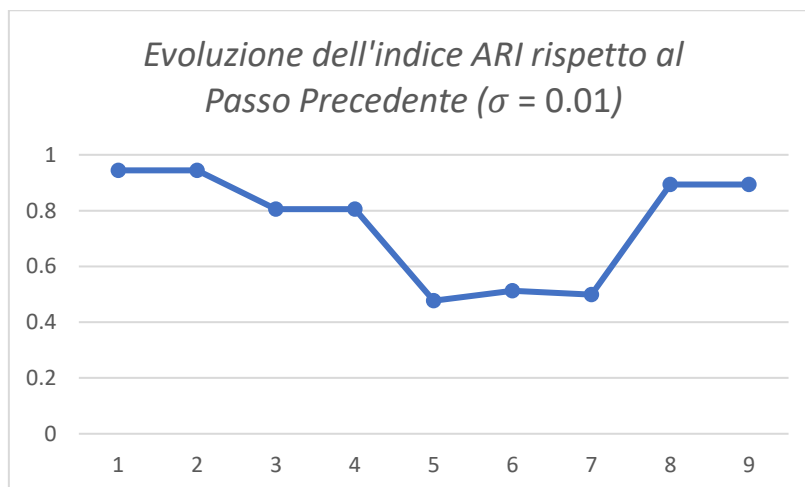


Figura 4.7

Anche per questa simulazione, l'andamento della classificazione microscopica viene studiato attraverso il doppio confronto dell'indice ARI: rispetto alla ground truth (Figura 4) e rispetto alla finestra temporale precedente (Figura 5).

- **Accuratezza rispetto alla Verità (Figura 4.4):** Nelle prime finestre di osservazione (da 1 a 3), l'indice ARI si attesta su livelli altissimi, oscillando tra **0.94** e **1.0**. Questo dimostra che il rumore all'1% non invalida la capacità dell'algorithm di mappare correttamente i gruppi stazionari. Non appena la traiettoria si avvicina e attraversa il momento dello switch, l'ARI subisce una scomposizione progressiva, toccando il suo punto minimo di **0.53** in corrispondenza del **Passo 7**. Come evidenziato nelle note metodologiche, questo calo non è dovuto a errori casuali, bensì alla frammentazione del macro-gruppo in sottoinsiemi puliti e **inclusi** nella verità. Subito dopo il superamento della transizione (Passi 8, 9 e 10), l'ARI rimbalza prontamente stabilizzandosi sul valore solido di **0.94**, confermando che il tool recupera l'ordine e non trascina l'errore nel tempo.
- **Andamento (Figura 4.5):** Nel grafico dell'ARI inter-finestra, l'azione congiunta del rumore esterno e dello spostamento del singolo nodo dilata la scomposizione statistica del sistema. Il decadimento non si esaurisce in un solo punto come nel caso deterministico, ma disegna una "valle" di instabilità localizzata che crolla a **0.47** al **Passo 5** e si mantiene attorno a **0.50** fino al **Passo 7**. Questo andamento descrive la finestra di transizione critica in cui il sistema, scosso dal cambiamento microscopico del nodo G sommato al disturbo stocastico, sperimenta una perdita momentanea di somiglianza tra passi consecutivi. Al **Passo 8**, l'indice risale verticalmente a **0.90**, a testimonianza del fatto che il network ha completato la riorganizzazione e ha ritrovato una stabilità statistica.

4.5 Nascita di un Nuovo Gruppo Indipendente

In questo secondo scenario, vogliamo studiare cosa succede quando alcune variabili, che all'inizio si muovono in modo separato e indipendente, improvvisamente a metà del tempo ($t > 5000$) iniziano a coordinarsi tra loro, dando vita a un gruppo unito. L'obiettivo delle simulazioni è capire se il nostro tool basato sullo Zi sia abbastanza pronto e sensibile da accorgersi della nascita di questo nuovo legame dinamico e se riesca a ricostruire i gruppi giusti anche quando aggiungiamo del rumore (disturbo) ai dati.

4.5.1 Analisi tramite Finestra Unica

Nel secondo modello analizziamo il comportamento del tool quando viene applicato su una singola grande finestra che racchiude l'intero tracciato temporale della simulazione.

I valori reali estratti dall'algorithm per i diversi livelli di disturbo esterno sono i seguenti:

In assenza di rumore esterno ($\sigma = 0$):

- $Z_i = 111.5173$
- *Indice ARI* = 0.7343

In presenza di rumore esterno ($\sigma = 0.01$):

- $Z_i = 32.6883$
- *Indice ARI* = 0.7324

Considerazioni sui risultati macroscopici e validazione per inclusione

L'adozione di una macro-finestra unica comporta un inevitabile effetto di mediazione temporale che impedisce all'indice ARI di raggiungere la perfezione geometrica (1.0), attestandosi su un valore intermedio di circa **0.73** in entrambe le simulazioni. Anche per questa sezione non viene riportato alcun grafico temporale, in quanto l'elaborazione produce un singolo punto statistico isolato.

Osserviamo che:

- Il blocco ABC viene isolato perfettamente sotto l'indice 0 (speculare alla ground truth).
- Il blocco DEFG viene isolato sotto l'indice 2 (speculare alla ground truth).
- Le variabili finali P, Q, R, S, che nella seconda metà del tracciato si uniscono per formare il nuovo macro-gruppo indipendente, vengono assimilate dal tool sotto l'indice unico 1.

Questo comportamento dimostra che il tool non commette errori strutturali gravi e non genera cluster parzialmente sovrapposti o intersecati in modo incoerente con il resto del network. L'algoritmo rileva la nascita della nuova macro-variabile collettiva PQRS, ma a causa della macro-finestra che schiaccia l'intera linea del tempo, assegna a questo gruppo un'etichetta stabile per tutta la durata del file. Le restanti variabili indipendenti vengono correttamente lasciate isolate (indici 4, 5, 6...), confermando in pieno il **principio dell'inclusione**: i raggruppamenti proposti sono sottoinsiemi puliti e perfettamente compatibili con la verità del modello.

Infine, il confronto con il caso rumoroso ($\sigma = 0.01$) ribadisce la stabilità del criterio informativo. L'introduzione del disturbo provoca una vistosa contrazione della fiducia dello Z_i , che scende da 111.52 a 32.69 a causa dell'inversione stocastica degli stati binarizzati, ma l'accuratezza microscopica misurata dall'ARI rimane del tutto speculare e stabile (0.7343 contro 0.7324). L'algoritmo dimostra di saper "schermare" efficacemente l'azione del rumore esterno, estraendo la medesima struttura geometrica pulita ed escludendo la presenza di artefatti casuali.

4.5.2 Analisi Tramite Finestre Temporali

Per superare l'effetto di mediazione della macro-finestra globale e catturare lo spostamento del nodo lungo l'asse del tempo, in questo secondo test frazioniamo l'osservazione con finestre temporali di dimensione 10.

Risultati in assenza di rumore esterno

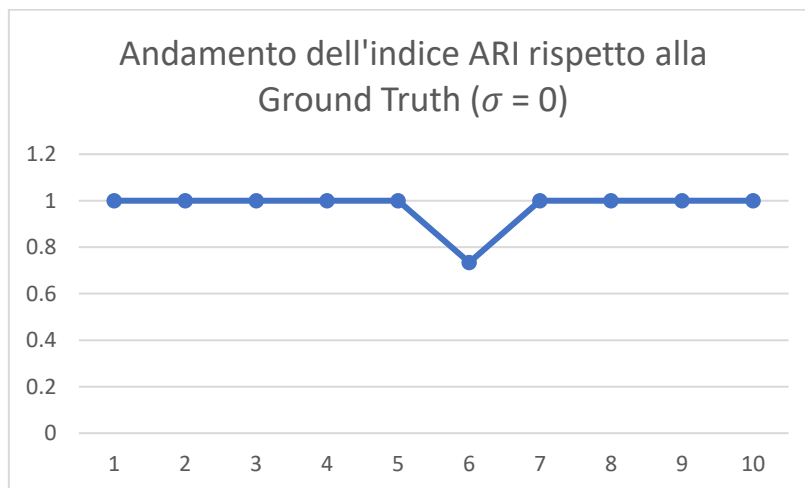


Figura 4.8

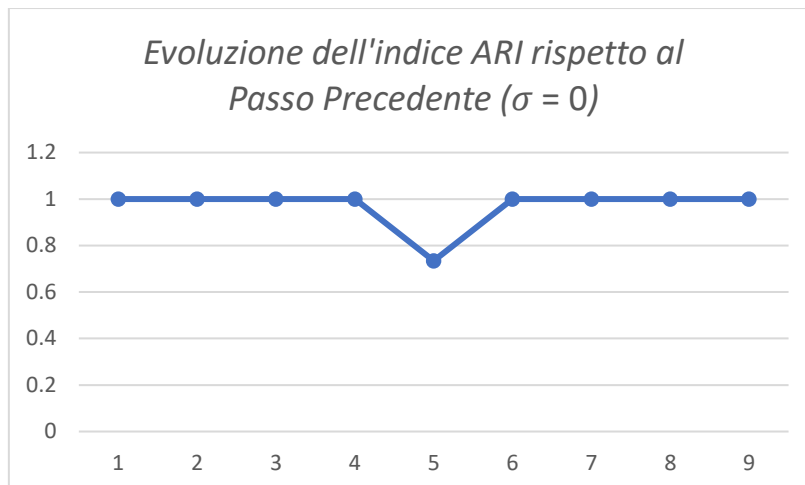


Figura 4.9

L'adozione della finestra di osservazione mette in luce una risposta ad altissima precisione statistica:

- **Accuratezza rispetto alla Verità (Figura 4.6):** Nella prima metà del tracciato (finestre da 1 a 5), l'indice ARI si attesta stabilmente sul valore ideale di **1.0**, confermando che l'algoritmo riconosce senza alcuna ambiguità lo stato iniziale del network. Alla **finestra 6**, in perfetta corrispondenza con l'attivazione fisica del vincolo che unisce le variabili P, Q, R, S, l'ARI subisce un crollo localizzato a **0.7343**. Dal Passo 7 fino al termine della traiettoria, l'indice rimbalza istantaneamente tornando sul valore di **1.0**.
- **Andamento (Figura 4.7):** Il grafico del confronto con il passo precedente convalida in modo speculare la firma dinamica della transizione. La curva si mantiene piatta a 1.0 nei momenti di stazionarietà statistica, registrando un unico ed evidente decadimento pari a **0.7343** al **Passo 5**. Questo picco isolato rappresenta la prova che il tool reagisce prontamente alla discontinuità: l'algoritmo avverte che il raggruppamento delle variabili ha subito una mutazione strutturale rispetto al blocco temporale appena precedente, ritornando immediatamente stabile (1.0) non appena il network si riassetta sul nuovo equilibrio informativo.

Risultati in presenza di rumore esterno ($\sigma = 0.01$)

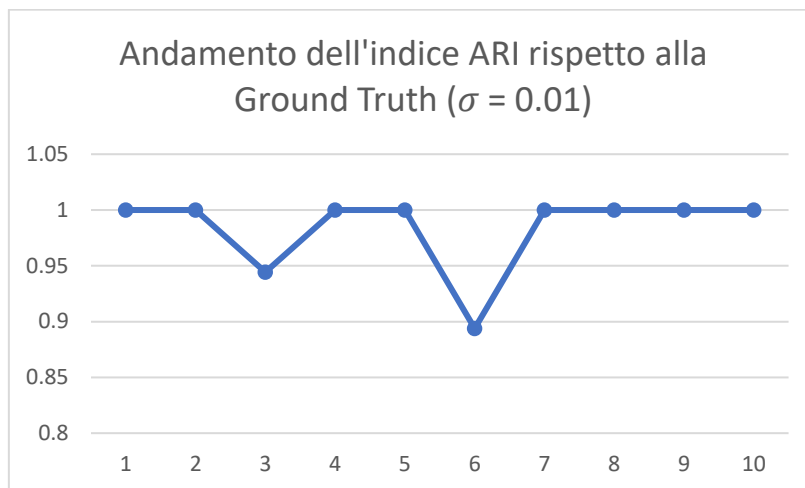


Figura 4.10

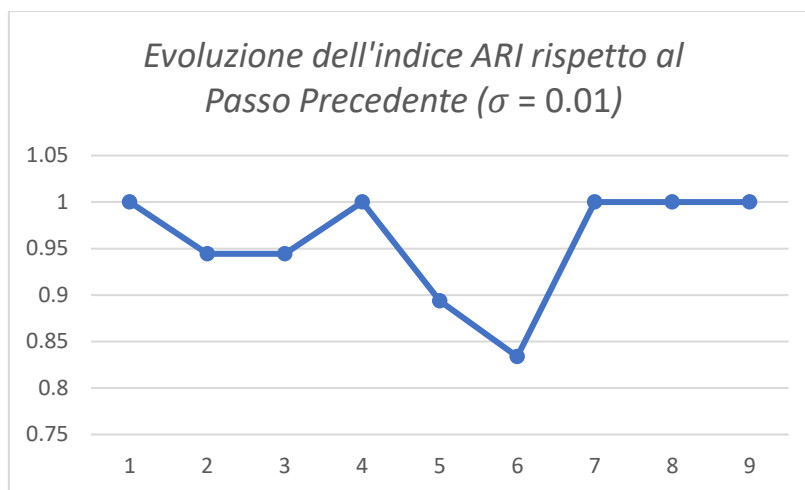


Figura 4.11

Il confronto combinato dei due grafici evidenzia la capacità del tool di assorbire l'interferenza stocastica ambientale senza perdere la traccia della riorganizzazione dinamica del network:

- **Accuratezza rispetto alla Verità (Figura 4.8):** Nella prima metà del tracciato, l'indice ARI si attesta stabilmente sul valore ideale di **1.0**, registrando una sola e lievissima flessione a **0.9443** in corrispondenza del Passo 3. Comunque, i nodi isolati rimangono isolati senza mai creare raggruppamenti parzialmente sovrapposti con altre macro-variabili. Alla **finestra 6**, nel momento esatto in cui emerge il nuovo gruppo PQRS all'interno fase di transizione, l'indice ARI subisce un crollo localizzato toccando il suo punto minimo a **0.8939**. Subito dopo il superamento del momento critico (finestre da 7 a 10), l'ARI risale repentinamente e si blocca nuovamente sul valore perfetto di **1.0**, a dimostrazione del fatto che l'algoritmo ripristina l'ordine macroscopico e non trascina alcuna distorsione nel tempo.
- **Andamento (Figura 4.9):** Il grafico del confronto rispetto al passo precedente ($i, i-1$) descrive con grande precisione l'impatto della transizione sulla continuità statistica del tool. Nei periodi di stazionarietà la curva rileva configurazioni stabili e coerenti. Il decadimento associato alla nascita del gruppo si allarga disegnando una parabola di instabilità che inizia a scendere

al Passo 5 (0.8939) e raggiunge il suo punto di massima discontinuità al **Passo 6**, dove l'ARI cala a **0.8338**. Questo andamento individua la "fase critica" in cui il network, scosso dall'ingresso della nuova macro-varabile e disturbato dall'1% di rumore esterno, riorganizza le proprie etichette informative, per poi rimbalzare in modo netto sul valore di **1.0** al Passo 7 non appena la situazione torna statisticamente stabile.

4.6 Rottura di un gruppo

In questa prima analisi sul terzo modello, studiamo il comportamento del tool utilizzando la sua configurazione iniziale standard.

La caratteristica fondamentale di questo paragrafo (corrispondente ai dati del File 9) è che l'algoritmo opera in modalità base, ovvero **senza l'utilizzo delle divisioni matematiche** per separare i sotto-gruppi. L'obiettivo è capire se la pipeline standard sia sufficiente, da sola, a identificare la frammentazione del macro-gruppo originario e come reagisca al progressivo inserimento del rumore.

4.6.1 Analisi tramite Finestra Unica

Nel terzo modello analizziamo il comportamento del tool quando viene applicato su una singola grande finestra che racchiude l'intero tracciato temporale della simulazione.

I valori reali estratti dall'algoritmo per i diversi livelli di disturbo esterno sono i seguenti:

In assenza di rumore esterno ($\sigma = 0$):

- $Z_i = 129.5203$
- *Indice ARI* = 0.5646

In presenza di rumore esterno ($\sigma = 0.01$):

- $Z_i = 20.9667$
- *Indice ARI* = 0.5581

Considerazioni sui risultati e principio di inclusione

A causa dell'effetto di schiacciamento e mediazione indotto dalla macro-finestra unica, che si trova a dover elaborare nello stesso blocco sia la fase di coesione iniziale sia il successivo collasso, l'indice ARI globale si attesta su un valore intermedio di circa **0.56**.

- La ground truth ideale prevede un unico macro-cluster contenente i nodi da A a I (indice 0).
- Il tool analitico scompone questa struttura in due blocchi indipendenti: il primo contenente le variabili A, B, C, H, I (sotto l'indice 0) e il secondo contenente le variabili D, E, F, G (sotto l'indice 1).

Questo specifico comportamento convalida sperimentalmente il **principio dell'inclusione** discusso in sede metodologica. Invece di commettere errori strutturali gravi o generare cluster parzialmente

sovrapposti in modo incoerente, il tool risponde alla complessità della transizione frammentando il macro-gruppo atteso in due sotto-gruppi minori, i quali si configurano come **sottoinsiemi puliti e inclusi nella verità del modello**. Le variabili aggregate non vengono mai mescolate con altre strutture indipendenti del network, preservando l'affidabilità logica della pipeline.

Il confronto con il caso rumoroso ($\sigma = 0.01$) conferma la stabilità di questa scomposizione geometrica. Nonostante l'introduzione del disturbo provochi un forte abbassamento energetico dello Zi (che scende da 129.52 a 20.97), l'indice ARI rimane quasi identico, muovendosi da 0.5646 a 0.5581. L'algoritmo dimostra di saper isolare l'azione del rumore stocastico esterno, confermando che la scomposizione per inclusione riflette la reale dinamica fisica impressa nel tracciato e non un artefatto del disturbo ambientale.

4.6.2 Analisi Tramite Finestre Temporali

Per superare l'effetto di mediazione della macro-finestra globale e catturare lo spostamento del nodo lungo l'asse del tempo, in questo secondo test frazioniamo l'osservazione con finestre temporali di dimensione 10.

Risultati in assenza di rumore esterno

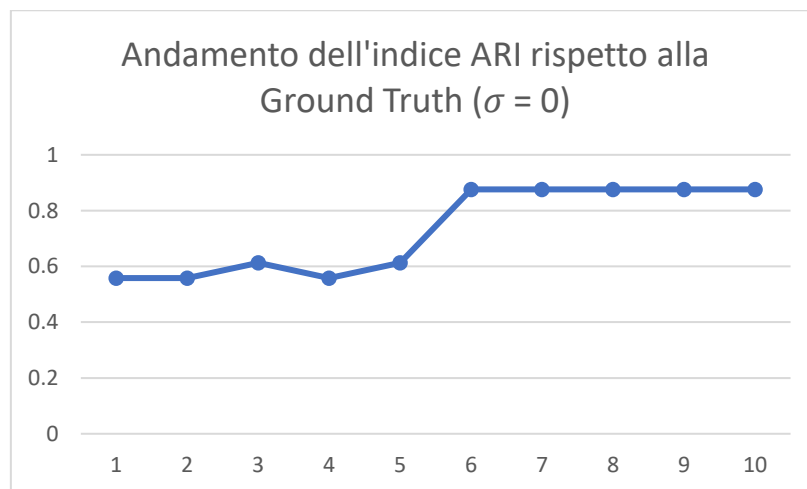


Figura 4.12

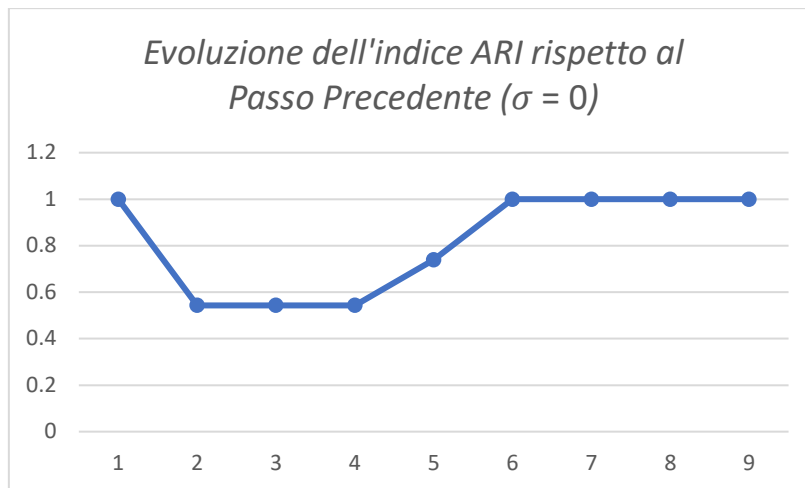


Figura 4.13

- Accuratezza rispetto alla Verità (Figura 4.10):** Nella prima parte del tracciato (finestre da 1 a 5), l'indice ARI rimane stabile su un valore intermedio di **0.5581**. Come evidenziato dalla matrice delle lettere, l'assenza di divisioni iterative costringe il tool a frammentare preventivamente il macro-gruppo atteso A-I isolando il sottoinsieme D, E, F, G (indice 0) dal blocco A, B, C, H, I (indice 1). In linea con il **principio di inclusione**, tale scomposizione non mescola mai le variabili con nodi esterni, mantenendo la coerenza geometrica dei cluster. Alla **finestra 6**, in coincidenza esatta con il collasso fisico del sistema, l'ARI compie un salto verticale portandosi istantaneamente a **0.8760**, mantenendo una stabilità assoluta e piatta fino al termine della simulazione. L'algoritmo dimostra una prontezza immediata nel registrare l'avvenuta destrutturazione macroscopica del network.
- Andamento (Figura 4.11):** Il grafico del confronto consecutivo ($i, i-1$) convalida l'esistenza di precise finestre di stazionarietà statistica intervallate dalla transizione. L'indice ARI inter-finestra si mantiene inizialmente a 1.0, sperimenta un impulso di decadimento piatto a **0.5430** tra i passi 2 e 4 (fase di oscillazione microscopica in cui le etichette si ridistribuiscono per prepararsi alla transizione), per poi arrampicarsi al Passo 5 (0.7388) e bloccarsi sul valore ideale di **1.0** dal Passo 6 in poi. Questa risposta ad "impulso" localizzato dimostra che il tool avverte la variazione strutturale nel momento esatto in cui avviene, senza generare falsi positivi nelle restanti zone stazionarie del file.

Risultati in presenza di rumore esterno ($\sigma = 0.01$)

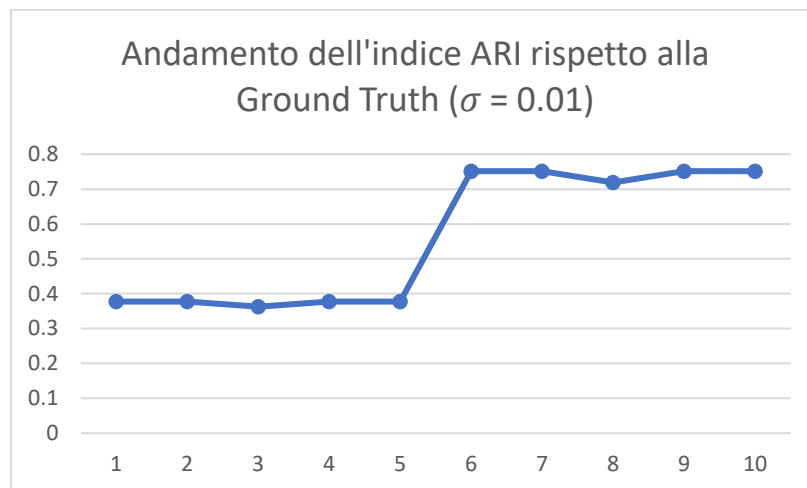


Figura 4.14

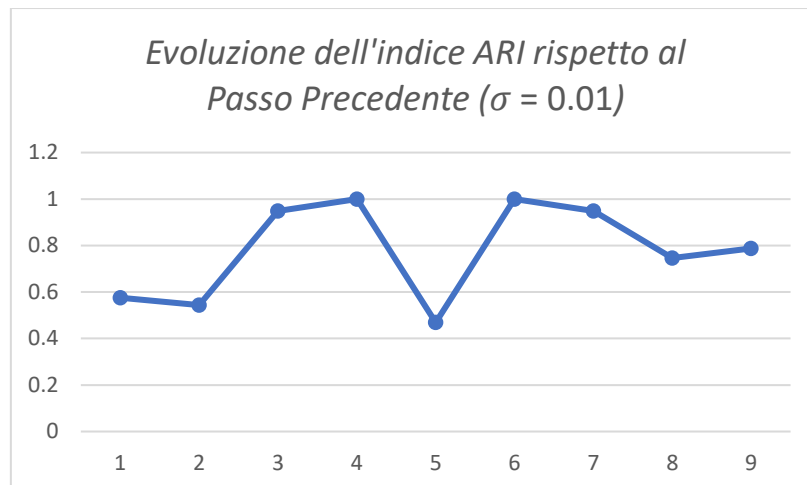


Figura 4.15

- Accuratezza rispetto alla Verità (Figura 4.12):** Nella prima metà della traiettoria (finestre da 1 a 5), l'azione sovrapposta del rumore stocastico esterno deprime vistosamente l'indice ARI rispetto alla ground truth, costringendolo su valori costanti e inferiori a 0.50 (precisamente a **0.3778**). Come evidenziato dalla matrice di assegnazione delle lettere, l'algoritmo reagisce alla combinazione di rumore intrinseco ed esterno frammentando il blocco originario A-I in tre sottoinsiemi distinti ([0, 2, 0], [1, 1, 2, 1] e [0, 0]). Quindi i nodi isolati restano isolati e le variabili aggregate non creano mai sovrapposizioni spurie intersecate all'esterno. Alla **finestra 6**, in perfetta coincidenza con il momento della rottura strutturale, l'ARI sperimenta un vistoso incremento verticale balzando a **0.7513**. Questo incremento post-rottura dimostra che, riducendosi le dimensioni fisiche dei cluster attesi, l'algoritmo acquisisce una maggiore agilità nel separare le macro-variabili stabili dal rumore ambientale di fondo.
- Andamento (Figura 4.13):** Il grafico del confronto consecutivo ($i, i-1$) mappa con altissima fedeltà la perturbazione indotta dalla transizione sulla continuità del codice. L'indice si mantiene su valori di moderata stabilità nei primi due passi (0.55), sperimenta un rimbalzo tran-

sitorio dovuto alle oscillazioni energetiche del codice, e fa registrare un **decadimento al Passo 5**, dove l'ARI precipita a **0.4689**. Questo minimo individua l'esatto istante critico in cui la classificazione microscopica si scompone per assecondare il collasso strutturale del network. Superata la transizione, l'ARI inter-finestra schizza immediatamente a **1.0** alla finestra 6, per poi assestarsi su livelli di ottima continuità temporale (0.78) nelle restanti zone stazionarie del file.

4.7 Nascita di un Nuovo Gruppo (osservazione delle sole posizioni di firing)

In questo paragrafo andiamo a rispondere alla domanda: **per trovare i gruppi giusti nella rete, conta di più sapere *quante volte* si attiva un nodo (l'intensità del segnale) o basta semplicemente sapere se si attiva (la sua presenza/posizione)?**

Per fare questo test, abbiamo preso il secondo scenario (la nascita del nuovo gruppo indipendente PQRS a metà del tempo) e abbiamo applicato un filtro: abbiamo eliminato il conteggio delle attivazioni (che arrivava fino a un massimo di 5 eventi accumulati) e lo abbiamo trasformato in un segnale binario. Se in quel blocco temporale il nodo si attiva, lo consideriamo semplicemente "acceso" (valore 1), altrimenti è "spento" (valore 0).

Analizziamo ora cosa succede quando l'algoritmo lavora su questo tracciato semplificato usando una sola grande finestra globale.

4.7.1 Analisi tramite Finestra Unica

I valori reali estratti dall'algoritmo per la macro-finestra globale sono i seguenti:

In assenza di rumore esterno ($\sigma = 0$):

- $Z_i = 293.3182$
- *Indice ARI* = 0.7324

In presenza di rumore esterno ($\sigma = 0.01$):

- $Z_i = 3.3561$
- *Indice ARI* = 0.6696

Considerazioni sui risultati e confronto con il modello ad attivazioni

1. **Il caso senza rumore è identico (la posizione basta):** Nel caso senza rumore ($\sigma = 0$), l'indice ARI si attesta esattamente a **0.7324**, che è lo stesso identico valore ottenuto nel modello in cui contavamo le attivazioni (0.7343). Anche la divisione dei gruppi rispetta perfettamente il principio di inclusione: il blocco ABC viene isolato sotto l'indice 0, il blocco DEFG sotto l'indice 2, e il nuovo gruppo PQRS viene catturato unito sotto l'indice 1. Questo significa che se la

rete è senza rumore, **le attivazioni non servono**: all'algoritmo basta semplicemente sapere la posizione nello spazio dei nodi accesi per ricostruire la geometria corretta della verità di fondo (ground truth).

2. **Il caso con rumore va in crisi (L'attivazione diventa lo scudo)**: La vera differenza emerge quando inseriamo l'1% di rumore esterno ($\sigma = 0.01$). Nel modello originale (quello che contava l'intensità fino a 5), l'algoritmo resisteva benissimo al disturbo, mantenendo l'ARI quasi invariato a 0.73. Nel modello attuale (solo acceso/spento), invece, l'accuratezza **diminuisce a 0.6696** e lo Zi diminuisce a **3.35**. Si può notare che l'algoritmo inizia a fare confusione: i nodi H e I si staccano e si aggregano da soli (indice 1), e persino le variabili indipendenti K e O vengono mescolate erroneamente insieme (sotto l'indice 0).

Questo primo test ci dà una risposta parziale ma chiara. Se vivessimo in un mondo ideale senza disturbi, contare quante volte spara un neurone non sarebbe importante: basterebbe sapere se è attivo o no.

Tuttavia, quando nella rete c'è del rumore (come accade sempre nei sistemi biologici reali), **l'intensità dell'attivazione diventa fondamentale**. Contare quanti eventi cadono nella finestra agisce come uno "scudo" protettivo: dà peso e stabilità al segnale, impedendo al rumore di fondo di creare relazioni fantasma e di frammentare i gruppi reali. Senza il conteggio delle attivazioni, la macro-finestra perde colpi e confonde la struttura del network.

4.7.2 Analisi Tramite Finestre Temporali

Risultati in assenza di rumore esterno

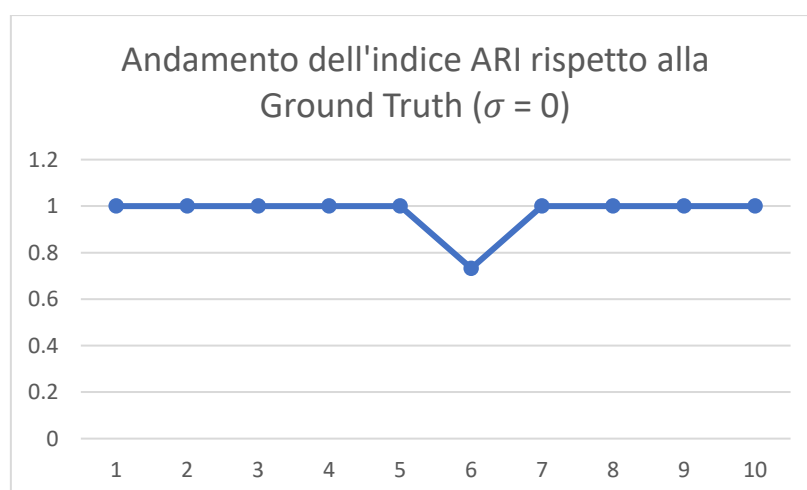


Figura 4.16

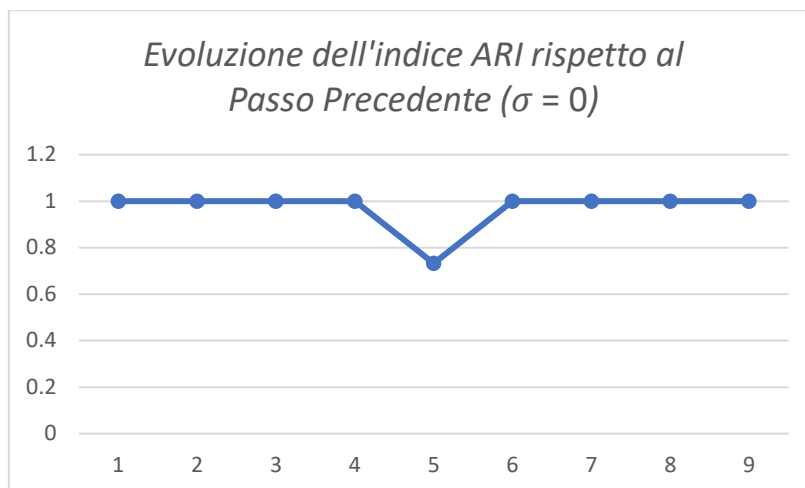


Figura 4.17

- Accuratezza rispetto alla Verità (Figura 4.14):** Nella prima metà del tracciato (dalla finestra 1 alla 5), l'indice ARI si mantiene costantemente sul valore ideale di **1.0**. L'algoritmo mappa lo stato iniziale del network.
 Esattamente alla **finestra 6**, nel momento in cui nel sistema si attiva il nuovo legame che unisce le variabili P, Q, R, S, l'indice ARI mostra un calo scendendo a **0.7324**. Superata questa perturbazione strutturale, l'indice rimbalza immediatamente al valore perfetto di **1.0** dalla finestra 7 fino alla fine della simulazione.
- Andamento (Figura 4.15):** Il grafico dell'ARI che confronta ogni blocco con quello appena precedente mostra un andamento speculare. La curva rimane piatta a **1.0** nei momenti di stabilità, registrando un unico crollo isolato a **0.7324** in corrispondenza del Passo 5. Questo picco negativo è la prova che il tool avverte istantaneamente la mutazione del network rispetto al passo precedente, riasestandosi non appena la situazione torna stabile.

Risultati in presenza di rumore esterno ($\sigma = 0.01$)

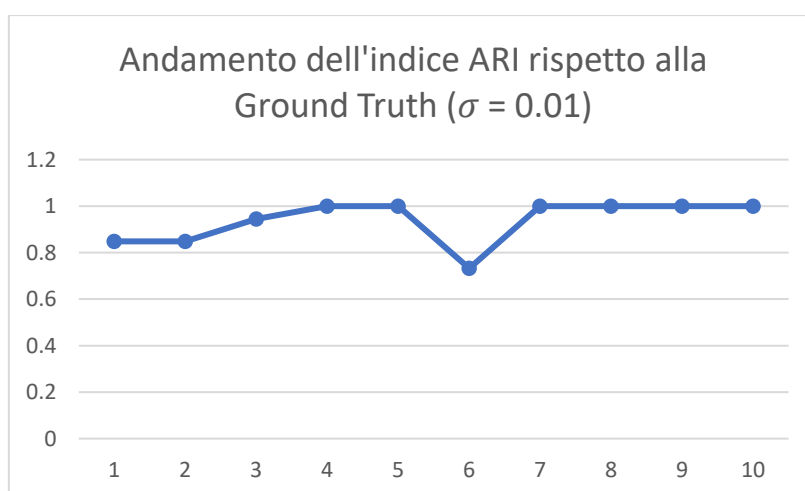


Figura 4.18

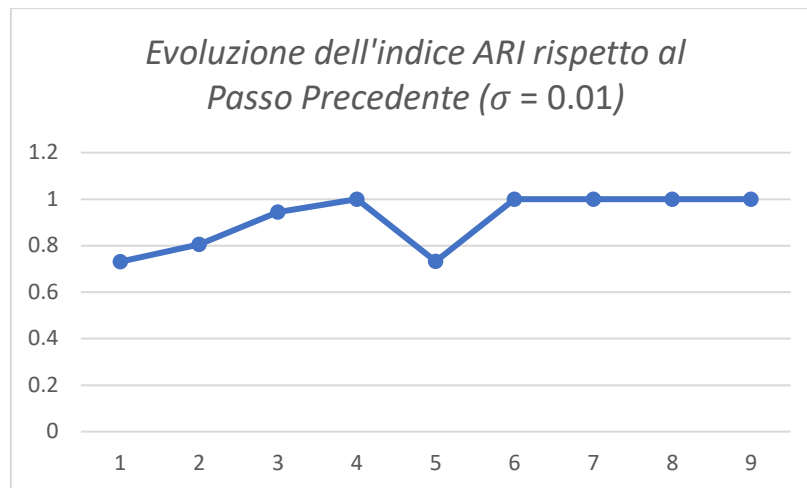


Figura 4.19

- **Accuratezza rispetto alla Verità (Figura 4.16):** Nelle prime finestre di osservazione (da 1 a 3) l'indice ARI non è più perfetto, ma si attesta comunque su livelli ottimi (oscillando tra **0.85** e **0.94**). I valori di fiducia dello Zi in questa fase sono molto bassi (attorno a 3.4). Non appena attraversiamo la transizione della rottura e della nascita dei gruppi, l'indice ARI mostra il suo calo critico alla **finestra 6 (pari a 0.7324)**, per poi risalire verticalmente e bloccarsi sul **valore perfetto di 1.0** stazionario fino al Passo 10. In questa seconda metà del file, anche il valore Zi recupera robustezza, oscillando stabilmente tra 12.8 e 16.7.
- **Andamento (Figura 4.17):** Il grafico dell'ARI inter-finestra (confronto tra il passo corrente i e il precedente $i-1$) traccia una "valle" di instabilità molto chiara. Nei primi tre passi, a causa del disturbo stocastico, l'indice oscilla tra **0.73** e **0.80**. Al **Passo 5** registriamo il crollo principale in cui l'ARI scende a **0.7324**, evidenziando la scomposizione temporale del sistema nel momento esatto dello switch strutturale. Subito dopo la transizione, l'indice si alza e si stabilizza sul valore ideale di **1.0** (dal Passo 6 in poi), dimostrando che il network ha digerito la riorganizzazione e ha ritrovato una stabilità statistica totale.

4.7.3 Conclusioni

I risultati combinati di questa sezione ci permettono di dare una risposta alla domanda di partenza: **quanto sono importanti le attivazioni numeriche accumulate nella finestra rispetto alla semplice posizione (accesso/spento) del nodo?**

1. **Nel mondo ideale (Senza rumore), la posizione basta:** Confrontando i dati delle finestre temporali del modello "solo acceso/spento" con quelli del modello originale "ad attivazioni", notiamo che in assenza di rumore l'andamento è **del tutto invariata**. In entrambi i casi l'ARI viaggia stabile a 1.0, crolla a 0.73 alla sesta finestra e rimbalza a 1.0. Questo significa che se il segnale è pulito, l'informazione geometrica della "posizione" è autosufficiente: l'algoritmo non ha bisogno di sapere *quanto* un neurone stia sparando, ma gli basta sapere *se* si accende insieme agli altri per ricostruire i cluster perfetti.

2. **Nel mondo reale (Con rumore), l'attivazione fa da scudo:** La vera e propria divergenza concettuale si manifesta nel caso rumoroso ($\sigma = 0.01$).
- Nel modello a finestre temporali originale (quello che contava l'intensità cumulativa fino a 5), l'algoritmo usava il valore numerico come un filtro d'area: i nodi che sparavano di più pesavano di più, permettendo all'ARI di rimanere alto (vicino a 0.90 o 1.0) anche in mezzo al rumore.
 - Nel modello attuale binario (0/1), l'algoritmo perde questa protezione. Nelle prime finestre l'accuratezza cala vistosamente fino a **0.85** e lo Zi scende a livelli critici (3.4). Questo accade perché il rumore inverte gli stati dei singoli zeri e uni in modo casuale, e l'algoritmo — non potendo più distinguere uno *spike* casuale isolato da una raffica coerente ad alta intensità — inizia a frammentare i gruppi reali inserendo variabili isolate.
3. **La resilienza post-transizione:** L'aspetto più rassicurante e sorprendente di questo modello semplificato è la sua eccezionale capacità di recupero. Nella seconda metà del tracciato temporale (finestre 7-10), nonostante la perdita del conteggio delle attivazioni e la presenza del rumore, l'indice ARI balza e si congela sul valore perfetto di **1.0**. Questo accade perché quando il macro-gruppo atteso si restringe ed emerge una struttura fortemente coesa come il blocco PQRS, la coincidenza spaziale diventa talmente netta da neutralizzare l'assenza del dato di intensità.

In conclusione, possiamo affermare che **la misura dell'attivazione integrata nel tempo non è strettamente necessaria per intercettare la nascita di nuove faglie o transizioni strutturali, ma è una proprietà indispensabile per conferire stabilità e robustezza statistica all'algoritmo di clustering di fronte al rumore ambientale di fondo**. Senza le attivazioni, lo spazio degli stati diventa fragile e sensibile alle perturbazioni stocastiche; con le attivazioni, il tool acquisisce una capacità di filtraggio che protegge la reale verità dinamica del sistema.

5. Analisi di dati biologici

5.1 Introduzione ai Dati Sperimentali

I dati biologici a disposizione coinvolgono registrazioni dell'attività cerebrale di topi e di scimmie durante alcuni test comportamentali, in cui i soggetti dovevano distinguere diverse situazioni (comparsa di punti colorati in un visore, ed in base al colore ed alla permanenza premere un tasto per avere una ricompensa).

Vi sono diverse sessioni che coinvolgono i topi, ed altre che coinvolgono le scimmie. Il design degli esperimenti e della stanza varia dai topi alle scimmie, e così anche la dimensione dell'apparato volto alla misura dell'attività cerebrale. Lo strumento di misura è in grado di identificare l'attività elettrica di singoli nodi, con una precisione temporale intorno al millisecondo. Per i topi però abbiamo a disposizione serie temporali ognuna contenente diverse centinaia di neuroni, mentre per le scimmie i neuroni a disposizione sono solo poche unità - fino ad una decina. In tutte le traiettorie (topi e scimmie) ogni singolo neurone è tipicamente poco attivo (la maggior parte del tempo in effetti i nodi sono silenti), ragione per cui - come vedremo a breve - abbiamo preferito concentrarci nell'analisi delle serie temporali dei topi, sommando le attività dei neuroni appartenenti alla medesima area cerebrale, in modo da avere segnali più continui. Questo tipo di analisi ci ha permesso quindi di osservare l'attività coordinata a livello di area cerebrale. In particolare, in questa tesi analizzeremo come esemplificazione l'analisi di una sequenza particolare, essendo tutte le altre sequenze strutturalmente simili a quella presentata.

I dati a disposizione derivano da una sessione sperimentale della durata complessiva di 2702.7 secondi. La rete neuronale monitorata è costituita da un totale di 716 nodi attivi. L'informazione originaria registra, per ciascun nodo, l'istante temporale esatto in cui è avvenuto un evento di emissione (sparo), con la precisione del millisecondo.

Al fine di rendere i dati computazionalmente trattabili e di poterne analizzare le proprietà macroscopiche, si è proceduto a un'operazione di integrazione temporale. I vettori di sparo dei singoli nodi sono stati raggruppati all'interno di finestre temporali discrete con un'ampiezza pari a 0.1 secondi. Di conseguenza, la struttura dati finale è un tracciato in cui ogni riga rappresenta un intervallo di un decimo di secondo e ogni colonna esprime l'attività (ovvero il conteggio discreto delle attivazioni) del rispettivo nodo all'interno di quella specifica finestra temporale.

Tempo	224a	172a	241a	225a	256a	259a	211a	208b	214a	206a	245a	196a	222a	210a	235a	196b	192a	228a	221a	235b	248a	212a	231a	228b
Tempo	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA	ACA
0.1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	3	0	0	0	0	0	1	0	0
0.2	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
0.3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.4	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0.5	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.7	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0.8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1.1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1.2	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1.3	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1.6	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
1.7	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0
1.8	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	2	0	0	1
1.9	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0

Figura 5.1 Esempio di porzione del tracciato dei dati sperimentali integrati nell'area ACA. Ogni colonna rappresenta un singolo nodo della rete neuronale (identificato dal proprio codice alfa-numerico), mentre ogni riga corrisponde a un intervallo discreto (bin) di tempo pari a 0.1 secondi. I valori numerici all'interno delle celle esprimono il conteggio delle attivazioni (spari) registrate da quel determinato nodo nella specifica finestra temporale, evidenziando visivamente la prevalenza di intervalli silenti (valore pari a 0, in verde), l'incidenza di singoli eventi isolati (valore pari a 1, in giallo) e la presenza di scariche multiple (valori superiori ad 1). Nella seconda riga è presente il nome dell'area del cervello cui appartiene il neurone.

5.2 Analisi del Trend Temporale delle Attivazioni

Una prima indagine esplorativa ha riguardato il comportamento globale del sistema lungo l'intero arco temporale dell'esperimento.

In primo luogo, si è calcolata la somma totale delle attivazioni registrate dall'intera rete per ogni frazione di secondo (0.1 s), valutandone l'andamento complessivo.

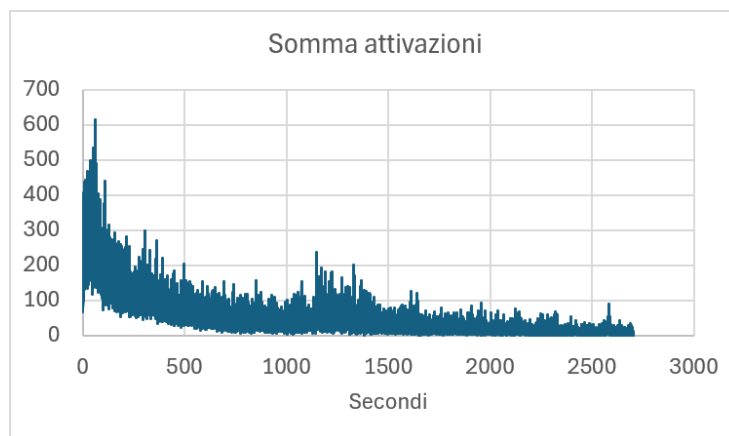


Figura 5.2

Come evidenziato dalla **Figura 5.2**, l'attività complessiva del sistema mostra un trend decrescente nel tempo. Se nei primi secondi dell'esperimento si registrano picchi che superano le 500 o 600 attivazioni complessive per decimo di secondo, procedendo verso la fine della sessione i valori medi si attestano stabilmente al di sotto delle 100 unità.

Per comprendere se tale contrazione fosse dovuta a una diminuzione generalizzata della frequenza di scarica di tutti i neuroni o, viceversa, al progressivo "spegnimento" di una parte di essi, si è analizzato il numero di nodi attivi in ogni istante di tempo. In questo secondo conteggio, l'attività del singolo nodo viene binarizzata: si assegna valore 1 se il nodo ha emesso almeno un impulso nel decimo di secondo considerato (indipendentemente dal fatto che abbia sparato 1, 3 o più volte), e 0 in caso contrario.

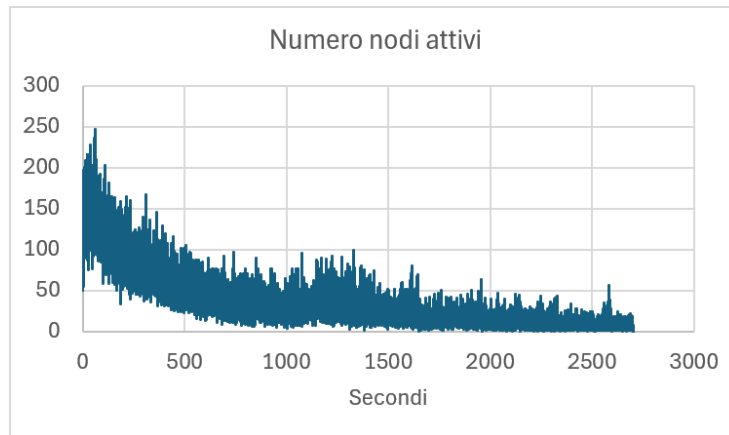


Figura 5.3

Il relativo andamento, illustrato in **Figura 5.3**, ricalca in modo coerente il profilo della somma totale delle attivazioni. Il numero di nodi simultaneamente attivi scende progressivamente da valori prossimi a 200-250 nelle fasi iniziali a valori inferiori a 50 nelle fasi avanzate. Ciò dimostra che il fenomeno osservato non è una semplice riduzione della frequenza di scarica dei singoli elementi, bensì una contrazione del numero di nodi attivamente coinvolti nelle dinamiche di rete.

Per isolare il contributo della frequenza di scarica pura rispetto al numero di neuroni partecipanti, si è calcolato il rapporto tra il numero totale di attivazioni e il numero di nodi attivi in ogni step temporale, calcolandone la media mobile su una finestra di 20 passi per attenuare le fluttuazioni stocastiche.

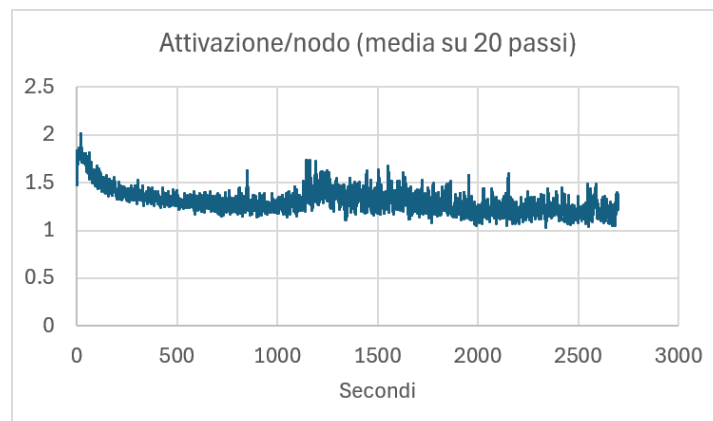


Figura 5.4

Dall'analisi della **Figura 5.4** emergono tre fasi distinte:

1. **Fase Iniziale (0 - 500 secondi):** Si osserva un valore di attivazione per nodo sensibilmente più elevato (prossimo a 2.0). In questa fase i singoli nodi non solo sono numericamente superiori, ma sparano anche con maggiore frequenza.
2. **Fase Intermedia (500 - 1100 secondi):** Il rapporto si stabilizza intorno a un valore basale compreso tra 1.2 e 1.3.

3. **Fase Centrale e Finale (1100 - 2700 secondi):** Intorno al secondo 1100 si registra un temporaneo incremento della frequenza di scarica per nodo, la quale tocca punte di 1.5-1.7, per poi tornare a contrarsi e stabilizzarsi verso la fine dell'esperimento.

Questo comportamento indica che, sebbene la storia globale della rete sia caratterizzata da un progressivo e inesorabile spegnimento (dovuto alla netta diminuzione dei nodi attivi), la frequenza interna dei nodi superstiti subisce una parziale fluttuazione centrale prima del definitivo declino.

5.3 Trattamento del Rumore di Fondo: Analisi "NoUni"

Un elemento critico nell'analisi di dati sperimentali di natura elettrofisiologica è la discriminazione tra i segnali biologici effettivi e il rumore di fondo della strumentazione. Esaminando il tracciato dei dati integrati, è emersa un'elevata incidenza di eventi in cui un nodo registrava un'unica attivazione (1 sparo) all'interno del decimo di secondo, senza alcuna attività nei periodi immediatamente precedenti o successivi. Per verificare l'ipotesi che tali attivazioni isolate rappresentassero esclusivamente rumore casuale, si è applicato un protocollo di filtraggio denominato **"NoUni"**, azzerando sistematicamente tutti i conteggi pari a 1 all'interno del tracciato.

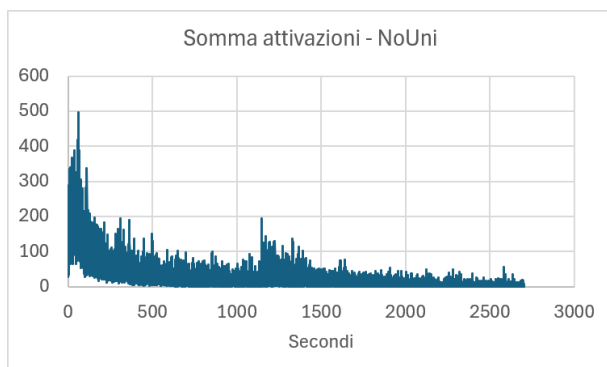


Figura 5.5

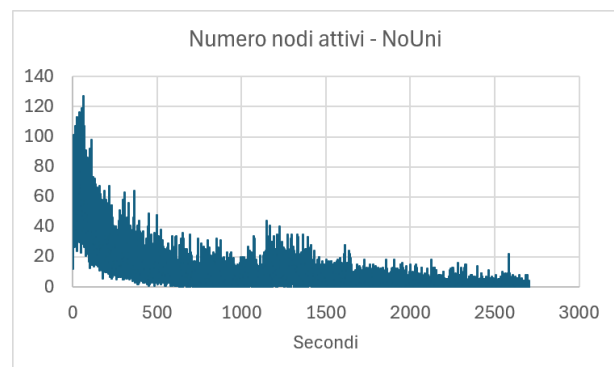


Figura 5.6

Il confronto tra i dati originali e quelli filtrati rivela un impatto significativo sulla scala quantitativa dei grafici, pur mantenendo inalterato l'andamento generale di fondo. Come mostrato in **Figura 5.5**, la rimozione degli "uni" comporta un abbassamento generale della curva della somma delle attivazioni. L'effetto è ancora più evidente nella **Figura 5.6**, relativa al numero di nodi attivi nello stesso istante: i picchi iniziali subiscono una contrazione drastica, passando da circa 250 nodi attivi nel dataset grezzo a circa 120 nel dataset filtrato. Tale discrepanza numerica indica che una quota rilevante (vicina al 50%) dei nodi ritenuti inizialmente attivi fosse in realtà guidata da fluttuazioni stocastiche assimilabili a rumore strumentale o a scariche isolate non funzionali.

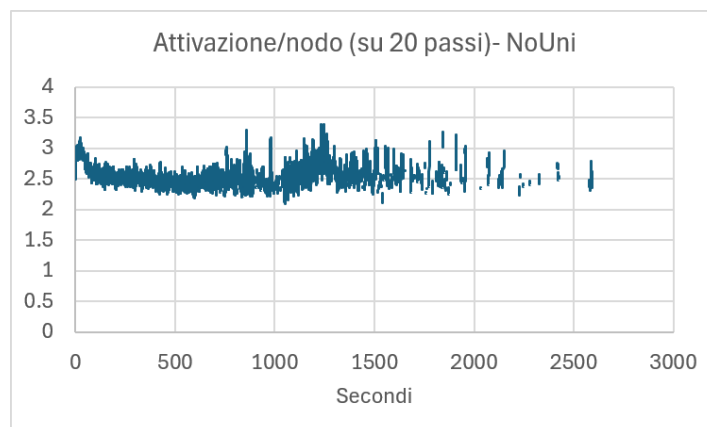


Figura 5.7

L'efficacia del filtro trova un'ulteriore conferma nella **Figura 5.7**, che mostra il rapporto di attivazione per nodo calcolato sul dataset "NoUni". Mentre nel caso originario il valore oscillava attorno a 1.5, l'eliminazione dei singoli spari sposta la media stabilmente sopra la soglia di 2.5 attivazioni per nodo. Questo incremento dimostra che, una volta filtrato il sistema dal rumore di fondo dei nodi debolmente attivi (monosparo), i neuroni effettivamente coinvolti nella rete operano a frequenze decisamente più sostenute e biologicamente significative.

5.4 Analisi delle Frequenze di Sparo e Vincoli Biologici

Per validare la coerenza dei dati rispetto alla fisiologia neuronale, i risultati sperimentali sono stati messi in relazione con le evidenze biologiche fornite in letteratura, le quali indicano per questo tipo un tempo di latenza (periodo refrattario) compreso tra i 10 e i 15 millisecondi.

Considerando la nostra finestra di integrazione temporale pari a 0.1 secondi (100 ms), un tempo di latenza di 15 ms impone un limite teorico superiore al numero di scariche consecutive eseguibili da un singolo neurone, pari a circa:

$$\frac{100 \text{ ms}}{15 \text{ ms}} \approx 6.67 \text{ spari}$$

Arrotondando tale valore per eccesso, si ottiene una soglia critica pari a **7 spari per decimo di secondo**. Eventi con frequenze superiori a questo limite indicherebbero una violazione del tempo di refrattarietà biologica, sollevando anomalie interpretative o suggerendo la presenza di artefatti sperimentali o di specifiche sottopopolazioni cellulari (ad esempio, neuroni a scarica rapida o *fast-spiking*).

Al fine di quantificare la distribuzione di tali eventi, sono stati elaborati gli istogrammi delle frequenze complessive di sparo sia in scala lineare che logaritmica.

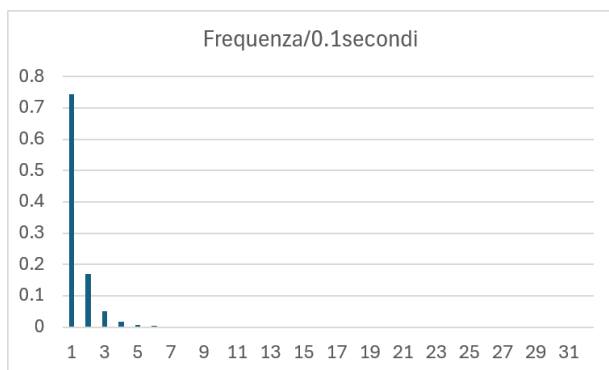


Figura 5.8

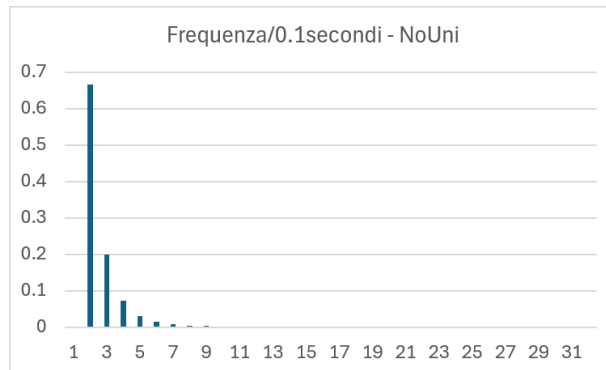


Figura 5.9

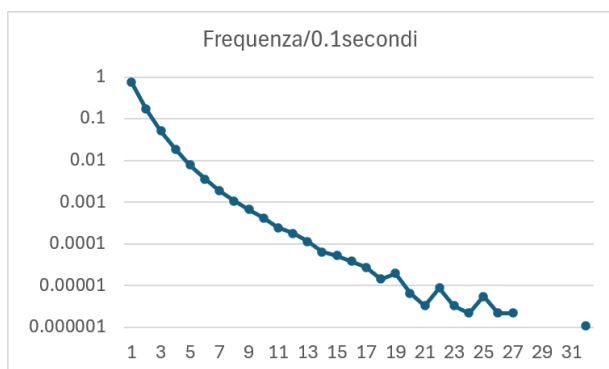


Figura 5.10

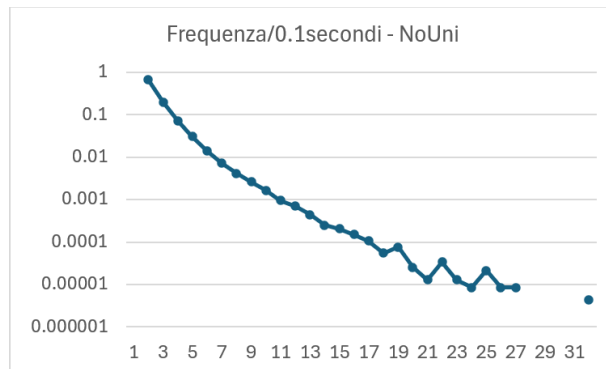


Figura 5.11

Dall'analisi degli istogrammi lineari (**Figure 5.8 e 5.9**) si osserva che la stragrande maggioranza degli eventi si concentra su valori di frequenza molto bassi. Nel dataset completo, circa il 75% delle attivazioni totali del sistema è rappresentato da singoli spari (valore pari a 1), mentre il 18% circa presenta un conteggio pari a 2. Il passaggio alla scala logaritmica (**Figure 5.10 e 5.11**) permette di analizzare nel dettaglio la coda della distribuzione: all'interno della rete, la frazione di nodi che registra una frequenza superiore a 7 spari per decimo di secondo è inferiore allo 0.1%.

Per approfondire la natura di queste scariche ad alta frequenza, si è condotta un'analisi mirata a livello dei singoli nodi, contando quante volte ciascun neurone superasse il limite di 7 spari nel corso dell'intero esperimento.

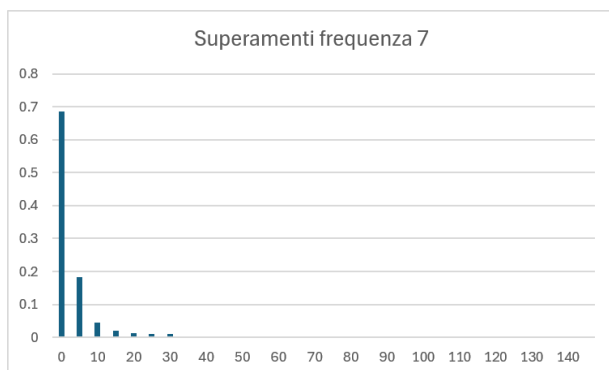


Figura 5.12

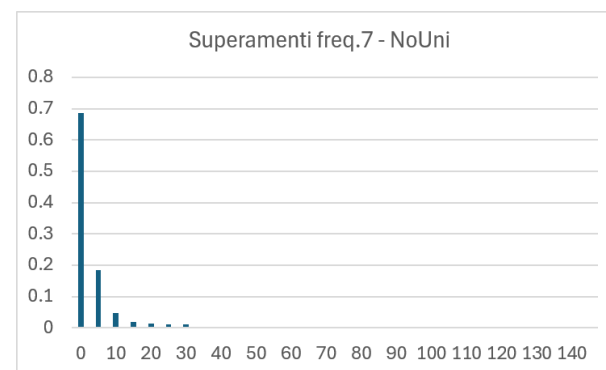


Figura 5.13

I risultati, illustrati nelle **Figure 5.12 e 5.13**, rivelano un quadro estremamente chiaro sulla distribuzione di questa iperattività:

- **Circa il 70% dei nodi** non supera mai, in nessun istante dell'esperimento, la soglia dei 7 spari. Per la grande maggioranza della rete, quindi, il vincolo biologico dei 15 millisecondi viene rigorosamente rispettato.
- **Circa il 18% dei nodi** mostra un superamento sporadico, registrando una frequenza superiore a 7 solo per un numero limitato di intervalli (da 1 a 5 volte in tutta la sessione).
- **Il 3%-4% dei nodi** si spinge leggermente oltre, superando la soglia tra le 5 e le 10 volte.
- **Una frazione piccolissima (appena 15 nodi su oltre 700)** manifesta invece un comportamento iperattivo sistematico, superando molto frequentemente la frequenza critica.

All'estremo di questa distribuzione troviamo un singolo nodo fortemente attivo in cui la scala dei superamenti arriva fino a **140 volte**. La persistenza e la regolarità con cui questi pochissimi nodi violano la soglia escludono l'ipotesi che si tratti di fluttuazioni casuali o rumore strumentale isolato; al contrario, è altamente probabile che questa ristretta sottopopolazione rappresenti neuroni con caratteristiche fisiologiche peculiari (come i già citati interneuroni *fast-spiking*) o che sia soggetta a dinamiche di stimolazione locali molto intense.

Un'osservazione metodologica fondamentale emerge dal confronto tra il dataset originale e quello depurato tramite il filtro "NoUni": **la distribuzione dei superamenti della soglia rimane identica**. L'esclusione dei conteggi pari a 1 non altera in alcun modo né la quota di nodi che rispettano la soglia (che resta ancorata al 70%), né il comportamento della coda dei nodi iperattivi.

Questo fenomeno costituisce un indizio statistico interessante: se gli "uni" avessero riflesso una reale dinamica informativa o funzionale interconnessa alla rete, la loro rimozione avrebbe modificato la struttura delle distribuzioni di sparo. Il fatto che l'architettura dei superamenti rimanga del tutto invariata conferma il forte sospetto iniziale che i singoli spari isolati rappresentino esclusivamente rumore di fondo additivo e scorrelato.

5.5 Anomalie Sperimentali e Dinamiche di Spegnimento

Un'osservazione rigorosa dei dati grezzi ha messo in luce alcune anomalie macroscopiche di natura prettamente sperimentale, legate alla stabilità della strumentazione di misura più che a reali fenomeni biologici.

La prima anomalia riguarda i segmenti iniziali della sessione di registrazione. Esaminando il tracciato nei primi 10 secondi, si osserva la presenza di repentine "partenze improvvise".

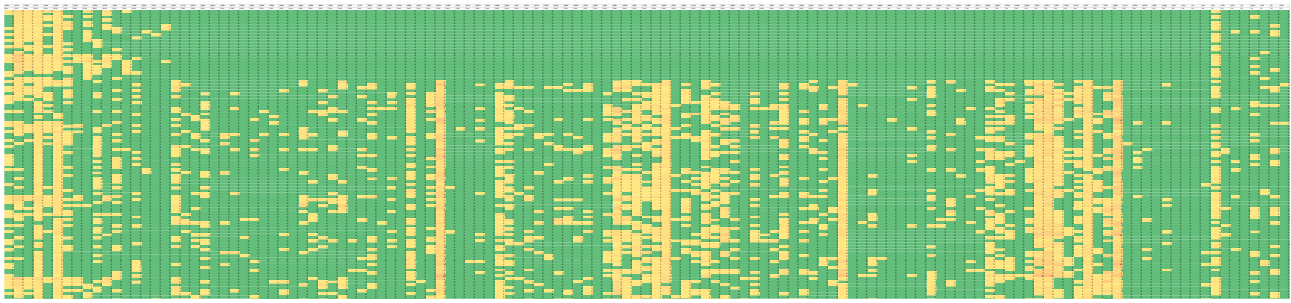


Figura 5.14 **Tracciato originale**: L'andamento evidenzia il fenomeno delle "partenze improvvise", in cui intere popolazioni di nodi distribuiti in diverse zone della matrice si attivano simultaneamente nello stesso esatto istante temporale. Questo comportamento solleva il forte sospetto di un artefatto strumentale legato all'accensione tardiva di alcuni specifici sensori dell'apparato di misura.

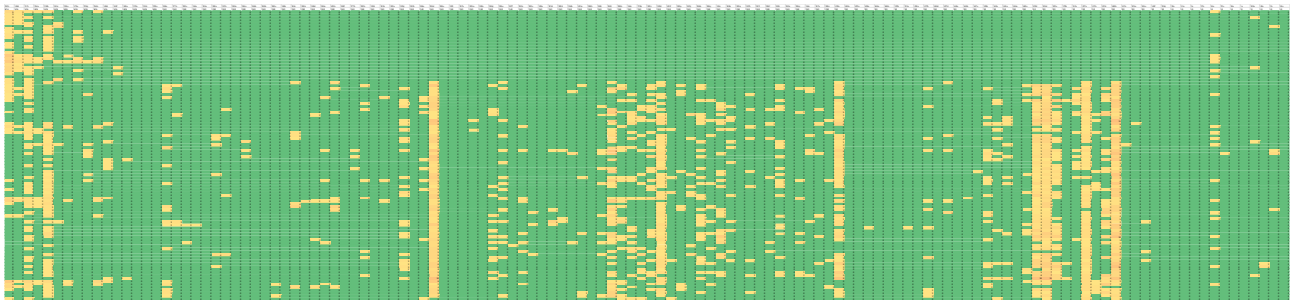


Figura 5.15 **Tracciato dopo l'applicazione del filtro "NoUni"**: Nonostante il filtraggio e l'azzeramento sistematico di tutte le attivazioni isolate (pari a 1 sparo per bin), la struttura a bande verticali delle attivazioni simultanee e sincrone rimane chiaramente identificabile. Tale persistenza convalida la scelta metodologica di scartare i primi 3 secondi dell'esperimento dall'analisi complessiva, al fine di escludere la transizione iniziale e lavorare esclusivamente a regime strumentale stabilizzato.

Come mostrato in **Figura 5.14**, vi sono interi gruppi di nodi che iniziano a registrare attività simultaneamente nello stesso identico istante. Da un punto di vista neurofisiologico, l'attivazione sincrona e immediata di una popolazione così vasta ed eterogenea risulta altamente improbabile. Il sospetto è che alcuni sensori dell'apparato sperimentale fossero inizialmente spenti o in fase di calibrazione e si siano attivati di colpo. Poiché tale finestra di accensione anomala permane inalterata anche dopo l'applicazione del filtro "NoUni" (**Figura 5.15**), si è deciso di **scartare i primi 3 secondi di analisi** dall'elaborazione complessiva dei dati, garantendo che l'indagine parta da una condizione di regime strumentale stabile.

Una seconda e cruciale evidenza riguarda il comportamento a lungo termine dei nodi. Per mappare il progressivo spegnimento della rete, si è calcolata per ciascun nodo la distanza temporale intercorsa tra il suo ultimo sparo e la fine dell'esperimento (al tempo 2702.7).

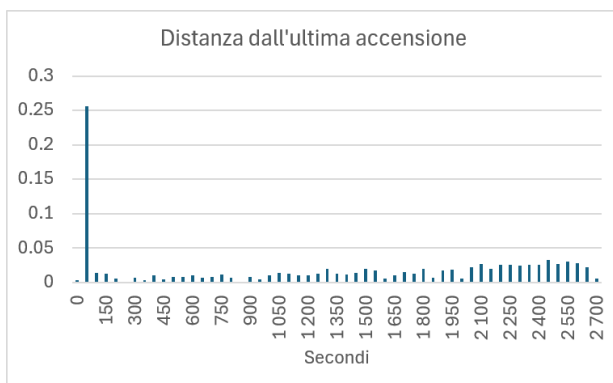


Figura 5.16

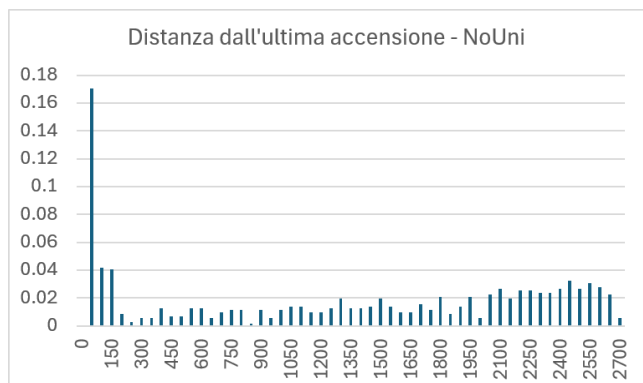


Figura 5.17

L'istogramma a passi di 50 secondi (**Figure 5.16 e 5.17**) evidenzia che circa il 50% dei nodi totali ha registrato la propria ultima attivazione nella prima metà della sequenza, rimanendo completamente silente per gli ultimi 1400 secondi. Questo dato è coerente con la progressiva riduzione del numero di nodi attivi analizzata nel paragrafo 5.2. Dal punto di vista statistico, il passaggio al dataset "NoUni" riduce la frazione di nodi che risultano essersi spenti proprio negli ultimi secondi (la prima colonna cala dal 25% al 17%), ma lascia invariato il profilo del restante 49% dei nodi spenti precocemente. Ciò ribadisce che la struttura macroscopica del fenomeno di spegnimento non viene alterata dal filtro, confermando l'ipotesi del rumore isolato.

L'analisi del tempo totale di attività di ogni singolo nodo (calcolato come intervallo tra la primissima e l'ultimissima accensione registrata) offre un'ulteriore conferma.

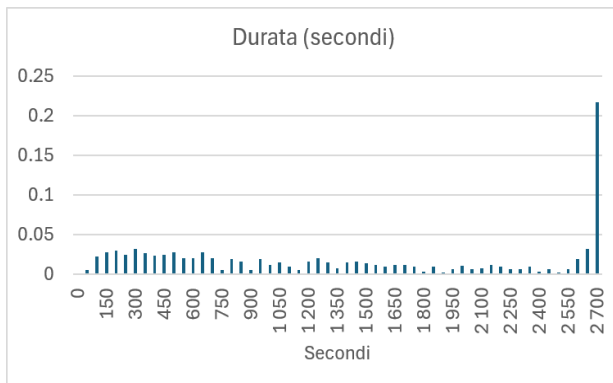


Figura 5.18

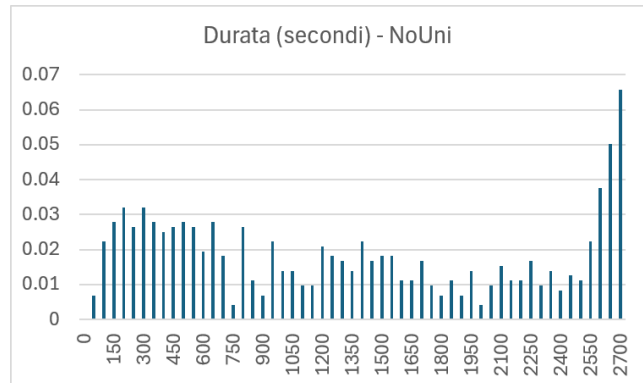


Figura 5.19

I relativi grafici (**Figure 5.18 e 5.19**) mostrano che nel dataset originale solo il 22% dei nodi copre l'intera durata dell'esperimento, mentre il 51% mostra una finestra di attività inferiore alla metà del tempo totale. Nel dataset "NoUni", a causa della rimozione degli spari isolati, la quota di nodi attivi per tutta la sequenza crolla appena al 6%. L'invarianza geometrica del resto della distribuzione convalida nuovamente l'accuratezza del filtraggio.

Infine, si è studiato l'andamento cumulativo dei nodi definitivamente spenti al passare del tempo.

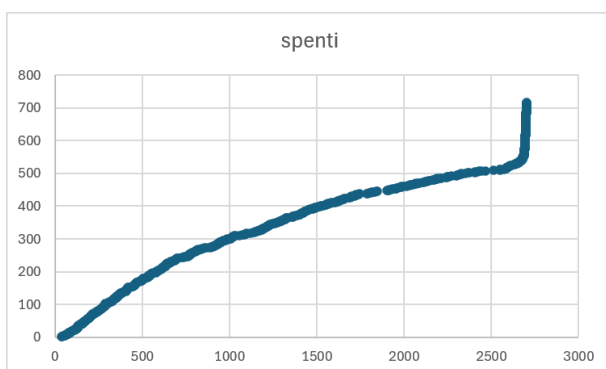


Figura 5.20

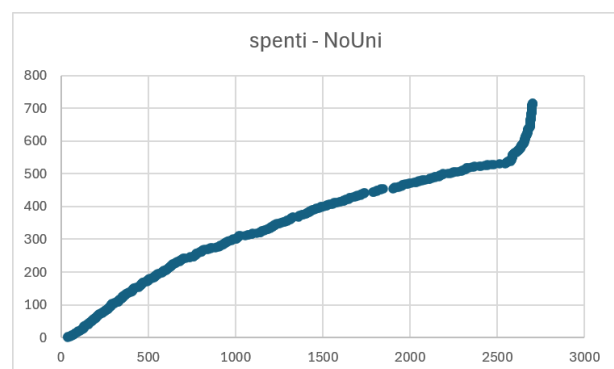


Figura 5.21

Le **Figure 5.20 e 5.21** mostrano un trend di spegnimento estremamente regolare e lineare per gran parte dell'esperimento. Tuttavia, negli ultimissimi secondi si osserva una netta e repentina impen-nata della curva. Questa anomalia finale, della durata di circa 20-30 secondi probabilmente non corrisponde a una dinamica biologica, bensì all'interruzione del sistema di misurazione e allo spegni-mento dell'apparato elettronico. Per evitare che questo artefatto influenzi i modelli matematici suc-

cessivi, si è scelto di **scartare la parte finale dell'esperimento**, fissando il termine della sequenza utile al secondo 2600- in modo da avere un totale di 2600 secondi di osservazione.

5.6 Suddivisione in Aree Cerebrali e Analisi dei Tracciati

Date le dimensioni della rete (716 nodi complessivi), uno studio condotto livello per livello sui singoli elementi risulterebbe cinematicamente ridondante e di difficile interpretazione collettiva. Si è pertanto sfruttata la classificazione anatomica strutturale basata sull'**Allen Mouse Brain Reference Atlas**, raggruppando i nodi all'interno di 7 macro-aree cerebrali distinte:

- **ACA (Anterior Cingulate Area - 109 nodi):** Regione corticale associata all'elaborazione emotiva, ai processi di apprendimento e alla memoria.
- **CA3 (Field CA3 - 68 nodi):** Sotto area dell'ippocampo fondamentale per la memoria a breve termine e il consolidamento spaziale.
- **DG (Dentate Gyrus - 34 nodi):** Porzione dell'ippocampo deputata alla neurogenesi e alla ricezione di input corticali.
- **LS (Lateral Septal complex - 139 nodi):** Complesso sottocorticale implicato nella regolazione emotiva, risposte allo stress e sistemi di ricompensa.
- **MOs (Somatomotor areas, supplementary - 113 nodi):** Aree motorie supplementari deputate alla pianificazione del movimento volontario.
- **SUB (Subiculum - 75 nodi):** Struttura di transizione ippocampale che funge da principale via di output per i segnali ippocampali.
- **VISp (Primary Visual Area - 178 nodi):** Corteccia visiva primaria, responsabile della prima elaborazione degli stimoli visivi.

ACA	CA3	DG	LS	MOs	SUB	VISp
109	68	34	139	113	75	178

Figura 5.22 Numero nodi nelle macroaree cerebrali, nell'esperimento commentato in questa tesi.

Per comprendere come queste aree interagiscano e se il fenomeno di spegnimento sia selettivo o generalizzato, si è tracciata l'attività temporale media di ciascuna area (calcolata come somma delle attivazioni dei nodi interni divisa per il numero di nodi dell'area stessa).

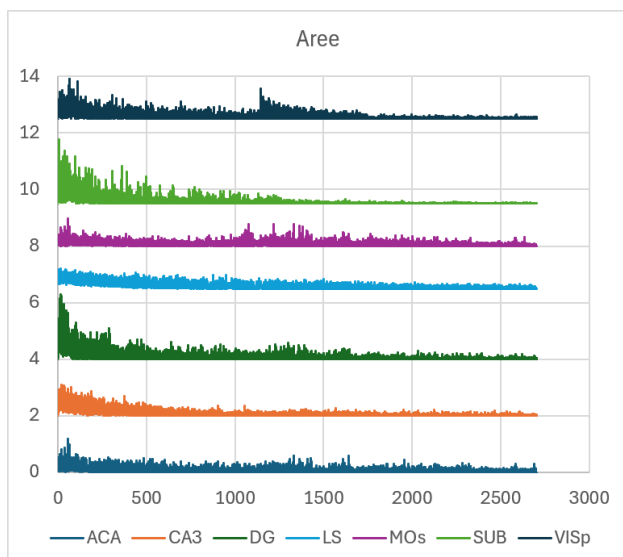


Figura 5.23

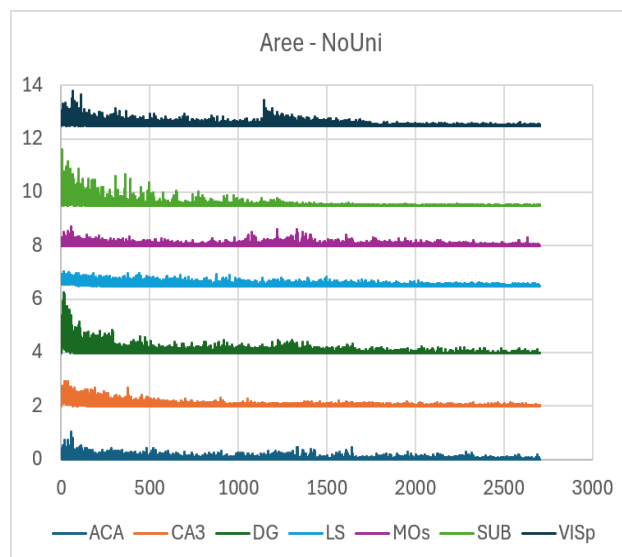


Figura 5.24

Il confronto tra la rete originaria (**Figura 5.23**) e quella filtrata (**Figura 5.24**) mostra variazioni minime nei profili generali, confermando l'opportunità di concentrare le analisi esclusivamente sul dataset "NoUni". I tracciati evidenziano che **il fenomeno della contrazione temporale dell'attività si verifica in modo sistematico in tutte e 7 le aree cerebrali**. Ciascuna macroarea, pur partendo da livelli basali di attivazione differenti (con la corteccia visiva VISp e l'area motoria MOs che mostrano la dinamica più energetica), segue un destino comune di progressivo spegnimento lungo l'asse temporale.

5.7 Discretizzazione dell'Attività: Approccio 1 vs Approccio 2

Per poter studiare le interazioni funzionali e le transizioni di stato della rete tramite indici statistici di coerenza, si rende utile utilizzare una ripartizione in pochi livelli di attivazione (basso, medio, alto), anziché utilizzare l'alto numero di differenti valori disponibili, di per sé poco significativi. Sulla base delle anomalie strumentali individuate nel paragrafo 5.5, la sequenza temporale complessiva è stata preventivamente ridotta, isolando l'intervallo stabile compreso **dal secondo 3 al secondo 2603** dell'esperimento (per una durata totale di 2600 secondi). Questo intervallo è stato suddiviso in **43 gruppi temporali da 60 secondi ciascuno** (estendendo l'analisi effettiva fino al secondo 2580).

Le variabili analizzate sono costituite dalla somma delle attivazioni delle 7 macroaree cerebrali individuate. Poiché la distribuzione dei dati di sparo non segue un andamento gaussiano ma si presenta fortemente asimmetrica, la classificazione dell'attività è stata strutturata su **tre livelli discreti (bassa, media, alta)** tramite l'applicazione di soglie, calibrate in modo che circa il 70% delle osservazioni ricada nel livello basso, il 10-15% nel livello medio e la restante parte nel livello alto.

Per capire come il progressivo spegnimento della rete possa influenzare l'individuazione degli stati funzionali, sono stati sviluppati e messi a confronto due differenti percorsi metodologici:

Approccio 1: Soglie Assolute sui Dati Filtrati ("NoUni")

Il primo metodo applica delle soglie di ampiezza fisse direttamente sulla somma delle attivazioni del dataset depurato dai singoli spari.

	ACA	CA3	DG	LS	MOs	SUB	VISp
SGL1	3	3	2	5	5	2	7
SGL2	6	6	3	12	7	6	18

Figura 5.25 Valori di Soglia (SGL1, SGL2) - Approccio 1

L'applicazione di queste soglie rigide genera la seguente distribuzione delle osservazioni nei tre livelli di attività:

	ACA	CA3	DG	LS	MOs	SUB	VISp
Basso	0.73	0.76	0.79	0.75	0.78	0.71	0.70
Medio	0.15	0.11	0.11	0.14	0.10	0.18	0.19
Alto	0.12	0.13	0.10	0.11	0.12	0.11	0.11

Figura 5.26 Frazione delle osservazioni nei tre livelli - Approccio 1

Approccio 2: Normalizzazione Temporale e Soglie

Poiché l'attività complessiva del sistema cala drasticamente nel tempo, l'utilizzo di soglie assolute (Approccio 1) rischia di falsare l'analisi nelle fasi avanzate dell'esperimento, classificando l'attività finale quasi esclusivamente come "bassa" a causa del decremento globale e non per un'effettiva variazione nei rapporti interni tra le aree.

Per ovviare a questo limite, il secondo approccio introduce una **fase di normalizzazione preliminare**: i valori di attivazione di ogni singola area vengono divisi per la somma totale delle attività del sistema in quel momento (calcolata come media mobile su 20 osservazioni). Questo processo compensa efficacemente il progressivo spegnimento del sistema, portando i tracciati a oscillare attorno a un valore medio costante lungo l'intero asse temporale.

Sui tracciati normalizzati e resi stabili vengono quindi calcolate e applicate le nuove soglie relative:

	ACA	CA3	DG	LS	MOs	SUB	VISp
SGL1	0.1	0.08	0.0001	0.16	0.22	0.03	0.35
SGL2	0.3	0.2	0.15	0.36	0.5	0.15	0.6

Figura 5.27 Valori di Soglia (SGL1, SGL2) - Approccio 2

La conseguente ripartizione delle osservazioni nei tre livelli mostra un bilanciamento simile al precedente, confermando la coerenza statistica dei criteri di soglia scelti:

	ACA	CA3	DG	LS	MOs	SUB	VISp
Basso	0.68	0.70	0.79	0.70	0.71	0.71	0.71
Medio	0.17	0.16	0.13	0.17	0.15	0.13	0.13
Alto	0.15	0.14	0.08	0.14	0.15	0.16	0.16

Figura 5.28 Frazione delle osservazioni nei tre livelli - Approccio 2

5.8 Studio della Stabilità dei Raggruppamenti tramite Indice ARI

Per valutare la consistenza e l'evoluzione temporale delle interazioni funzionali tra le 7 aree cerebrali, si è calcolato l'**Indice ARI (Adjusted Rand Index)** tra ciascun raggruppamento temporale e il precedente lungo la sequenza dei 43 gruppi. L'indice ARI misura il grado di somiglianza e di stabilità tra partizioni di dati: un valore prossimo a 1 indica una forte costanza strutturale nel tempo, mentre valori vicini allo 0 o negativi indicano una riconfigurazione disordinata dei gruppi.

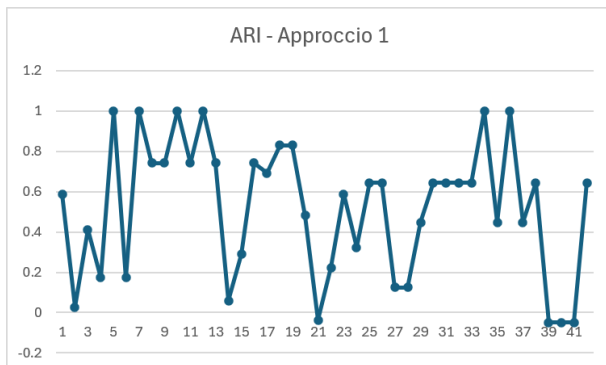


Figura 5.29

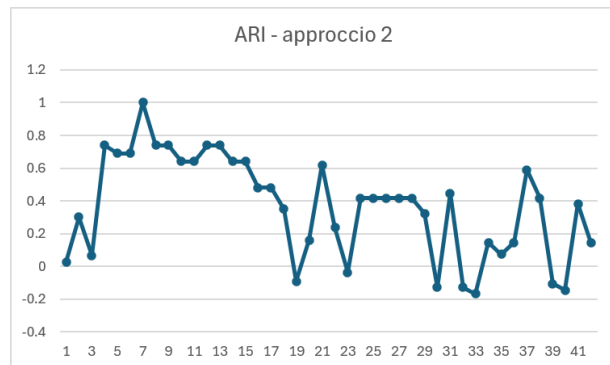


Figura 5.30

Il confronto tra i profili temporali dell'indice ARI ottenuti con il primo approccio (**Figura 5.29**) e con il secondo approccio (**Figura 5.30**) mostra, in modo estremamente interessante, una sostanziale compatibilità macroscopica. L'evoluzione della rete neuronale può essere linearmente suddivisa in **quattro macro-fasi distinte**, identificate con precisione da entrambi i metodi:

1. **Fase Iniziale Disordinata (primi 5-6 minuti):** Entrambi gli approcci registrano forti oscillazioni dell'indice ARI (con picchi verso lo zero), evidenziando una fase di instabilità in cui le aree cerebrali non hanno ancora strutturato aggregazioni funzionali di lunga durata.
2. **Fase di Massima Stabilità Strutturale:** Segue un lungo intervallo in cui i raggruppamenti rimangono costanti. Questa stabilità si protrae **fino al minuto 14** per l'Approccio 1 e si estende **fino al minuto 18** per l'Approccio 2 (il quale, detraendo la tendenza allo spegnimento, riesce a isolare la costanza funzionale più a lungo).
3. **Fase di Transizione e Cambiamento:** Si osserva una marcata caduta dei valori ARI, sintomo di una profonda riorganizzazione interna della rete. Tale finestra di cambiamento si colloca **tra i minuti 14 e 21** per il primo approccio, e **tra i minuti 18 e 23** per il secondo.
4. **Fase Finale:** Negli ultimi minuti dell'osservazione utile, il sistema si attesta su una condizione identificata come debolmente stabile nel caso del primo approccio (influenzato dal basso livello di attivazione generale) o disordinata nel secondo approccio.

Un'indagine più dettagliata a livello delle singole componenti permette di decodificare l'origine biologica e geometrica di queste fluttuazioni dell'indice ARI, rivelando quali specifiche aree guidino la stabilità o il disordine del sistema nelle diverse fasi:



Figura 5.31 Aggregazioni delle aree - Approccio 1



Figura 5.32 Aggregazioni delle aree - Approccio 2

I grafici di dettaglio (**Figure 5.31 e 5.32**) rendono meglio comprensibili le dinamiche che generano il profilo dell'indice ARI, mostrando una netta evoluzione bipartita della rete:

- **Dinamica della Prima Parte:** Nella fase stabile iniziale, la rete è dominata da un nucleo funzionale compreso da un **gruppo di 3 aree fortemente coeso e stabile**, composto dalle regioni corticali **ACA, MOs e VISp**. In parallelo, si rileva un'aggregazione secondaria a composizione più debole e fluttuante, guidata stabilmente dalla coppia **DG ed LS**, a cui si associano in modo intermittente le aree SUB o CA3.
- **Dinamica della Seconda Parte (Post-Transizione):** Nella seconda metà dell'esperimento, in concomitanza con il progredire dello spegnimento, il macro-gruppo da 3 aree si dissolve. La corteccia visiva primaria (VISp) interrompe quasi completamente le proprie interazioni, lasciando come unico nucleo corticale superstite e stabile la **coppia (ACA, MOs)**. Specularmente, l'aggregazione ippocampale e sottocorticale si consolida in un trio compatto formato da **CA3, DG ed LS**, escludendo progressivamente il subicolo (SUB) dalle proprie dinamiche.

Le fluttuazioni e i cali relativi dell'indice ARI (visibili nei grafici 5.30 e 5.31) derivano proprio dalla compresenza di queste componenti: da un lato vi sono strutture identificate in modo rigidamente stabile (il trio corticale nella prima parte e la coppia corticale nella seconda), dall'altro vi sono moduli ippocampali intrinsecamente più variabili e fluidi nelle loro interconnessioni.

Rimarchevolmente, l'identificazione passo per passo di queste aggregazioni composite coincide in modo quasi perfetto tra i due approcci, a testimonianza della robustezza statistica dell'analisi. L'unica differenza significativa si registra proprio nella fase finale: nell'Approccio 2, grazie alla normalizzazione che preserva le proporzioni relative anche a bassi regimi di sparo, emerge una parziale e tardiva interazione residua tra le aree periferiche. In questa coda finale, il subicolo (SUB) torna a dialogare debolmente con il blocco ippocampale (CA3, DG, LS) e la corteccia visiva (VISp) mostra un parziale riavvicinamento alla coppia (ACA, MOs).

In conclusione, l'andamento dell'indice ARI evidenzia che il progressivo spegnimento della rete non dà luogo a un collasso caotico e casuale. Al contrario, si osserva una transizione ordinata e coerente, in cui le grandi squadre di aree cerebrali che lavoravano insieme all'inizio si dividono, lasciando spazio a gruppi più piccoli, isolati e indipendenti.

Conclusione

Il presente lavoro di tesi ha affrontato il problema dell'identificazione e della tracciabilità delle unità funzionali (*cell assemblies*) all'interno di reti neurali, proponendo un percorso organico che ha compreso l'approccio al tema, l'individuazione di un framework in grado di estrarre conoscenza dai dati, lo sviluppo di modelli – uno dedicato alla comprensione del sistema in esame ed un altro alla validazione del tool statistico – ed infine l'applicazione delle conoscenze così acquisite ad un caso sperimentale rilevante.

Per mappare le strutture interessanti (i *cell assemblies*) è stato utilizzato un framework basato sullo z-score dell'integrazione e su un algoritmo iterativo di *sieving*, capace di individuare gruppi ampi e strutture gerarchiche (capitolo 2).

I test condotti sui modelli di rete artificiale hanno mostrato come tool basati su misure di correlazione – come atteso – non permettono una ricostruzione puramente causale delle connessioni, ma introducono gruppi fondati su attivazioni sincronizzate. La discriminazione di legami causali necessita di altri tipi di misura (come ad esempio la Transfer Entropy) - approcci che però a loro volta possono essere indotti in errore dai tempi di ritardo presenti nel sistema in esame. L'introduzione di una finestra di integrazione può ridurre la dipendenza dai ritardi (senza eliminarla del tutto), spostando il problema dall'individuazione di fenomeni di emissione all'individuazione di fenomeni di frequenza di emissione (strategia utilizzata in questo lavoro). Il rumore dinamico può essere una fonte interessante di informazione, attivando schemi di reazione altrimenti non osservabili; il rumore strumentale (errori nella misura degli spike) invece riduce l'osservabilità dei gruppi dinamici (capitolo 3).

La validazione del tool basato sullo Zi in traiettorie con dinamiche *Leader-Follower* ha mostrato la precisione e resilienza dello strumento nell'isolare la *ground truth* strutturale anche durante profonde transizioni e instabilità dinamiche (migrazioni di nodi, collassi o emersione di gruppi – capitolo 4).

Infine, nel capitolo 5 abbiamo applicato il framework a registrazioni ad alta densità dell'attività neurale di topi durante compiti di discriminazione. L'introduzione di una procedura di integrazione temporale in finestre discrete da 0.1 secondi ha permesso di trasformare i singoli eventi in segnali connessi alla frequenza di emissione. L'indagine esplorativa ha rivelato una chiara tendenza temporale decrescente dell'attività globale del sistema, dimostrando che la contrazione del segnale lungo gli oltre 2600 secondi di sessione non è legata a una semplice riduzione della frequenza di scarica, bensì al progressivo "spegnimento" e riduzione numerica dei nodi attivamente coinvolti nelle dinamiche di rete, con gruppi caratterizzati da fluttuazioni transitorie nella seconda parte delle osservazioni.

Appendice A

Topologia circolare

L'output della simulazione mostra la traiettoria temporale dei 5 nodi. Al passo temporale iniziale ($t = 0$), i nodi N000, N001 e N003 sono attivi (valore 1), mentre N002 e N004 sono spenti (0).

Tuttavia, l'evoluzione temporale mostra che **la rete si spegne immediatamente dopo il primo passo**, rimanendo fissa a 0 per tutti i successivi time-step. Questo fenomeno è la prova empirica di quanto discusso precedentemente: a causa della forte presenza di pesi sinaptici negativi (connessioni inibitorie) e in assenza di una sollecitazione costante o di rumore di fondo, la rete decade istantaneamente in un attrattore statico di stop (stato zero).

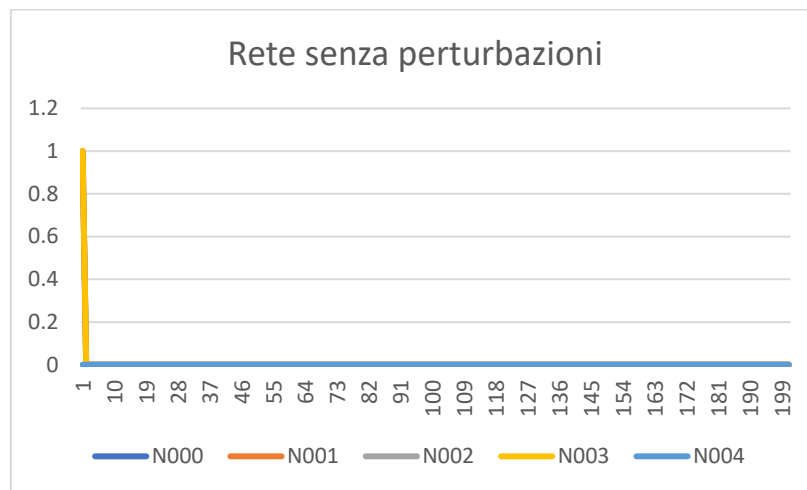


Figura A.1

Nonostante lo spegnimento immediato, l'algoritmo di analisi statistica applicato tramite lo script `ri_coordinaN.py` è programmato per cercare andamenti coordinati in tutto l'intervallo di osservazione. L'output ha generato il seguente risultato:

8.20358871	2	_N000_N001_	N003
------------	---	-------------	------

Figura A.2

Questo ci dice che l'algoritmo ha rilevato una forte significatività statistica (Z-score elevato, pari a **8.20**) nell'attivazione simultanea del gruppo formato da **N000 e N001 insieme a N003** all'interno di una finestra molto stretta (ampiezza 2). Questo evento coordinato viene registrato proprio al primo passo ($t = 0$), dove i tre nodi si trovano accesi nello stesso momento.

Modifica delle probabilità e inserimento del rumore

Per evitare che la rete cadesse immediatamente nell'attrattore statico di zero, la fase successiva dell'esperimento ha previsto la modifica manuale della **probabilità di attivazione spontanea (rumore)** su tutti i nodi.

L'obiettivo di questa variazione è stato osservare se l'introduzione di impulsi casuali potesse "riaccendere" l'anello circolare, permettendo al segnale di propagarsi lungo la catena di nodi nonostante i pesi inibitori, alterando così i valori di coordinazione estratti dallo script di analisi.

1. Perturbando il nodo 0 con una probabilità pari a 0.2: a differenza del caso precedente in cui la rete si spegneva totalmente dopo il primo time-step, l'introduzione del rumore sul Nodo 0 permette a quest'ultimo di "riaccendersi" in modo stocastico lungo tutto l'arco della simulazione (si osservano impulsi isolati o consecutivi, come ai passi 3, 4, 5, 6, 12, ecc.).

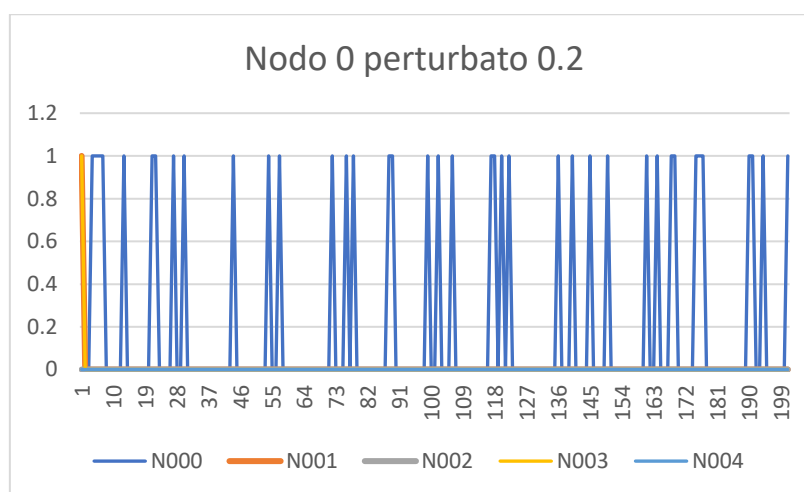


Figura A.3

Un dato estremamente rilevante è che, nonostante il Nodo 0 si attivi ripetutamente grazie alla perturbazione, **tutti gli altri nodi (N001, N002, N003, N004) rimangono fissi a 0**. Questo accade perché le forti connessioni inibitorie della struttura ad anello (in particolare il legame distruttivo da Nodo 0 a Nodo 1 con peso -0.89) bloccano sul nascere la propagazione dell'impulso. Il rumore riesce quindi a mantenere in vita l'attività locale del nodo perturbato, ma non è sufficiente a propagarsi e a riaccendere l'intera dinamica collettiva della rete circolare.

Per quanto riguarda lo Zi:

8.20358871	2	_N001_N003_	N000
------------	---	-------------	------

Figura A.4

Il valore statistico dell'indice Zi rimane identico (**8.2035**), ma **la struttura della coordinazione si è invertita**. Nel caso deterministico originario, l'algoritmo individuava la coppia **_N000_N001_** legata al target N003. Con l'introduzione del rumore sul Nodo 0, la configurazione rilevata vede la coppia **N001 e N003 agire in coordinazione verso il target N000**. Questo spostamento dimostra empiricamente come la redistribuzione del rumore e delle

probabilità di attivazione all'interno di una topologia closed-loop possa alterare l'ordine di attivazione apparente dei moduli neuronali, modificando i vettori di sincronizzazione scoperti dagli algoritmi di analisi statistica.

2. Perturbando il nodo 1 con una probabilità pari a 0.2: In questo secondo test, abbiamo attivato il rumore (con una probabilità di 0.2) **soltanto sul Nodo 1**, lasciando tutti gli altri nodi normali.

Guardando i dati della traiettoria, si nota che il Nodo 1 si accende più volte in modo casuale durante il tempo proprio a causa di questa stimolazione. Tuttavia, **il suo segnale rimane totalmente isolato**. Tutti gli altri nodi (N000, N002, N003, N004) rimangono spenti a 0.

Il motivo di questo blocco è legato ai collegamenti (pesi) della rete: il Nodo 1 è collegato al Nodo 2 con un peso negativo (inibitorio) pari a -0.31. Questo significa che ogni volta che il Nodo 1 si accende, manda un segnale di "spegnimento" al Nodo 2. Visto che nella rete non ci sono altre spinte o aiuti, questo blocco è sufficiente a fermare l'informazione, impedendo al segnale di fare il giro dell'anello.

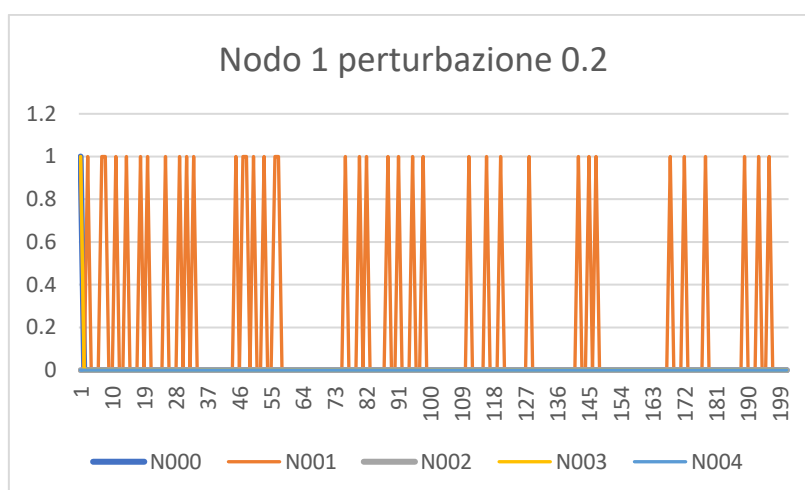


Figura A.5

Per quanto riguarda lo Zi:

8.20358871	2	_N000_N003_	N001
------------	---	-------------	------

Figura A.6

Il valore statistico (l'indice Z) è rimasto identico al caso precedente (**8.2035**), perché l'accensione iniziale al passo zero (dove i nodi sono partiti attivi insieme) è talmente forte da dominare i calcoli.

La cosa interessante, però, è che aggiungendo il rumore sul Nodo 1, i ruoli si sono invertiti: adesso il programma legge i nodi **N000 e N003 come la coppia che si coordina per attivare il Nodo 1**. Questo ci dimostra che basta spostare il rumore su un singolo neurone per cambiare il modo in cui l'algoritmo legge la direzione dei segnali nel cervello artificiale.

3. Perturbando il nodo 2 con una probabilità pari a 0.2: a differenza dei casi precedenti, qui succede qualcosa di nuovo: **il segnale non rimane isolato, ma viaggia**. Guardando la traiettoria, ogni volta che il Nodo 2 si accende a 1 per via del rumore, **esattamente al passo successivo si accende anche il Nodo 3**. Il Nodo 2 è collegato al Nodo 3 con un **peso positivo (eccitatorio) molto alto, pari a +0.82**. Quando il Nodo 2 si attiva, dà una forte spinta al Nodo 3 e lo fa accendere subito dopo. Il segnale poi si ferma sul Nodo 4 perché il collegamento successivo torna a essere negativo (peso -0.17).

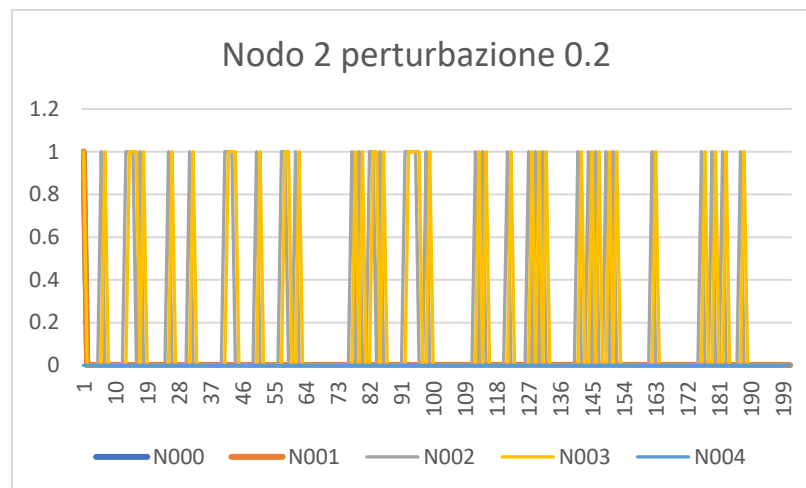


Figura A.7

Anche in questo caso, il programma ha calcolato i pattern di attivazione e ha generato questo risultato:

8.20359	3	_N000_N001_	N002	N003
1.55498	2	__N000_N001__N003_	N002	

Figura A.8

4. Perturbando il nodo 3 con una probabilità pari a 0.2: i dati della traiettoria mostrano che il Nodo 3 risponde bene alla stimolazione, accendendosi molte volte in modo casuale durante la simulazione. Tuttavia, proprio come è successo per i nodi 0 e 1, **il suo segnale non riesce a muoversi e rimane intrappolato**. Tutti gli altri nodi (N000, N001, N002, N004) rimangono spenti a 0 per tutto il tempo. il Nodo 3 è collegato al Nodo 4 con un **peso negativo (inibitorio) pari a -0.17**. Anche se questo valore inibitorio è numericamente piccolo, in questa rete basta a "spegnere" sul nascere il Nodo 4. Di conseguenza, l'informazione si ferma subito e non può proseguire lungo l'anello.

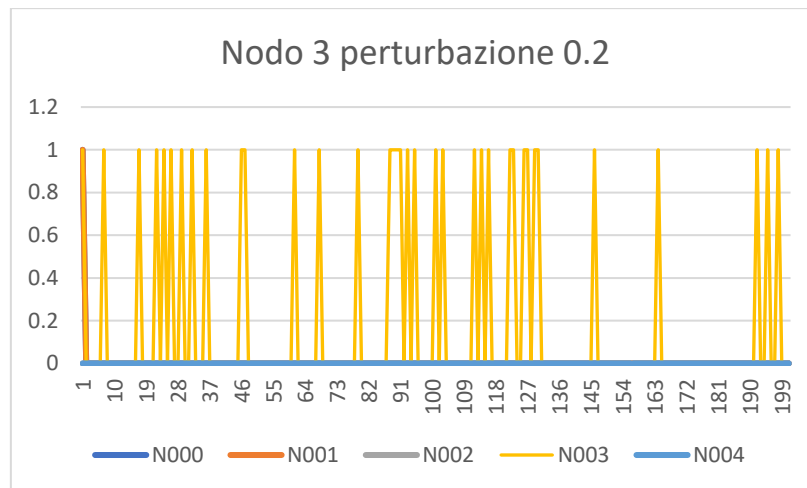


Figura A.9

Non essendoci stata una vera propagazione oltre al rumore isolato, il programma ha estratto solo il pattern di base:

8.20359	2	_N000_N001_	N003
---------	---	-------------	------

Figura A.10

L'indice Zi rimane fermo a **8.2035** a causa dell'impatto della scarica iniziale al passo zero. In questo caso specifico, l'algoritmo legge i nodi **N000 e N001 come la coppia che punta al target N003**. Questo ci conferma che, quando il rumore colpisce un nodo che si trova subito dopo un blocco inibitorio, la rete non riesce a generare nuove dinamiche di gruppo e i pattern statistici tornano alla configurazione standard di partenza.

- Perturbando il nodo 4 con una probabilità pari a 0.2: i dati mostrano che il Nodo 4 si attiva molte volte in modo stocastico. Tuttavia, proprio come nei casi 0, 1 e 3, **il segnale non riesce a fare il giro dell'anello**. Il Nodo 4 non riesce a trasmettere lo stimolo al Nodo 0, e di conseguenza tutta la rete rimane ferma a 0. Il motivo è sempre lo stesso: il Nodo 4 si collega al Nodo 0 tramite un **peso inibitorio (negativo) pari a -0.31**. Ogni volta che il Nodo 4 si accende, blocca il Nodo 0. Il segnale muore sul nascere a causa di questo sbarramento.

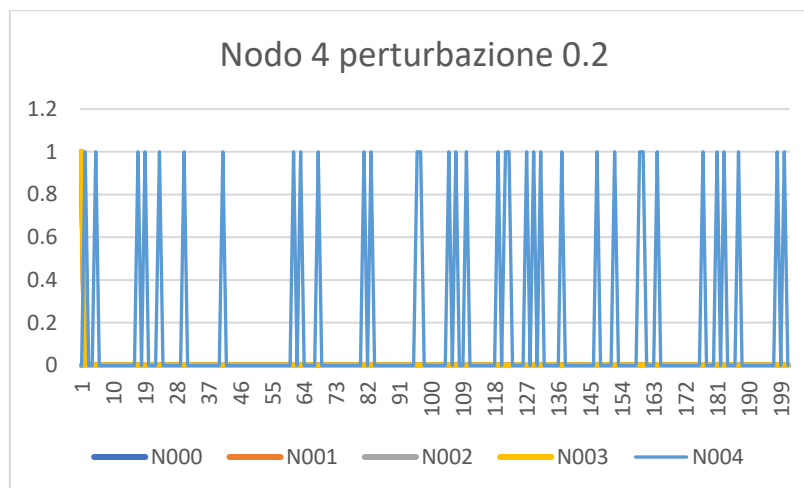


Figura A.11

Anche se il segnale non si propaga, l'algoritmo rileva che l'attività del Nodo 4 si inserisce in una struttura temporale legata all'evento iniziale. Ha estratto due pattern significativi:

8.203589	3	_N000_N001_	N003	N004
8.203589	2	__N000_N001__N003_	N004	

Figura A.12

Entrambi i valori hanno lo Z-score fisso a **8.2035** a causa del fortissimo peso del passo zero. Rispetto al Nodo 2, qui l'algoritmo non sta vedendo una nuova propagazione di segnale, ma cattura semplicemente il fatto che il Nodo 4 continua ad accendersi da solo all'interno dell'intervallo temporale in cui rileva la memoria storica della rete.

6. Perturbando contemporaneamente il nodo 2 e il nodo 4 con probabilità 0.2:

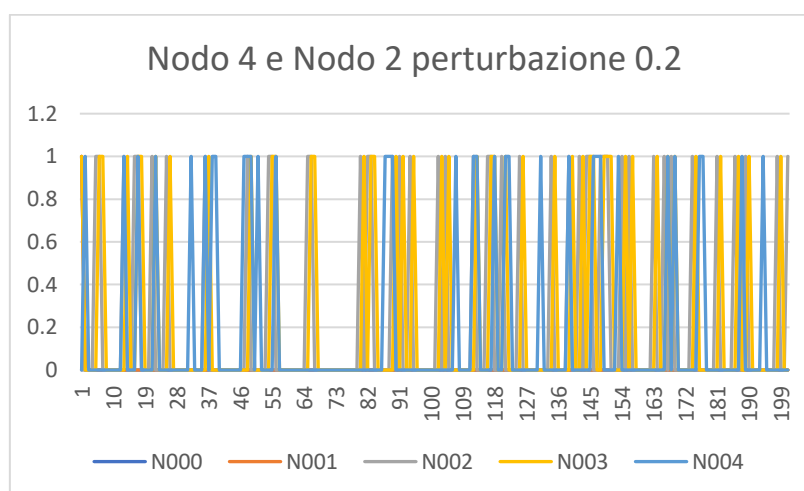


Figura A.13

Il Nodo 2 si accende spesso a causa del rumore. Come avevamo già scoperto nei test singoli, questo neurone è il vero motore dell'anello: ogni volta che si attiva a 1, trascina quasi sempre con sé il Nodo 3 al passo successivo.

Anche il Nodo 4 si accende molte volte a causa del rumore (da solo o insieme al Nodo 2). Tuttavia, il suo segnale si spegne all'istante.

Ciò significa che il rumore simultaneo crea una rete a due velocità: una metà dell'anello continua a far viaggiare impulsi (da 2 a 3), mentre l'altra metà subisce un blocco totale (da 4 a 0) che impedisce al segnale di completare il giro del cerchio.

Per quanto riguarda lo Zi:

8.20358871	4	_N000_N001_	N002	N003	N004
1.42177262	3	__N000_N001__N003_	N002	N004	
-0.19356454	2	___N000_N001__N003__N004_	N002		

Figura A.14

Appendice B

Topologia ad albero

Osservando la traiettoria dei dati ottenuta, notiamo un fenomeno molto netto. Al passo iniziale ($t = 0$), la rete si accende in un punto ben preciso: risultano attivi contemporaneamente i nodi **N002**, **N008**, **N009**, **N010** e **N012**.

Tuttavia, già a partire dal passo successivo ($t = 1$) e per tutto il resto della simulazione, **la rete si spegne completamente**, mostrando zeri in tutte le colonne. Non c'è alcuna attività stocastica successiva o propagazione nel tempo. Il segnale legato alla scarica iniziale muore immediatamente.

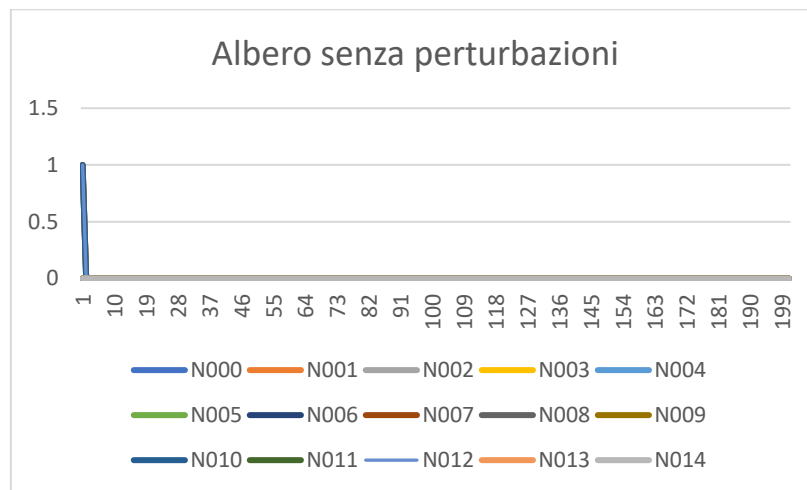


Figura B.1

La spiegazione di questo spegnimento istantaneo si trova guardando i pesi sinaptici dei nodi coinvolti: molti collegamenti chiave di questo albero hanno pesi negativi molto vicini a -1. Ad esempio, il legame da Nodo 4 a Nodo 10 è pari a -0.5199, e quello da Nodo 5 a Nodo 12 è -0.9432. Inoltre, non è stata inserita una perturbazione continua (rumore) capace di tenere vivi i neuroni artificiali dopo il passo zero. Senza uno stimolo stocastico esterno, le forti inibizioni locali bloccano sul nascere qualsiasi tentativo di trasmissione del segnale verso i rami più profondi dell'albero.

Nonostante l'attività si esaurisca subito, il programma di analisi è riuscito a estrarre **tre pattern significativi**, tutti con lo stesso identico Z-score di **8.2035**:

8.203589	4	_N002_N008_	N009	N010	N012
8.203589	3	_N002_N008_N009_	N010	N012	
8.203589	2	___N002_N008_N009_N010_	N012		

Figura B.2

Modifica delle probabilità e inserimento del rumore

1. Perturbando il nodo 0 con una probabilità pari a 0.1:

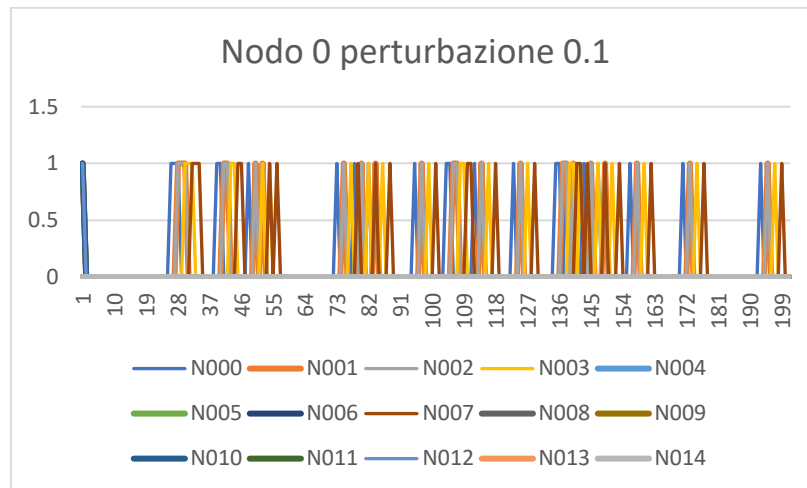


Figura B.3

L'introduzione di una componente stocastica continua (rumore a 0.1) sul Nodo 0 ha radicalmente cambiato il comportamento della rete ad albero, attivando una vera e propria catena di propagazione del segnale.

Analizzando la traiettoria temporale, si nota chiaramente che ogni volta che il Nodo 0 si attiva a causa del rumore si innesca un preciso "effetto domino" che attraversa l'albero passo dopo passo.

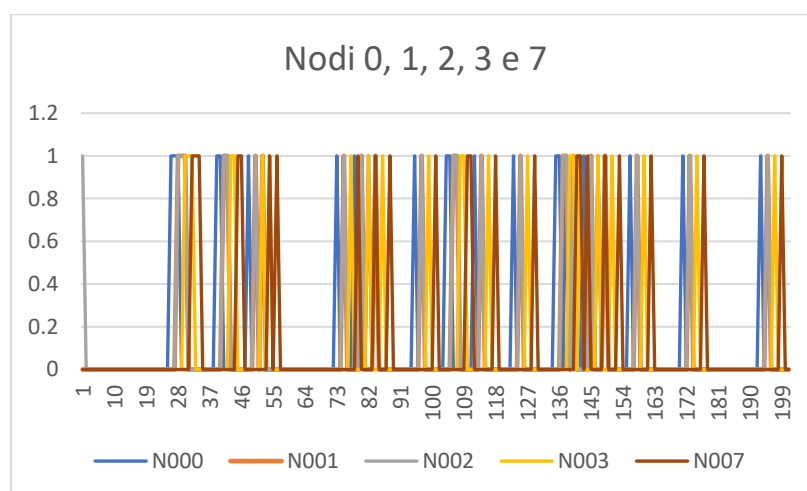


Figura B.4

Per quanto riguarda lo Zi:

94.51687414	8	_N001_N002_	N000	N003	N007	N008	N009	N010	N012
8.203588715	7	_N008_N010_	_N001_N002_	N000	N003	N007	N009	N012	
8.203588715	6	_N008_N010_N009_	_N001_N002_	N000	N003	N007	N012		
8.203588715	5	_N008_N010_N009_N012_	_N001_N002_	N000	N003	N007			
5.30081321	4	_N008_N010_N009_N012_N001_N002_	N000	N003	N007				
3.507583019	3	_N000_N007_	_N008_N010_N009_N012_N001_N002_	N003					
-0.84870278	2	_N008_N010_N009_N012_N001_N002_N003_	_N000_N007_						

Figura B.5

2. Perturbando sempre il nodo 0, ma con probabilità 0.2:

Rispetto al test con rumore a 0.1, in cui i blocchi di propagazione erano intervallati da lunghi periodi di silenzio, qui il Nodo 0 si attiva ripetutamente a brevi distanze.

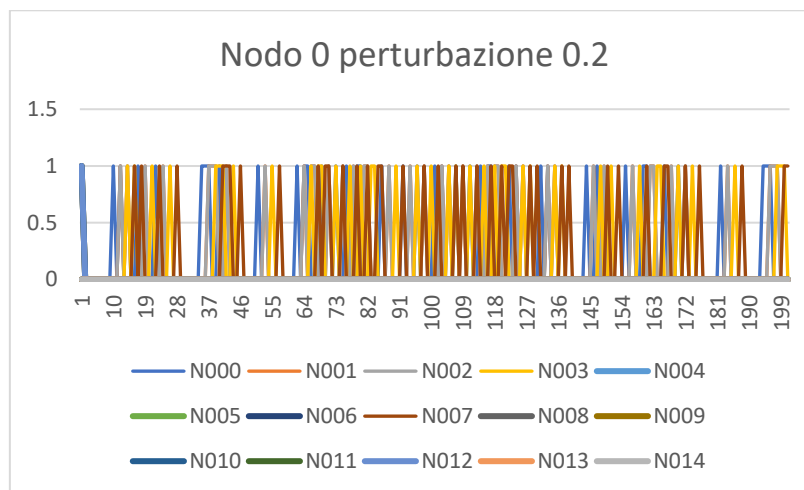


Figura B.6

Quando il Nodo 0 si accende per più passi consecutivi il segnale non viaggia più come un singolo impulso isolato, ma si sovrappone nel tempo. Si nota infatti che i nodi figli (N001, N002, N003) rimangono accesi contemporaneamente per più istanti, creando una propagazione "sostenuta" ed estremamente solida.

Il resto dell'albero continua a rimanere totalmente spento.

Lo Z-score del pattern principale compie un balzo straordinario, passando dal già elevato 94.51 dell'esperimento precedente all'impressionante valore di **138.2774**. La significatività statistica del pattern è aumentata perché l'evento coordinato si è ripetuto molte più volte lungo la traiettoria.

138.2774	8	_N001_N002_	N000	N003	N007	N008	N009	N010	N012
8.203589	7	_N008_N010_	_N001_N002_	N000	N003	N007	N009	N012	
8.203589	6	_N008_N010_N009_	_N001_N002_	N000	N003	N007	N012		
8.203589	5	_N008_N010_N009_N012_	_N001_N002_	N000	N003	N007			
5.300813	4	_N008_N010_N009_N012_N001_N002_	N000	N003	N007				
-0.50745	2	_N008_N010_N009_N012_N001_N002_N000_N007_	N003						

Figura B.7

- Perturbando il nodo 1 con probabilità 0.1: poiché il Nodo 0 non riceve più alcuna stimolazione, rimane spento a 0 per tutta la simulazione. Il Nodo 2 (insieme ai nodi 8, 9, 10 e 12) risulta inizialmente attivo (1), nonostante il rumore sia applicato al Nodo 1.

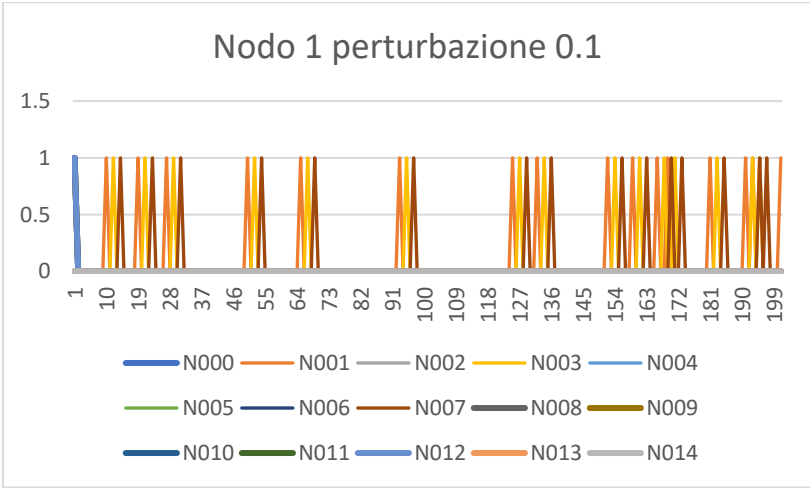


Figura B.8

Per quanto riguarda lo Zi:

8.20359	7	_N008_N010_	N001	N002	N003	N007	N009	N012
8.20359	6	_N008_N010_N002_	N001	N003	N007	N009	N012	
8.20359	5	_N008_N010_N002_N009_	N001	N003	N007	N012		
8.20359	4	_N008_N010_N002_N009_N012_	N001	N003	N007			
1.12474	3	_N001_N007_	_N008_N010_N002_N009_N012_	N003				
-0.5971	2	_N008_N010_N002_N009_N012_N003_	_N001_N007_					

Figura B.9

- Perturbando il nodo 1 con probabilità 0.2:

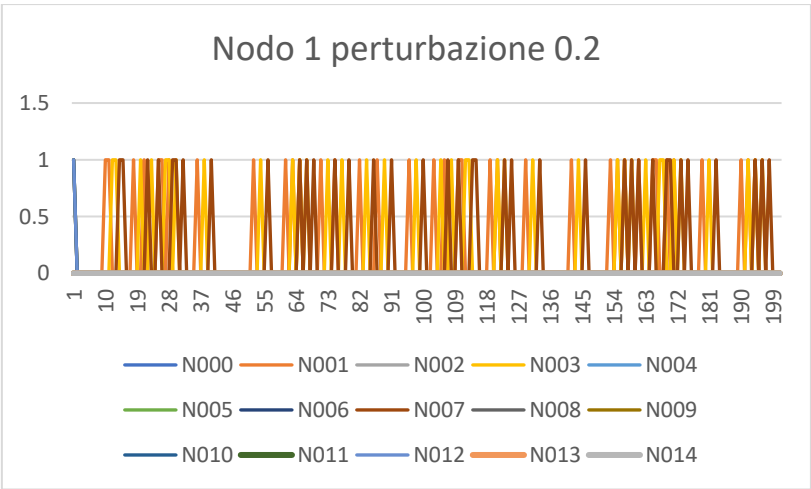


Figura B.10

Il Nodo 1 si attiva ripetutamente a brevi distanze. Quando la perturbazione si concentra in passaggi ravvicinati: l'attivazione continua del nodo N001 mantiene eccitato il nodo N003, che inizia a mostrare scariche prolungate rimanendo acceso per due passi consecutivi. Di riflesso, l'output terminale della catena, ovvero il **Nodo 7**, risponde con la medesima densità, attivandosi molto più frequentemente rispetto al test con rumore a 0.1.

Per quanto riguarda lo Zi:

8.20359	7	_N008_N010_	N001	N002	N003	N007	N009	N012
8.20359	6	_N008_N010_N002_	N001	N003	N007	N009	N012	
8.20359	5	_N008_N010_N002_N009_	N001	N003	N007	N012		
8.20359	4	_N008_N010_N002_N009_N012_	N001	N003	N007			
-0.0004	3	_N001_N003_	_N008_N010_N002_N009_N012_	N007				
-0.4186	2	_N008_N010_N002_N009_N012_N007_	N001_N003_					

Figura B.11

5. Perturbando il nodo 3 con probabilità 0.1:

I nodi da N000 a N002 rimangono completamente spenti a 0 (tranne il nodo N002 passo iniziale).

Si può osservare che ogni volta che il Nodo 3 si attiva in modo casuale a causa del rumore il segnale si propaga istantaneamente al Nodo 7 al passo successivo.

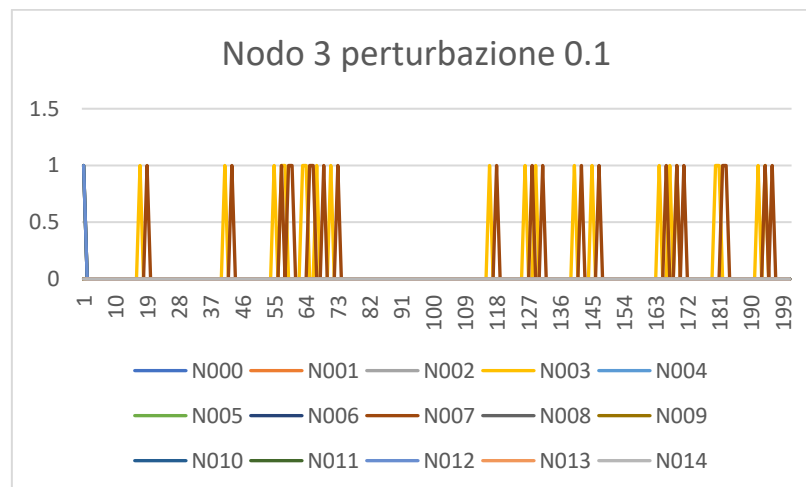


Figura B.12

Per quanto riguarda lo Zi:

8.203588715	6	_N008_N009_	N002	N003	N007	N010	N012
8.203588715	5	_N008_N009_N002_	N003	N007	N010	N012	
8.203588715	4	_N008_N009_N002_N012_	N003	N007	N010		
8.203588715	3	_N008_N009_N002_N012_N010_	N003	N007			
-0.4671087	2	_N003_N007_	_N008_N009_N002_N012_N010_				
0	0		0	0			

Figura B.13

6. Perturbando sempre il nodo 3 ma con probabilità 0.2:
Rispetto alla perturbazione 0.1, il Nodo 3 genera impulsi in modo molto più fitto e ravvicinato. Questa stimolazione densa innesca una forte sommazione temporale sul Nodo 7, che risponde con accensioni prolungate e ripetute.

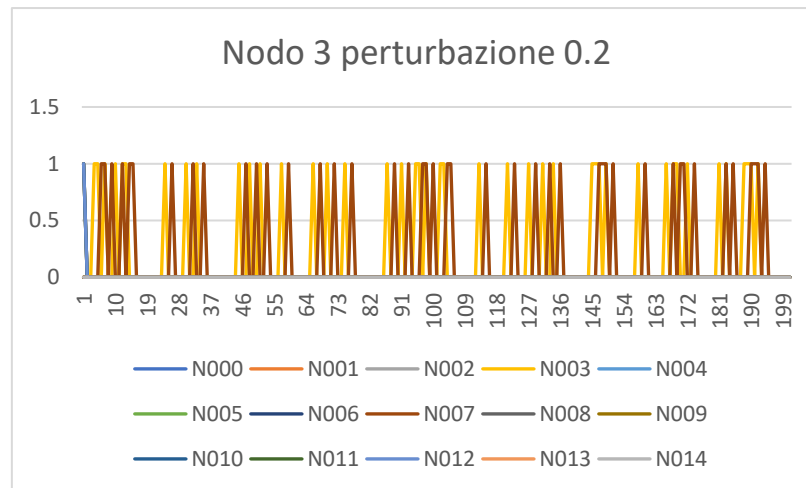


Figura B.14

Per quanto riguarda lo Zi:

8.20358871	6	_N008_N009_	N002	N003	N007	N010	N012
8.20358871	5	_N008_N009_N002_	N003	N007	N010	N012	
8.20358871	4	_N008_N009_N002_N012_	N003	N007	N010		
8.20358871	3	_N008_N009_N002_N012_N010_	N003	N007			
-0.36573057	2	_N008_N009_N002_N012_N010_N003_	N007				

Figura B.15

7. Perturbando il nodo 4 con probabilità 0.1:
Guardando la serie temporale dei dati, notiamo subito un comportamento radicalmente diverso rispetto a quando avevamo stimolato il Nodo 3: il Nodo 4 risponde alla perturbazione e si accende più volte in modo casuale durante il tempo. Tuttavia, il suo segnale si spegne all'istante e non viaggia mai. Tutti i restanti 14 nodi della rete rimangono fissi a 0, nonostante i nodi N002, N008, N009, N010 e N012 siano accesi al primo passo.
Il motivo di questo sbarramento totale si trova guardando i figli biologici a cui il Nodo 4 è collegato, ovvero il Nodo 9 e il Nodo 10. La nostra rete ci dice che la sinapsi da Nodo 4 a Nodo 9 ha un peso **negativo** pari a -0.2743 e la sinapsi da Nodo 4 a Nodo 10 ha un peso **negativo** pari a -0.5199.
Questo significa che il Nodo 4 controlla un ramo interamente **inibitorio**. Ogni volta che si accende a causa del rumore, anziché svegliare i suoi figli, manda loro un comando di spegnimento. Bloccando sul nascere i nodi 9 e 10, il segnale muore sul colpo all'interno del neurone sorgente, impedendo all'impulso di scendere verso la periferia di quel ramo dell'albero.

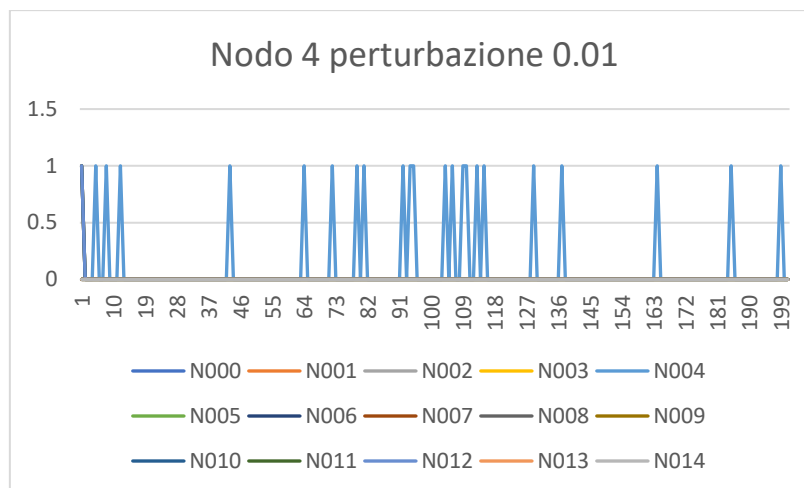


Figura B.16

Per quanto riguarda lo Zi:

8.20358871	5	_N008_N010_	N002	N004	N009	N012
8.20358871	4	__N008_N010__N002_	N004	N009	N012	
8.20358871	3	___N008_N010___N002___N012_	N004	N009		
8.20358871	2	____N008_N010____N002____N012____N009_	N004			

Figura B.17

I quattro pattern estratti mantengono tutti lo Z-score pari a 8.2035.

8. Perturbando sempre il nodo 4 ma con probabilità 0.2:
 Rispetto al caso con probabilità 0.1, il Nodo 4 si attiva con frequenza maggiore, ma comunque i suoi figli, ovvero il Nodo 9 e il Nodo 10, sono accesi al primo passo per poi spegnersi e non accendersi più.

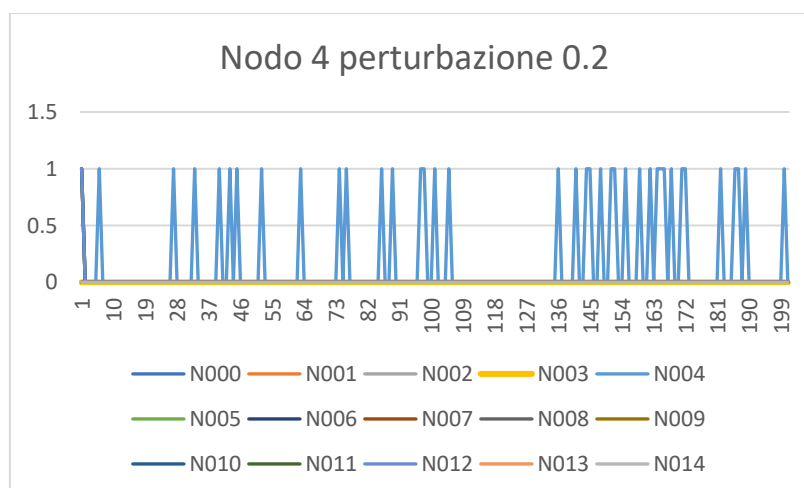


Figura B.18

Per quanto riguarda lo Zi:

8.20358871	5	_N008_N010_	N002	N004	N009	N012
8.20358871	4	_N008_N010_N002_	N004	N009	N012	
8.20358871	3	__N008_N010__N002__N012_	N004	N009		
8.20358871	2	___N008_N010__N002__N012__N009_	N004			

Figura 39

Bibliografia

- **Amari, S. I. (2001).** Information geometry on hierarchy of probability distributions. *IEEE Transactions on Information Theory*, 47(5), 1701-1711.
<https://doi.org/10.1109/18.930911>
- **Buzsáki, G. (2010).** Neural syntax: cell assemblies, synapses, and concepts. *Neuron*, 68(3), 362-385.
<https://doi.org/10.1016/j.neuron.2010.09.023>
- **Cover, T. M., & Thomas, J. A. (2006).** *Elements of Information Theory* (2nd ed.). John Wiley & Sons.
- **D'Addese, G.** *Individuazione di organizzazioni dinamiche in sistemi complessi: aspetti teorici e applicazioni*. Tesi di Laurea Magistrale in Informatica, Università degli Studi di Modena e Reggio Emilia, Dipartimento di Scienze Fisiche, Informatiche e Matematiche. Relatori: M. Villani, L. La Rocca.
- **D'Addese, G., Sani, L., La Rocca, L., Serra, R., & Villani, M. (2021).** *Asymptotic Information-Theoretic Detection of Dynamical Organization in Complex Systems*. *Entropy*, 23(4), 398.
<https://doi.org/10.3390/e23040398>
- **Harris, K. D. (2005).** Neural signatures of cell assemblies. *Current Opinion in Neurobiology*, 15(5), 518-523.
<https://doi.org/10.1016/j.conb.2005.08.010>
- **Hebb, D. O. (1949).** *The Organization of Behavior: A Neuropsychological Theory*. John Wiley & Sons.
- **Hubert, L., & Arabie, P. (1985).** Comparing partitions. *Journal of Classification*, 2(1), 193-218.
<https://doi.org/10.1007/BF01908075>
- **Humphries, M. D. (2011).** Spike-train communities: Finding groups of co-active neurons in high-density recordings. *Journal of Neuroscience*, 31(7), 2321-2336.
<https://doi.org/10.1523/JNEUROSCI.4742-10.2011>
- **Londei, F., Ceccarelli, F., Arena, G., Ferrucci, L., Russo, E., Brunamonti, E., & Genovesio, A. (2024).** *Out of the single-neuron straitjacket: neurons within assemblies change selectivity and their reconfiguration underlies dynamic coding*. bioRxiv preprint.
- **Lopes-dos-Santos, V., Ribeiro, S., & Tort, A. B. (2013).** Detecting cell assemblies in large neuronal populations. *Journal of Neuroscience Methods*, 220(2), 149-166.
<https://doi.org/10.1016/j.jneumeth.2013.04.010>
- **Makdah, G., Wiener, S. I., & Pompili, M. N. (2024).** *Detection of cell assemblies in high-density extracellular electrophysiological recordings*. bioRxiv preprint.
<https://doi.org/10.1101/2024.01.26.577338>
- **Newman, M. E. (2018).** *Networks* (2nd ed.). Oxford University Press.
<https://doi.org/10.1093/oso/9780198805090.001.0001>
- **Quaglio, P., Rostami, V., Torre, E., & Grün, S. (2018).** *Methods for identification of spike patterns in massively parallel spike trains*. *Biological Cybernetics*, 112, 57-80.
<https://doi.org/10.1007/s00422-018-0755-0>
- **Rand, W. M. (1971).** Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336), 846-850.

<https://doi.org/10.1080/01621459.1971.10482356>

- **Tononi, G., Sporns, O., & Edelman, G. M. (1994).** A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences*, 91(11), 5033-5037.
<https://doi.org/10.1073/pnas.91.11.5033>
- **Villani, M., Roli, A., Filisetti, A., & Serra, R. (2015).** *Interpreting the dynamics of complex systems through the organization of their relevant subsets*. Workshop on Artificial Life and Evolutionary Computation (WIVACE 2015).