

UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI INGEGNERIA
Dipartimento di Ingegneria dell'Informazione
Corso di Laurea Magistrale in Ingegneria Informatica e dell'Automazione



TESI DI LAUREA

**Un framework basato su multilayer network per analizzare la
resilienza di social network multimodali**

**A multilayer network-based framework for investigating the
resilience of multimodal social networks**

Relatore

Prof. Domenico Ursino

Correlatori

Dott. Francesco Cauteruccio

Dott. Enrico Corradini

Candidato

Edoardo Giannelli

ANNO ACCADEMICO 2022-2023

Sommario

Nella letteratura recente, i multilayer network vengono sempre più utilizzati per modellare e gestire scenari e sistemi complessi, uno dei quali è rappresentato dai multimodal social network, ovvero una categoria di reti in cui i nodi e gli archi possono essere di diverse tipologie. In questo caso, ciascun layer del multilayer network può essere utilizzato per modellare un tipo di nodi del multimodal social network. Successivamente, i layer vengono connessi al loro interno da archi intra-layer e al loro esterno da archi inter-layer. In questa tesi, vogliamo dimostrare la realizzabilità di questa idea, proponendo un framework per studiare i multimodal social network attraverso l'analisi di un multilayer network ad esso corrispondente. In particolare, utilizziamo il nostro framework per affrontare due questioni significative, ovvero l'evoluzione temporale dei multilayer network e la loro resilienza alle perturbazioni intra- e inter-layer. Dopo aver introdotto il modello e descritto i dettagli tecnici del nostro framework, presentiamo diversi esperimenti che dimostrano l'efficacia del presente studio.

Keywords: Multilayer Network; Multimodal Social Network; Evoluzione temporale delle reti; Resilienza delle reti; Modularità delle reti; Cross-Correlation Coefficient.

Introduzione	1
1 Multilayer network	3
1.1 Introduzione ai multilayer network	3
1.2 Dai grafi semplici ai multilayer network	4
1.2.1 Network e grafi singoli	4
1.2.2 I multilayer network	7
1.3 Processi dinamici nei multilayer network	10
1.3.1 Percolation e multilayer network	10
1.3.2 Processi di diffusione nei multilayer network	11
1.4 Esempi di utilizzo dei multilayer network	12
1.4.1 Ecologia	12
1.4.2 Biologia	12
1.4.3 Trasporti	13
1.4.4 Cervello umano	14
1.4.5 Economia	14
1.4.6 Teoria dei giochi	14
2 Multimodal social network	15
2.1 Introduzione alla Social Network Analysis	15
2.2 Verso i multimodal social network	16
2.2.1 1-mode network: definizione e proprietà	16
2.2.2 2-mode network: definizione e applicazioni	19
2.3 Multimodal network	21
2.4 Applicazioni reali nella Social Network Analysis	23
2.4.1 Misure di centralità	24
2.4.2 Triadi	25
3 Related Literature	28
4 Descrizione del framework	32
4.1 Descrizione del modello basato su un multilayer network	32
4.1.1 Definizione dell'istanza temporale di un multilayer network	33
4.2 Misure associate al framework	34
4.2.1 Misure relative alla struttura di rete	34

4.2.2	Misure relative alle comunità	35
4.2.3	Misure relative alla resilienza	37
5	Risultati sperimentali	41
5.1	Raccolta e pre-elaborazione dei dati	41
5.2	Analisi esplorativa	42
5.3	Analisi strutturale	44
5.3.1	Analisi intra-layer	44
5.3.2	Analisi inter-layer	47
5.4	Analisi delle comunità	48
5.4.1	Analisi intra-layer	48
5.4.2	Analisi inter-layer	49
5.5	Analisi della resilienza	50
5.5.1	Analisi intra-layer	50
5.5.2	Analisi inter-layer	53
5.6	Discussione sui risultati ottenuti	56
	Conclusioni	58
	Bibliografia	59
	Ringraziamenti	63

Elenco delle figure

1.1	Esempio di network	4
1.2	Esempio di multilayer network	5
1.3	Esempio di matrice di adiacenza per un grafo orientato	6
1.4	Estensione della matrice di adiacenza per un multilayer network	8
1.5	Esempio di percolation nel caso multilayer	10
1.6	Multilayer network applicato al trasporto urbano	13
2.1	Esempio di matrice di adiacenza pesata	16
2.2	Esempio di lista di adiacenza	17
2.3	Differenza tra BFS e DFS	18
2.4	Cammino minimo tra A e F	18
2.5	Esempio di small worls network	19
2.6	Esempio di 2-mode network	20
2.7	Esempio di un 2-mode network che descrive la relazione tra medici e pazienti, sia in forma grafica che con matrice di adiacenza	21
2.8	Esempio di multimodal network	22
2.9	Multimodal network in formato matriciale	22
2.10	Visualizzazione grafica di un multimodal network composto da 5 tipi di nodi	23
2.11	Triade indiretta (sinistra) e triade diretta (destra)	26
4.1	Rappresentazione del multilayer network proposto	33
5.1	Trend temporale dei commenti nel dataset	43
5.2	Distribuzione dei commenti rispetto agli utenti in scala semi-logaritmica	43
5.3	Distribuzione dei commenti rispetto agli argomenti in scala semi-logaritmica	44
5.4	Valori di r_L per ogni layer nel tempo	44
5.5	Valori di γ_L per ogni layer nel tempo	45
5.6	Valori di \mathcal{L}_L per ogni layer nel tempo	46
5.7	Valori di r_{L_1, L_2} per coppie di layer nel tempo	47
5.8	Valori di Q_L per ogni layer nel tempo	48
5.9	Valori di Q_M per ogni layer nel tempo	49
5.10	Valori di $\Delta D_L[\tau_1, \tau_2]$ per il layer <i>Authors</i> nel tempo	50
5.11	Valori di $\Delta D_L[\tau_1, \tau_2]$ per il layer <i>Comments</i> nel tempo	51
5.12	Valori di $\Delta D_L[\tau_1, \tau_2]$ per il layer <i>Topics</i> nel tempo	51
5.13	Valori di $\Delta \gamma_L[\tau_1, \tau_2]$ per il layer <i>Authors</i> nel tempo	52

5.14	Valori di $\Delta\gamma_L[\tau_1, \tau_2]$ per il layer <i>Comments</i> nel tempo	52
5.15	Valori di $\Delta\gamma_L[\tau_1, \tau_2]$ per il layer <i>Topics</i> nel tempo	53
5.16	Valori di $\Delta r_{L_1, L_2}[\tau_1, \tau_2]$ per i layer <i>Authors</i> e <i>Comments</i> nel tempo	54
5.17	Valori di $\Delta D_E[\tau_1, \tau_2]$ nel tempo	55
5.18	Valori di $\Delta Q_{\mathcal{M}}[\tau_1, \tau_2]$ nel tempo	55
5.19	Valori di $\Delta D_E[\tau_1, \tau_2]$ nel tempo	56

Elenco delle tabelle

5.1 Statistiche di base del dataset proposto 42

Viviamo in un'epoca di interconnessione e di proliferazione di dati senza precedenti, in cui stiamo assistendo alla nascita e allo sviluppo di sistemi sempre più complessi, i quali sono diventati un tema centrale nello studio di numerose discipline scientifiche. In questo contesto, la network analysis è emersa come un framework potente ed efficace per indagare e comprendere le intricate relazioni tra le entità del mondo reale. Gli esempi includono le interazioni sociali [23, 12, 36], i processi biologici e medici [16, 42, 22] e le applicazioni tecnologiche [15, 9].

Gli approcci tradizionali alla network analysis si sono concentrati principalmente sullo studio dei pattern di connettività e delle proprietà strutturali delle reti argomento di indagine; tuttavia, poiché queste ultime sono diventate progressivamente più sfaccettate (e, quindi, multimodali [41]) e interconnesse, sorge la necessità di sviluppare framework analitici più sofisticati, in grado di catturare la complessità e l'eterogeneità proprie di questi sistemi.

Grazie alla capacità di integrare più fonti di dati e di modellare l'eterogeneità delle interdipendenze, il multilayer network ha guadagnato, in questo contesto, un'attenzione significativa come strumento efficace per rappresentare e analizzare sistemi complessi e intricati, caratterizzati da molteplici tipi di interazioni o layer [35, 22, 27]. I multilayer network rappresentano, quindi, un notevole miglioramento rispetto ai tradizionali single-layer network utilizzati finora, i quali non sono in grado di catturare la natura eterogenea delle relazioni che caratterizza molti sistemi complessi, spesso caratterizzati da molteplici tipologie di interazioni e da più layer che operano simultaneamente. Queste reti permettono, quindi, di facilitare l'identificazione di proprietà emergenti, la scoperta di modelli nascosti e sono in grado di supportare lo sviluppo di nuove strategie per ottimizzare le prestazioni di sistema, progettare interventi efficienti e gestire i rischi [18].

Per i motivi appena citati, negli ultimi anni le caratteristiche e le applicazioni dei multilayer network sono state studiate approfonditamente dalla comunità scientifica [1, 8, 38, 17, 11, 13, 28]. Tuttavia, riteniamo che ci siano due aspetti di queste reti che sono stati ancora poco approfonditi; nello specifico, essi riguardano la capacità di modellare l'evoluzione temporale e la resilienza.

In merito al primo aspetto, pensiamo che la modellazione delle interazioni variabili nel tempo tra le entità di un multilayer network consente un'analisi efficace delle dinamiche temporali che coinvolgono la rete e il sistema modellato attraverso di essa. Molti sistemi del mondo reale, infatti, presentano un comportamento evolutivo strettamente dipendente dal tempo; ne sono un esempio i social network online, le reti di trasporto e i modelli di comunicazione [24, 20, 37]. Catturando l'evoluzione temporale delle interdipendenze, i multilayer network possono facilitare la comprensione di come questi sistemi crescono, si

evolvono, si adattano e si riconfigurano nel tempo.

Per quanto riguarda la resilienza, esistono studi che analizzano questo aspetto all'interno dei multilayer network quando questi vengono utilizzati in un dominio preciso, come i sistemi ingegneristici [40, 39]. Tuttavia, tali studi sono specifici per quel settore e difficilmente sono stati estesi a contesti applicativi molto diversi dai multilayer network, come i multimodal social network. Per questo, riteniamo che uno studio generale sulla resilienza dei multilayer network, che non dipenda dal campo in cui vengono applicati, potrebbe aiutare a far luce sulle intricate interdipendenze esistenti all'interno dei vari fenomeni che queste reti modellano in diversi ambiti.

Nel contesto appena delineato, il presente lavoro intende fornire un importante contributo; in particolare, esso propone un framework per studiare l'evoluzione temporale e la resilienza di un multilayer network associato a un multimodal social network, ovvero una rete sociale dedicata alla condivisione e alla diffusione di informazioni. La struttura presentata opera su un multilayer network composto da tre layer che, nel nostro caso, rappresentano autori, commenti e topic di un multimodal social network. Tuttavia, è necessario precisare che, sebbene il nostro framework e le analisi ad esso associate siano state definite per un multilayer network che rappresenta una rete sociale con determinate caratteristiche, questi sono abbastanza generali da poter essere applicati a qualsiasi multilayer network che modella qualunque scenario applicativo, anche diverso dai social network.

Il framework proposto comprende tre serie di analisi; la prima riguarda la struttura dei multilayer network, che ci permette di studiare l'evoluzione temporale delle proprietà strutturali della rete, come, ad esempio la degree-degree cross-correlation; la seconda si concentra sull'organizzazione e sulla modularità della rete attraverso lo studio delle comunità; infine, la terza permette di analizzare la resilienza della rete in presenza di perturbazioni intra- e inter-layer. Per eseguire queste analisi, introduciamo una serie di metriche abbastanza generali da poter essere utilizzate per qualsiasi multilayer network.

Riassumendo, i principali contributi di questa tesi sono: (i) la definizione di un framework e di una serie di metriche per studiare l'evoluzione temporale e la resilienza di un multilayer network; (ii) lo sviluppo di un metodo per mappare un multimodal social network in un multilayer network; (iii) l'applicazione del quadro proposto su un caso di studio reale.

La tesi in oggetto è composta da cinque capitoli, strutturati come segue:

- Il Capitolo 1 introduce il lettore alla definizione, alle proprietà e agli esempi applicativi reali dei multilayer network.
- Il Capitolo 2 descrive la struttura, le metriche e le caratteristiche dei multimodal social network.
- Il Capitolo 3 presenta la letteratura correlata.
- Il Capitolo 4 descrive il framework proposto.
- Il Capitolo 5 presenta la campagna sperimentale che abbiamo condotto per valutare l'efficacia delle analisi svolte.

I sistemi complessi sono caratterizzati da molte unità interagenti, codificate in modo particolarmente efficace attraverso la network analysis, ovvero l'analisi delle reti. Negli ultimi due decenni la network science ha fornito molteplici spunti di riflessione e di approfondimento sui sistemi naturali, sociali, biologici e tecnologici. Tuttavia, i sistemi reali sono spesso interconnessi, con molte interdipendenze che non sono adeguatamente catturate dai single network, o reti singole. Per tenere conto di questa fonte di complessità, è necessario una visione più generale, in cui network diversi si evolvono o interagiscono tra loro. Queste reti sono note come multilayer network, o reti multistrato. In questo capitolo forniremo una panoramica delle principali metodologie utilizzate per descrivere i sistemi multilayer e di alcuni processi dinamici rappresentativi che hanno luogo su di essi. Terminiamo, infine, con una sintesi di alcune applicazioni in diversi campi della scienza.

1.1 Introduzione ai multilayer network

Nel 1972 Philip W. Anderson pubblicò il celebre articolo *Why More Is Different* in cui criticava la teoria riduzionista secondo la quale tutto, dalla materia alla vita, può essere descritto interamente a partire da un unico insieme di leggi fondamentali. Secondo la sua visione, l'approccio riduzionista esclude, però, quello costruzionista: in altre parole, essere in grado di ridurre qualcosa alle sue leggi fondamentali non implica la capacità di ricostruirlo a partire da esse.

Questa idea semplice, ma potente, è il concetto cardine alla base dei sistemi complessi, ovvero quei sistemi che mostrano fenomeni emergenti, che non possono essere spiegati o previsti unicamente dalle parti costitutive del sistema. Tale definizione implica, quindi, una visione gerarchica del mondo, antitetica alla limitatezza dell'approccio riduzionista. Un semplice quanto efficace esempio a favore della visione di Anderson è la molecola d'acqua, le cui proprietà non possono essere previste a partire da quelle dell'idrogeno e dell'ossigeno presi singolarmente, ma solo attraverso la loro interazione mediante legami chimici. Infatti, per affrontare correttamente alcuni sistemi è necessaria una visione olistica, che tenga conto sia dei loro costituenti sia delle interazioni di questi ultimi.

Lo strumento più efficace per codificare le caratteristiche dei sistemi complessi in un oggetto matematico è la teoria dei grafi, attraverso i quali è stato possibile valicare i sistemi della fisica classica, i quali per primi ne hanno fatto uso, e studiare il comportamento di una serie molto più ampia di problemi in campi diversi, come le scienze economiche, sociali e biologiche. Lo studio in questi campi ha portato allo sviluppo di uno strumento molto potente per l'analisi dei sistemi complessi: la network science, o scienza delle reti.

Nella sua forma più elementare, una rete riduce un sistema a una struttura astratta costituita da alcune entità semplificate, chiamate nodi, che conservano le proprietà dei componenti originali del sistema, e dai modelli di connessione tra essi, i collegamenti (Figura 1.1). Questa rappresentazione non solo permette una visione più chiara e globale del sistema, ma svela anche alcune proprietà e caratteristiche non previste mediante la semplice analisi delle interazioni dei suoi costituenti.

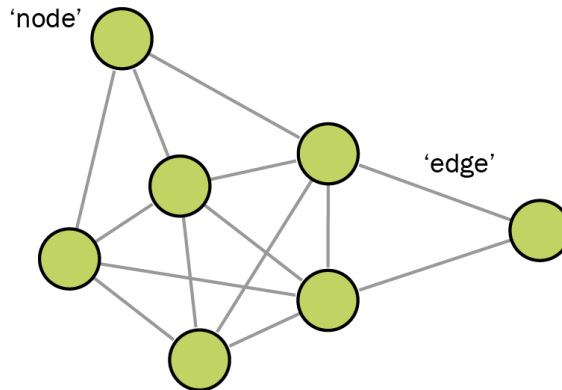


Figura 1.1: Esempio di network

In realtà, tale rappresentazione matematica potrebbe ancora essere una riduzione del sistema originale, ricadendo, così, nello stesso riduzionismo da cui ci si vuole allontanare. Infatti, è possibile aggiungere ulteriori informazioni a questi modelli, mantenendoli connessi e rendendoli, allo stesso tempo, più potenti; ciò è possibile classificando in gruppi le interazioni presenti in un sistema, in base alle loro caratteristiche. Questa classificazione produce un insieme di reti, una per ogni tipo di interazione, collegate tra loro. Estendendo l'analisi al modo in cui queste reti sono connesse tra loro, le entità vengono rappresentate dai nodi e come questi interagiscono tra loro; tutto ciò porta alla codifica di un nuovo insieme di reti che va oltre il concetto di semplici grafi; si tratta dei multilayer network, o reti multistrato (Figura 1.2).

1.2 Dai grafi semplici ai multilayer network

Il formalismo dei multilayer network è un'estensione di quello dei single network (reti singole), cioè delle reti che sono completamente descritte da un solo insieme di entità (nodi) e dalle loro interazioni (link). Per questo motivo, forniamo dapprima una breve introduzione alla nozione di single graph (grafi singoli), per poi trattare più approfonditamente i multilayer network.

1.2.1 Network e grafi singoli

Nella loro forma più generale, le reti sono rappresentate matematicamente dai grafi. Un grafo è definito come una tupla $G = (V, E)$, dove V è l'insieme dei nodi (entità del sistema) ed $E \in V \times V$ è l'insieme dei collegamenti (o archi) tra coppie di nodi quando questi interagiscono in qualche modo. Se E è un insieme non ordinato il grafo si dice non diretto; se è ordinato il grafo è diretto e, in quest'ultimo caso, i cambiamenti nel nodo sorgente in cui inizia il collegamento influenzano il nodo destinazione in cui termina, ma non viceversa.

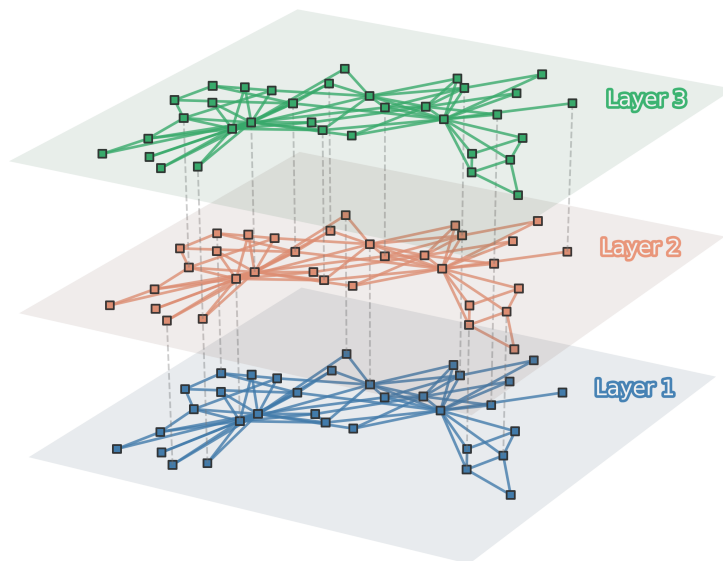


Figura 1.2: Esempio di multilayer network

Inoltre, in alcuni casi i collegamenti hanno un numero associato chiamato peso (weight), il quale rappresenta l'intensità dell'interazione.

La rappresentazione abituale delle reti è la matrice di adiacenza (Figura 1.3). Nello specifico, la matrice di adiacenza A di un single network non diretto è una matrice simmetrica $N \times N$ (dove N è il numero di nodi della rete) con elementi a_{ij} tali che $a_{ij} = 1$ se esiste un arco (edge) tra i nodi i e j , altrimenti $a_{ij} = 0$. Se, al contrario, il single network è diretto allora la matrice non è più simmetrica e $a_{ij} = 1$ solo quando l'arco va da i a j . Se al posto degli 1 nella matrice si trovano dei numeri, questi sono da interpretare come il peso attribuito a ciascun collegamento; ad esempio, se l'insieme dei vertici del grafo rappresenta una serie di punti su una carta geografica, il peso degli archi può essere interpretato come la distanza dei punti che questi connettono.

La misura più elementare nelle reti è il grado di un nodo, definito come il numero di collegamenti che possiede:

$$k_i = \sum_{j=1}^N a_{ij}$$

Definiamo la degree distribution di una rete P_k come la frazione di nodi della rete con grado k ; quindi, se ci sono complessivamente n nodi in una rete e n_k di questi hanno grado k :

$$P_k = \frac{n_k}{n}$$

Quasi tutte le reti del mondo reale hanno una degree distribution con una coda di nodi di grado elevato.

Un altro concetto portante nella scienza delle reti è la nozione di percorso. Un percorso è una sequenza di nodi tale che ogni coppia consecutiva di essi è connessa da un collegamento nella rete. Un esempio particolarmente interessante di percorso è la passeggiata aleatoria (random walk), la formalizzazione dell'idea di prendere passi successivi in direzioni casuali, ovvero un percorso in cui ogni passo è scelto uniformemente a caso dall'insieme dei collegamenti al nodo corrente. Si può dimostrare che nel limite di tempo la probabilità di trovare

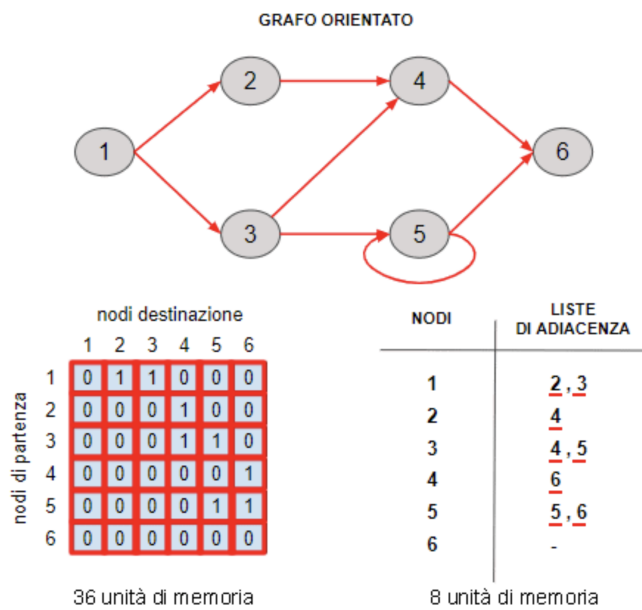


Figura 1.3: Esempio di matrice di adiacenza per un grafo orientato

un random walk al nodo i è proporzionale al suo grado k : $p_i = \frac{k_i}{2m}$, dove m è il numero di collegamenti nella rete.

Esistono altre misure della struttura delle reti che sono molto utili per comprendere il loro comportamento, come il concetto di centralità. Questo è utilizzato per determinare il nodo più importante in una rete, ovvero quel nodo le cui caratteristiche sono l'abilità di comunicare direttamente con altri nodi, la closeness centrality (vicinanza ad essi) e la sua betweenness centrality (ruolo di mediatore) tra diverse parti di una rete. In particolare, il grado di centralità semplice misura l'abilità di un nodo di comunicare direttamente con altri nodi. Il grado stesso di un nodo, k_i , può essere considerato una misura della centralità, nota come centralità di grado, in quanto i nodi con più collegamenti possono produrre effetti in grandi gruppi di nodi. Talvolta, tuttavia, un nodo può essere considerato importante non tanto per la sua importanza intrinseca, quanto perché è collegato ad altri nodi che sono a loro volta importanti. La closeness centrality di un nodo i indica quanto è vicino agli altri nodi della rete, ed è misurata in termini di cammino più breve:

$$CC(i) = \frac{n - 1}{s(i)}$$

dove la distanza totale $s(i)$ è calcolata come:

$$s(i) = \sum_{j \in V(G)} d(i, j)$$

con $d(i, j)$ la distanza tra il nodo i e il nodo j e $V(G)$ l'insieme dei nodi della rete G vicini al nodo i .

La betweenness centrality indica, invece, quanto un nodo è importante nella comunicazione tra diverse parti della rete, ed è definita come:

$$s(i) = \sum_j \sum_k \frac{p(i, j, k)}{p(j, k)}, i \neq j \neq k$$

dove $p(j, k)$ è il numero di cammini più brevi che collegano il nodo j al nodo k , mentre $p(i, j, k)$ è il numero di cammini più brevi tra i due nodi che passano anche per il nodo i . Se la rete è orientata, il termine $p(i, j, k)$ si riferisce al numero di cammini orientati dal nodo j al nodo k che passano per il nodo i , e $p(j, k)$ al numero totale di cammini orientati dal nodo j al nodo k .

Un'altra proprietà interessante delle reti è l'esistenza di scale intermedie che possono svolgere un ruolo chiave per il funzionamento del sistema a livello globale; non solo i singoli nodi sono importanti, ma è anche possibile trovare gruppi di nodi che condividono proprietà comuni. Il gruppo più semplice è quello formato da tre nodi: se tutti sono collegati reciprocamente la relazione si dice transitiva; per quantificare il livello di transitività di una rete definiamo il clustering coefficient, calcolato come la frazione di percorsi chiusi di lunghezza tre.

Allo stesso modo, potremmo essere in grado di classificare i nodi in gruppi o classi. Se i nodi di una determinata classe tendono a connettersi a nodi appartenenti alla stessa classe, diciamo che la rete è assortativa; di solito è utile misurare se l'assortatività è maggiore di quanto ci si aspetterebbe se i nodi, diversamente, fossero connessi in modo casuale.

Un'ulteriore proprietà strutturale riguarda la possibilità di identificare o meno in una rete le cosiddette comunità (o moduli), ovvero insiemi di nodi che sono più densamente connessi tra loro di quanto non lo siano i nodi che non fanno parte dello stesso modulo. Esistono diversi metodi ed euristiche per l'individuazione delle comunità; i più diffusi sono quelli basati sulla massimizzazione della modularità della rete, definita come:

$$s(i) = \frac{1}{2m} \sum_{ij} (a_{ij} - \frac{k_i k_j}{2m}) \delta(c_i, c_j)$$

dove c_i è la classe di appartenenza di i . Finora abbiamo considerato le reti come oggetto di studio isolati, i cui nodi sono collegati tra loro da un unico tipo di legame, trascurando altre possibili interazioni all'interno dell'insieme E . Tuttavia, in molti casi, questa codifica può rappresentare una semplificazione eccessiva; ad esempio, trascurare la dipendenza dal tempo distrugge l'ordine dei processi di trasmissione; allo stesso modo, ignorare la presenza di più tipi di collegamenti può introdurre errori quando i processi dinamici dipendono dal tipo di interazione. Pertanto, al fine di progredire nello studio dei sistemi complessi, dobbiamo addentrarci nell'analisi di strutture di rete più avanzate, ossia i multilayer network.

1.2.2 I multilayer network

Per rappresentare sistemi costituiti da reti con più tipologie di collegamenti e interazioni consideriamo strutture che, oltre ai nodi e ai loro collegamenti, sono suddivise in strati (layer). Nella sua forma più generale, un multilayer network ha un nodo u nel layer α che può essere collegato a qualsiasi nodo v in qualsiasi layer β . I layer rappresentano aspetti o proprietà che caratterizzano i nodi o i collegamenti che appartengono a essi. Di conseguenza, possiamo suddividere l'insieme delle connessioni in collegamenti intra-layer, tra nodi situati nello stesso layer, e collegamenti inter-layer (o di accoppiamento), tra nodi situati in layer diversi. È possibile estendere la rappresentazione della matrice di adiacenza, descritta per i single network, per i multilayer network attraverso la costruzione di una matrice a blocchi in cui ogni blocco diagonale corrisponde alla matrice di adiacenza di ogni layer, A^α , e le matrici fuori diagonale codificano l'accoppiamento, $C_{\alpha\beta}$ (Figura 1.4).

I multilayer network codificano due classi primarie di sistemi: le multiplex network (reti multiple) e le network of network (reti di reti). Le prime sono reti in cui lo stesso insieme di nodi è rappresentato in ogni layer; in ciascun layer l'interazione tra i nodi è diversa; ad esempio, due nodi possono essere connessi in un layer e non in un altro. Questo è il caso

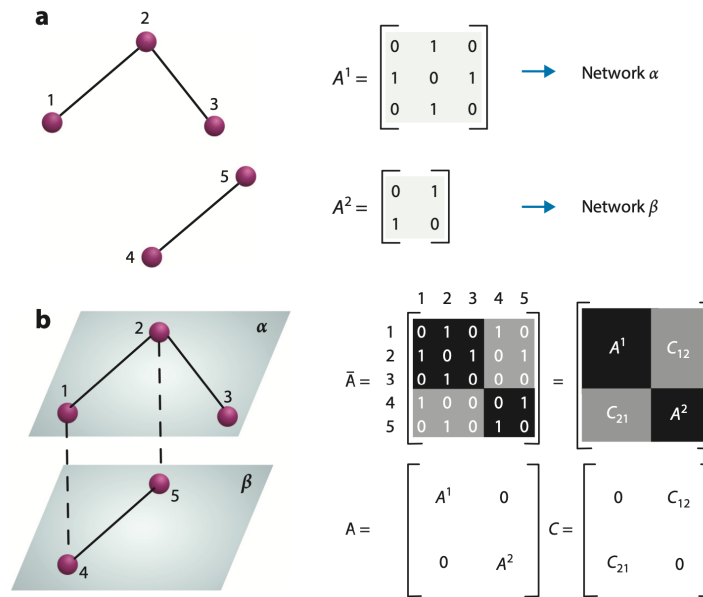


Figura 1.4: Estensione della matrice di adiacenza per un multilayer network

dei social network, in cui un dato utente può avere un account Twitter (layer 1) e un profilo Facebook (layer 2). L'insieme dei follower/amici non coincide in generale per entrambi i layer, il che porta a due matrici di adiacenza intra-layer diverse. Al contrario, un network of network è formato da reti interconnesse tra loro, ma costituite da diversi tipi di nodi. Quando si costruiscono multilayer network, spesso non è del tutto chiaro come definire ogni layer, quali siano le interazioni tra i layer e quale sia il loro numero. Si affronta, quindi, il problema della riducibilità, cioè della definizione del numero di layer che un multilayer network deve avere per rappresentare accuratamente la struttura del sistema. Ad esempio, utilizzando l'entropia di Von Neumann su un multilayer network è possibile determinare se la rappresentazione multilayer è distinguibile dalla rete aggregata; nello specifico, se l'aggregazione di due layer non comporta una diminuzione dell'entropia relativa rispetto al multiplex network in cui sono separati, allora essi dovrebbero essere mantenuti aggregati. Allo stesso modo, Kleineberg et al. nell' articolo pubblicato nel 2016 sulla rivista *Nature Physics*, *Hidden geometric correlations in real multiplex networks*, dimostrano che i multiplex network non sono solo combinazioni casuali di singoli layer della rete, ma piuttosto possiedono significative correlazioni geometriche nascoste. È, anche, possibile estendere gli strumenti sviluppati precedentemente per le reti single-layer a quelle multilayer.

Definiamo il grado del nodo i nel layer α come:

$$k_i^\alpha = \sum_j a_{ij}^\alpha$$

Di conseguenza, il grado nel multilayer network non è più uno scalare, ma il vettore $k_i = k_i^1, \dots, k_i^L$, che risulta in un grado totale o in una sovrapposizione di gradi di $o_i = \sum_\alpha k_i^\alpha$. È, quindi, possibile creare nuove misure, come l'edge overlap (sovrapposizione degli archi), $o_{ij} = \sum_\alpha a_{ij}^\alpha$, che tiene conto del numero di layer in cui esiste lo stesso collegamento. Dal momento che i nodi sono ora caratterizzati da vettori, invece che da scalari, dobbiamo sviluppare nuovi strumenti per semplificare la loro descrizione. Nello specifico, definiamo la distribuzione del grado di un nodo tra i vari layer utilizzando l'entropia del grado del multiplex come:

$$H_i = \sum_{\alpha=1}^L \frac{k_i^\alpha}{o_i} \ln \left(\frac{k_i^\alpha}{o_i} \right)$$

Questo è anche detto coefficiente di partecipazione. Esistono altre misure simili all'edge overlap che funzionano in un contesto multilayer; ne è un esempio l'interdipendenza del nodo i , definita come $\lambda = \sum_{i \neq j} \frac{\psi_{ij}}{\sigma_{ij}}$, dove σ_{ij} è il numero totale di percorsi più brevi tra i nodi i e j e ψ_{ij} è il numero di percorsi più brevi tra il nodo i e il nodo j che sono collegati in due o più layer. Questo parametro misura, quindi, la dipendenza di un nodo dal multiplex network in termini di raggiungibilità; allo stesso modo, è possibile estendere questa definizione anche ai layer per tenere conto dell'importanza di un determinato layer nell'intero sistema.

Cozzo et al. nell'articolo *Structure of triadic relations in multiplex networks* pubblicato nel 2015 sul *New Journal of Physics* hanno ampliato la nozione di relazione triadica, la relazione transitiva menzionata in precedenza, alle reti multiplex distinguendone cinque differenti possibilità: una in cui i tre collegamenti si trovano nello stesso layer, tre in cui due collegamenti si trovano in un layer e il terzo in un layer diverso e, infine, una in cui i tre collegamenti si trovano in tre layer diversi.

Allo stesso modo, per studiare l'assortatività nei multilayer network, intesa come la tendenza dei nodi ad essere collegati ad altri che si assomigliano, ci sono diverse possibilità: possiamo analizzare l'assortatività tra nodi di diversi layer, tra gli stessi layer o all'interno di ogni layer separatamente.

In merito ad altre proprietà definite per i single network, sappiamo che i random walk sono spesso utilizzati per esplorare una rete utilizzando solo informazioni locali, grazie alla loro semplicità; tuttavia, non è semplice estenderli ai multilayer network in quanto le proprietà dei nodi non sono più scalari, ma vettoriali, così come il loro grado.

Anche altre misure di centralità utilizzate nei single network possono essere estese al contesto dei multilayer network. Come hanno dimostrato De Domenico et al. nell'articolo *Ranking in interconnected multilayer networks reveals versatile nodes*, quando si calcolano misure di centralità nel contesto dei multilayer network, è possibile trovare nodi versatili che non possono essere estratti direttamente dalla rete aggregata; è consigliato utilizzare processi dinamici per caratterizzare la posizione relativa dei nodi nei multilayer network.

Come brevemente discusso in precedenza, le proprietà spettrali delle reti giocano un ruolo molto importante nella dinamica che esse rappresentano; utilizzando la teoria delle perturbazioni, ad esempio, alcuni studi hanno dimostrato che l'autovalore più grande del multiplex network è uguale a quello della matrice di adiacenza del layer dominante del sistema con un'approssimazione del primo ordine.

Analogamente, Sánchez-García et al. nell'articolo *Dimensionality reduction and spectral properties of multilayer networks* del 2014 hanno studiato gli autovalori del Laplaciano nel contesto dei multilayer network dimostrando che gli autovalori del quoziente sono intrecciati con gli autovalori della rete madre. Questo fatto ha profonde conseguenze, in quanto, ad esempio, il tempo di rilassamento di una rete multiplex è minore o uguale a quello della rete aggregata, il che, a sua volta, può portare a processi di diffusione più rapidi nelle reti multiplex rispetto alle loro controparti aggregate.

Infine, esaminando lo stato del rilevamento delle comunità nei multilayer network, Mucha et al. nell'articolo *Community Structure in Time-Dependent, Multiscale, and Multiplex Networks* del 2010 hanno generalizzato il concetto di modularità, considerato in precedenza in relazione ai multilayer network, che può essere utilizzato per rilevare le comunità non solo nelle reti statiche, ma anche in quelle dipendenti dal tempo, impostando lo stato della rete a ogni finestra temporale in un layer. Sono stati sviluppati anche metodi per rilevare comunità sovrapposte e per studiare la rilevabilità di strutture comunitarie in modelli stocastici a blocchi.

Dopo aver esteso i single network ai multilayer network, nella sezione successiva riassumiamo i risultati ottenuti per diversi processi dinamici paradigmatici svolti sui multilayer network.

1.3 Processi dinamici nei multilayer network

Questa sezione è divisa in due sottosezioni: in primo luogo, esploriamo il problema della percolation (percolazione), il cui studio è iniziato negli anni '40 e per il quale esiste un ampio quadro teorico ancora oggi molto rilevante per la sua importanza nella descrizione delle reti. Infatti, la percolation non solo può essere utilizzata per analizzare la resilienza delle reti, che rappresenta la sua applicazione più immediata, ma può anche essere di supporto nello studio di problemi in campi molto diversi, come la vaccinazione e l'immunità di gregge. Concludiamo questa sezione, nella sua seconda parte, con la descrizione dei processi di diffusione nei multilayer network, fornendo in aggiunta alcune informazioni su problemi in cui questi sono particolarmente rilevanti, come i casi di epidemia e di diffusione di opinioni.

1.3.1 Percolation e multilayer network

In fisica statistica e matematica, la teoria della percolazione descrive il comportamento di una rete quando vengono aggiunti o rimossi nodi e collegamenti. Nei single-layer network non diretti, una componente connessa è l'insieme massimo di nodi connessi tra loro attraverso un unico percorso. Nei multilayer network questa definizione può essere estesa includendo anche i collegamenti inter-layer, incrementando inevitabilmente il numero di nodi (Figura 1.5). Per questo, è importante implementare nuovi algoritmi per l'individuazione delle componenti connesse nei multilayer network.

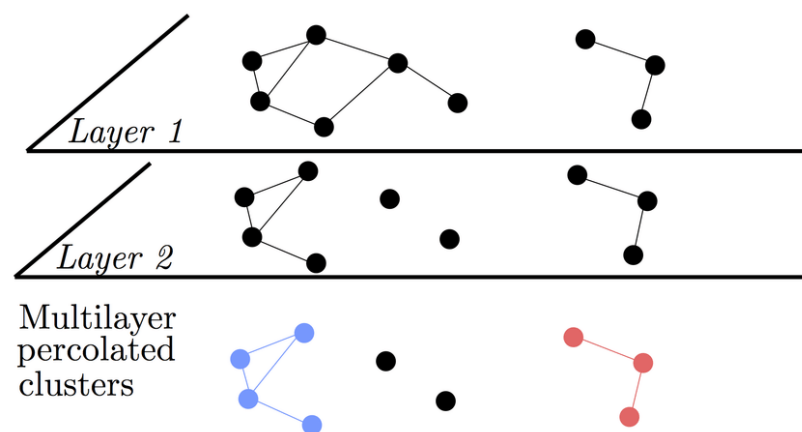


Figura 1.5: Esempio di percolation nel caso multilayer

Una delle principali applicazioni della teoria della percolazione nelle reti è lo studio della robustness (robustezza), definita come la capacità di preservare la struttura della rete quando è soggetta a guasti o attacchi, sia nei nodi sia nei collegamenti. Sebbene la robustezza delle reti a layer singolo sia stata analizzata a fondo e le domande principali abbiano già trovato risposta, nel caso dei multilayer network si presentano nuove ed entusiasmanti sfide. Ad esempio, a differenza delle reti a layer singolo in cui, in caso di rimozione casuale dei nodi, la dimensione della componente connessa più grande subisce una transizione continua, la transizione nei multilayer network è ibrida: al di sotto del punto critico si verifica una discontinuità simile a una transizione del primo ordine, ma al di sopra di esso i sistemi presentano un comportamento critico tipico di una transizione del secondo ordine.

Analogamente, un'ulteriore proprietà il cui comportamento è diverso nella transizione da single-layer a multilayer network è il clustering: nelle reti a layer singolo l'effetto del clustering sulla robustezza della rete è molto basso, mentre riduce sostanzialmente la robustezza di alcuni multilayer network. In questo contesto, Buldyrev et al. nell'articolo *Catastrophic cascade of failures in interdependent networks* pubblicato sulla rivista *Science* hanno studiato una rete interdipendente di reti elettriche e computer in cui possiamo considerare i collegamenti inter-layer come collegamenti di dipendenza. In questo contesto, per funzionare in un layer, un nodo ha bisogno del supporto di un altro nodo che si trova in un layer diverso e, quindi, dipende da un collegamento inter-layer. In questo caso, l'arresto dei generatori elettrici ha portato al guasto dei nodi della rete di comunicazione, che, a sua volta, ha causato un ulteriore danno ai generatori. Dopo aver modellato questo processo con una rete multilayer interdipendente a due layer, gli autori hanno dimostrato che queste reti possono essere meno robuste dei single network. Infatti, un nodo molto importante in un layer potrebbe essere collegato a un nodo più piccolo e più fragile nell'altro e, se quest'ultimo scompare, si avrà un impatto radicale sul primo layer, poiché spegnerà il nodo importante.

Un'estensione di questo modello è stata affrontata da Min et al. nel loro studio *Network robustness of multiplex networks with interlayer degree correlations* analizzando il problema che si verifica quando il guasto di un nodo in un layer non causa direttamente un guasto in un nodo dell'altro layer, ma invece esaurisce le risorse che il nodo ottiene; se questo valore scende al di sotto di una certa soglia, il nodo fallisce. Con questo modello, gli autori hanno individuato che il sistema presenta un'isteresi che produrrebbe un aumento del costo del processo di recupero. Questi processi in reti interdipendenti sono noti anche come processi a cascata; infatti, il malfunzionamento di un nodo del layer *A* provocherà l'arresto di alcuni nodi del layer *B* che, a sua volta, produrrà ulteriori malfunzionamenti nel layer *A*. Questa ricorsione continuerà iterativamente fino a quando tutti i nodi colpiti non saranno rimossi dalla rete.

Le reti interdipendenti non sono le uniche ad essere state affrontate; altri studi hanno esplorato la robustezza dei multiplex network quando i loro nodi vengono rimossi, sia per guasti casuali che per attacchi mirati.

Riguardo la determinazione degli effetti che ogni proprietà della rete ha sulla sua robustezza, Min et al. nella pubblicazione *Large deviation theory of percolation on multiplex networks* hanno analizzato il problema della robustezza della rete in presenza di correlazioni di grado tra i layer e hanno affrontato il problema nel contesto dei layer della rete commerciale internazionale.

1.3.2 Processi di diffusione nei multilayer network

I processi di diffusione sono stati a lungo studiati nel contesto delle reti ed è stato dimostrato che i processi stocastici, come i giochi evolutivi o la diffusione epidemica, presentino comportamenti molto diversi quando vengono applicati a popolazioni strutturate in rete rispetto a popolazioni che non presentano alcuna struttura. Ne è un esempio la soglia epidemica, grandezza chiave nell'epidemiologia matematica, che svanisce in alcune configurazioni di rete.

Pertanto, è indispensabile conoscere la struttura sottostante della popolazione in cui si svolge il processo e, poiché la struttura dei multilayer network codifica nuove interdipendenze strutturali, è necessario studiare quali nuovi effetti inducono nel comportamento dinamico del sistema.

I due principali processi di diffusione studiati nei multilayer network sono la diffusione epidemica e la diffusione dell'informazione, per le quali talvolta si utilizzano le stesse equazioni, perché in alcuni casi sono molto simili.

Come esempio di dinamica delle opinioni, lo studio di Amato et al. *Opinion competition dynamics on multiplex networks* del 2017 ha esaminato la competizione delle opinioni in una

rete a due layer permettendo allo stesso individuo di avere un'opinione diversa in ogni layer, rappresentando, così, la possibilità che gli individui abbiano opinioni diverse a seconda del contesto (i layer rappresentano scenari diversi).

Come già detto in precedenza, i processi di tipo epidemico sono stati utilizzati anche per studiare la diffusione delle malattie, per le quali l'aggiunta di più layer al sistema può avere effetti drammatici sull'esito di un'epidemia. Ad esempio, Buono et al. nel paper del 2014 *Epidemics in Partially Overlapped Multiplex Networks* hanno dimostrato che con una sovrapposizione parziale, cioè con solo una frazione dei nodi di un layer che si trova anche nell'altro layer, sia la soglia epidemica sia la frazione totale di individui infetti cambiano significativamente anche per bassi valori di sovrapposizione.

1.4 Esempi di utilizzo dei multilayer network

In questa sezione forniamo un breve resoconto di alcune applicazioni in diversi settori in cui i multilayer network sono stati recentemente utilizzati.

1.4.1 Ecologia

I sistemi naturali presentano tipicamente più tipi di interazioni; ad esempio una stessa specie vegetale interagisce sia con gli impollinatori sia con gli erbivori; pertanto potrebbe essere molto efficace separare queste interazioni e organizzarle in multilayer network.

Infatti, ci sono alcune modalità di stratificazione che possono fornire una visione profonda del comportamento e dell'evoluzione dei sistemi ecologici. In particolare, i layer possono rappresentare lo stesso sistema in punti diversi dello spazio o del tempo, il che è utile per studiare l'evoluzione a lungo termine dei sistemi ecologici; oppure, se concentriamo la nostra attenzione sui legami, possiamo definire ogni layer in base al tipo di interazione. Al contrario, se siamo più interessati all'organizzazione dei nodi, possiamo codificare i layer in base al livello di organizzazione all'interno del sistema.

Per esempio, Finn et al. nel loro studio *Can Multilayer Networks Advance Animal Behavior Research?* risalente al giugno del 2018 hanno utilizzato quest'ultimo approccio per esaminare il ruolo degli individui in una società di babbuini, costruendo un multilayer network composto da due layer il primo rappresenta il grooming (ovvero la pratica di spulciare la pelliccia di altri esemplari), mentre l'altro rappresenta l'associazione (cioè la loro vicinanza). Successivamente gli autori hanno studiato la centralità dei babbuini utilizzando questa rappresentazione e confrontandola con i risultati ottenuti per la rete aggregata (a layer singolo); è interessante notare che l'individuo più centrale nella rappresentazione multilayer non è così centrale in quella aggregata e l'individuo più centrale non è affatto centrale nei singoli layer. Gli autori hanno concluso che, quando si studiano relazioni che dipendono da più tipi di interazione, potrebbe essere necessario utilizzare la struttura multilayer per catturare i ruoli sociali degli individui.

1.4.2 Biologia

Gosak et al. nell'articolo *Network science of biological systems at different scales: A review* del 2018 hanno riassunto lo stato della scienza delle reti nello studio dei sistemi biologici a diverse scale in cui, oltre a fornire una panoramica sull'uso delle reti a layer singolo nei sistemi biologici, discutono anche le possibilità offerte dai multilayer network. Per esempio, si può costruire un multilayer network di interazione proteina-proteina in cui ogni layer rappresenta lo stadio di vita di un batterio. Questa ricerca non è più limitata a sistemi piccoli e semplici; ne è un esempio la ricerca condotta da Zitnik et al. *Predicting multicellular function*

through multi-layer tissue networks; in essa gli autori hanno costruito un multilayer network di interazioni molecolari in cui ogni layer rappresenta un diverso tessuto umano; imponendo una gerarchia tissutale, essi sono riusciti a migliorare il potere predittivo di questa descrizione rispetto al caso a layer singolo.

1.4.3 Trasporti

I sistemi di trasporto sono un esempio immediato in cui la struttura multilayer si presenta in modo naturale, in quanto possono esistere diverse modalità di trasporto tra due determinati luoghi. Queste modalità possono avere proprietà molto diverse, come la velocità o la capacità di trasporto; è, quindi, importante essere in grado di distinguere ciascuna di esse nello studio dell'intero sistema. Esistono diversi modi per modellare i sistemi di trasporto come multilayer network, a seconda dell'obiettivo dello studio; ad esempio, a livello di Paese, si può modellare ogni modalità di trasporto, cioè pullman, treni, voli, etc. come layer e le città come nodi.

Un approccio simile e meno complesso è relativo al trasporto urbano (Figura 1.6), in cui i nodi possono essere luoghi della città e ogni layer può rappresentare una modalità di trasporto, come autobus, metropolitana, ferrovia, etc. Tuttavia, in quest'ultimo tipo di sistema, è rilevante non solo analizzare l'interazione tra le diverse modalità, ma anche tra le linee di ciascuna modalità. Infatti, per considerare correttamente il tempo di trasferimento da una linea metropolitana all'altra, ad esempio, è necessario considerare ogni linea come un singolo layer. Successivamente, i layer di ciascuna modalità di trasporto possono essere raggruppati in un'entità di livello superiore, i superlayer, a seconda delle proprietà che si è interessati a svelare. Inoltre, poiché i trasporti urbani sono incorporati nella città, potrebbe essere interessante aggiungere anche il modello di strada come layer aggiuntivo, al fine di fornire nuovi spunti di riflessione sulla pianificazione urbana.

Analogamente, la metodologia multilayer è stata applicata anche allo studio delle reti di volo. Ne è un esempio la realizzazione di un multilayer network del trasporto aereo europeo in cui ogni layer rappresenta una compagnia aerea. In questo modo, è stato dimostrato che la resilienza del multilayer network è inferiore a quella della rete aggregata. Altre reti di trasporto aereo che sono state analizzate utilizzando il framework multilayer includono la rete dell'aviazione greca e la rete dell'aviazione cinese.

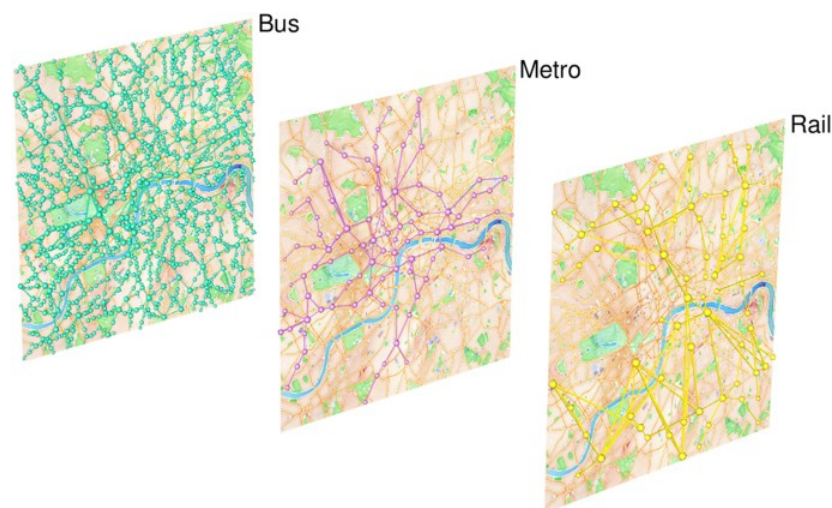


Figura 1.6: Multilayer network applicato al trasporto urbano

1.4.4 Cervello umano

Possiamo distinguere tre scale nelle reti cerebrali: la scala spaziale, che si riferisce alla granularità con cui è definita la rete, la scala temporale, che va dai submillisecondi all'intera durata della vita, e la scala topologica, che va dai singoli nodi alla rete nel suo complesso. La maggior parte delle analisi della rete cerebrale umana effettuate finora fissano queste tre caratteristiche a un determinato valore, tuttavia è ora possibile affrontare problemi come la connettività variabile nel tempo del cervello utilizzando multilayer network. Inoltre, se si affrontano gli stessi problemi, come l'identificazione dei nodi centrali, in reti cerebrali a layer singolo e nei multilayer network, i risultati sono molto diversi. Analogamente, Battiston et al. nello studio *The new challenges of multiplex networks: Measures and models* del marzo del 2017 hanno studiato le strutture dei motivi in un multilayer network composto dalle reti anatomiche e funzionali del cervello e hanno trovato motivi anatomofunzionali congiunti che differiscono da quelli trovati nelle reti a layer singolo.

1.4.5 Economia

La formulazione multilayer consente nuovi approcci allo studio dell'economia, come l'applicazione di un approccio multilayer per studiare la struttura dei mercati finanziari costruendo un multilayer network in cui ogni layer rappresenta gli stessi dati, ma i collegamenti sono costruiti utilizzando diverse misure di correlazione: Pearson, Kendall, Coda e Parziale.

1.4.6 Teoria dei giochi

È il modello matematico per lo studio delle "situazioni competitive", in cui, cioè, sono presenti più persone (o gruppi di persone, o organizzazioni) dette appunto "giocatori", con autonoma capacità di decisione e con interessi contrastanti. In questo contesto, poiché i multilayer network possono rendere conto di diversi contesti sociali allo stesso tempo, stanno diventando molto utilizzati nella comprensione della cooperazione umana. Ne è un esempio importante la reciprocità della rete interdipendente, che può mantenere livelli sani di cooperazione anche in condizioni estremamente avverse. Tuttavia, questo campo è ancora agli albori ed è importante non essere troppo ottimisti con questi risultati; questi autori hanno, infatti, dimostrato che l'interdipendenza della rete può effettivamente promuovere la cooperazione oltre i limiti imposti dalle reti isolate, ma solo nel caso in cui il multilayer network soddisfi alcune condizioni specifiche. Pertanto, il coinvolgimento in diversi contesti sociali da solo non è sufficiente a promuovere la cooperazione.

Le reti sociali modellano le nostre relazioni con amici e parenti, le nostre connessioni con gruppi e organizzazioni e la struttura sociale del nostro stesso quartiere, stato e nazione. I ricercatori studiano da tempo questo tipo di reti attraverso lo strumento della Social Network Analysis (SNA). In questo capitolo affronteremo lo studio della SNA attraverso i multimodal network, di cui analizzeremo proprietà, metriche e implementazioni.

2.1 Introduzione alla Social Network Analysis

La Social Network Analysis (SNA) indica una famiglia di metodi utilizzati per analizzare le modalità di relazione e di comunicazione tra individui, attraverso la teoria dei grafi.

Questo strumento è diventato sempre più rilevante grazie alle moderne tecnologie informatiche, le quali consentono agli esseri umani di connettersi e di relazionarsi in modi del tutto nuovi e facilmente osservabili. Si tratta di una metodologia estremamente versatile, che precede Twitter e Facebook di almeno 30 anni e che è stata poco studiata e scarsamente considerata fino all'avvento degli online social network; questi ultimi raggruppano centinaia di milioni di utenti connessi attraverso varie modalità di relazione, tanto che, a causa della loro ampiezza e complessità, esse richiedono un supporto software sempre maggiore per il calcolo, la statistica e il processo decisionale.

La SNA si basa sul concetto per il quale le nostre relazioni, prese nel loro insieme, definiscono chi siamo e come agiamo; pertanto, osservando e studiando tali relazioni, è possibile dare una risposta a molte domande sul nostro comportamento sociale. In un contesto interpersonale, una relazione può essere rappresentata da amicizia, influenza, affezione, fiducia o, al contrario, dissomiglianza e conflitto.

Le relazioni possono essere binarie o con valori; ad esempio, "Marco segue Giorgia su Twitter" è una relazione binaria, mentre "Marco ha risposto a 4 tweet di Giorgia" è una relazione con valore.

È possibile notare che alcune relazioni (ad esempio, la relazione insegnante-studente) sono asimmetriche per natura, così come il following su Twitter e Instagram, i quali non richiedono una mutua conferma. Altre relazioni, invece, sono simmetriche, come le amicizie su Facebook e le connessioni su LinkedIn, in quanto richiedono una mutua conferma; in questi casi, il software forza una simmetria anche quando la relazione umana corrispondente è asimmetrica.

Un social network può essere visto come una collezione di frasi che descrivono relazioni; ad esempio, «Alice – Ama – Bob». Una frase di questo tipo rappresenta l'unità di base della

SNA ed è chiamata diade; ciascuna diade rappresenta una relazione, ovvero un arco nella teoria dei grafi, dove i nodi (o vertici) collegati rappresentano le persone coinvolte nella relazione.

Nella SNA i nodi hanno un tipo (o modalità), in quanto possono rappresentare persone, commenti, post, hashtag, etc.

Se una rete (o un grafo) contiene nodi di un solo tipo è chiamata 1-mode network, se contiene nodi di due tipi essa è bimodale ed è chiamata 2-mode network. In molti contesti del mondo reale, però, le reti sono costituite non solo da uno, ma da diversi tipi di nodi; in questi contesti si parla di multimodal network (reti multimodali). Ne sono un esempio i co-authorship network, i quali contengono gli autori di articoli, insieme alle sedi che frequentano e alle riviste in cui pubblicano.

2.2 Verso i multimodal social network

2.2.1 1-mode network: definizione e proprietà

Introduciamo, ora, la definizione, le proprietà e le metriche di valutazione dei grafi che descrivono le reti unimodali, chiamati 1-mode graph. Queste rappresentano lo strumento di base per la Social Network Analysis su network costituiti da un solo tipo di nodi e possono essere estesi facilmente per costruire i 2-mode e i multimodal network.

Come già accennato in precedenza, un 1-mode graph (grafo unimodale) è un grafo in cui tutti i nodi appartengono a un solo tipo, rappresentando, quindi, la stessa categoria di entità. In questo contesto, stiamo assumendo, per semplicità, che tutti gli archi della rete utilizzano un singolo verbo consistente; questa assunzione, molto limitativa nella pratica, può funzionare agevolmente negli online social network, dove la semantica di una relazione è definita fortemente dagli sviluppatori. Ne è un esempio un grafo che rappresenta una rete di amicizie tra persone, dove i nodi rappresentano le persone e gli archi rappresentano le connessioni di amicizia tra di loro.

Il modo più comune per fornire una rappresentazione matematica di un online social network è tramite una matrice di adiacenza (Figura 2.1), di cui abbiamo parlato nel Capitolo 1.

	A	B	C	D	E
A	0	2	0	5	5
B	2	0	0	1	0
C	0	0	0	3	4
D	5	1	3	0	0
E	5	0	4	0	0

Figura 2.1: Esempio di matrice di adiacenza pesata

Il principale problema delle matrici di adiacenza è che gli 0 richiedono la stessa quantità di memoria delle altre celle, portando, così, a un grande spreco di risorse; infatti, la maggior parte degli online social network hanno una densità, calcolata come il rapporto tra celle non-zero e celle zero, dello 0.1%.

Per sopperire a tale limitazione, si predilige implementare una visualizzazione attraverso le liste di adiacenza (Figura 2.2).

A	B	2
A	D	5
A	E	5
B	A	2
B	D	1
C	D	3
C	E	4
D	A	5
D	B	1
D	C	3
E	A	5
E	C	4

Figura 2.2: Esempio di lista di adiacenza

Nonostante tale implementazione richieda un'occupazione di memoria molto inferiore alle matrici di adiacenza, il suo svantaggio è quello di consentire unicamente l'iterazione sugli archi, non permettendo, quindi, di effettuare ricerche o attraversamenti veloci del grafo.

Un ulteriore miglioramento è rappresentato dal formato: $A(B, 2), (D, 5), (E, 5)$. In questo caso, la ricerca, l'aggiunta e la rimozione di nodi sono veloci; tuttavia, questa struttura è difficile da elaborare attraverso i database.

La metrica più significativa per un grafo è la sua dimensione (size), pari al numero dei suoi nodi.

Se il grafo è indiretto, il grado (degree) di un nodo v_i è pari al numero di archi che lo connettono agli altri nodi del grafo; se il grafo è diretto, invece, il numero degli archi entranti in v_i rappresenta il suo indegree, mentre il numero di archi uscenti da v_i rappresenta il suo outdegree.

In aggiunta, se il grafo è indiretto, possiamo anche definire il suo mean degree, inteso come il rapporto tra la somma dei gradi dei suoi nodi e il numero di nodi che esso contiene; in modo analogo, se il grafo è diretto, possiamo definire il mean indegree e il mean outdegree.

Se il grafo non è pesato, definiamo la sua densità come il rapporto degli archi esistenti rispetto al numero dei suoi possibili archi; se il grafo è pesato, invece, la densità è definita come il rapporto tra la somma dei pesi degli archi esistenti e il numero degli archi possibili.

Per crawling (o attraversamento) di un grafo si intende a un algoritmo che, partendo da un punto iniziale A , percorrendo i collegamenti tra i nodi, raggiunge una destinazione finale B ; il crawling viene affrontato con il fine di percorrere il percorso più breve o di attraversare l'intero grafo per comprendere o campionare la sua struttura.

I principali algoritmi di crawling sono:

- *Depth-First Search (DFS)*: è un algoritmo di ricerca in profondità che attraversa ordinatamente tutti i nodi fino a quando non raggiunge il suo obiettivo. Nel contesto di un albero, DFS privilegia un attraversamento in profondità piuttosto che in ampiezza, scendendo, cioè, iterativamente verso il figlio di un figlio, per poi, tramite backtracking, passare ai fratelli. Tuttavia, in un grafo non ci sono figli o fratelli, ma solo vicini; in questo caso, l'algoritmo può produrre un albero a partire dai nodi del grafo che ha visitato.
- *Breadth-First Search (BFS)*: si tratta di un algoritmo di ricerca in ampiezza, opposto a DFS; infatti, esso consiste, a partire dalla radice dell'albero (o da qualche nodo arbitrario di un grafo, a volte indicato come "chiave di ricerca"), nell'esplorare i nodi vicini prima di passare ai vicini di livello successivo. La differenza tra BFS e DFS è mostrata in Figura 2.3.

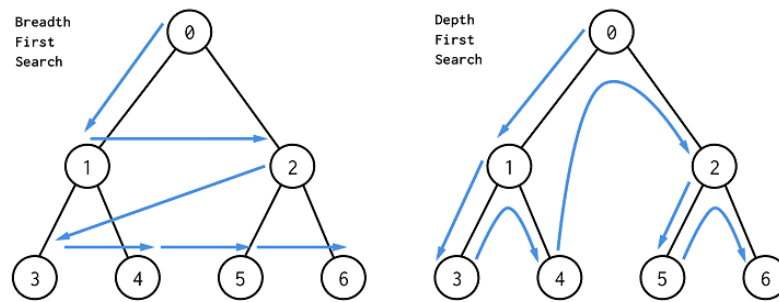


Figura 2.3: Differenza tra BFS e DFS

- *Walk (cammino)*: è un percorso che passa per una sequenza alternata di nodi ed archi che li connettono; se il nodo di partenza A e quello destinazione B sono diversi il cammino si dice aperto, altrimenti è chiuso. Se il grafo non è pesato, la lunghezza di un walk è pari al numero di archi che esso attraversa; se, invece, il grafo è pesato, la lunghezza di un walk è data dalla somma dei pesi degli archi percorsi. Un walk aperto e semplice (dove nessun nodo è attraversato due volte) è detto path, mentre si definisce ciclo un walk semplice e chiuso; un path può avere lunghezza 0, un ciclo no. Inoltre, un walk che attraversa ogni arco del grafo al più una volta prende il nome di trail, mentre un tour è un walk chiuso che comprende ciascun arco del grafo almeno una volta. Dati due nodi v_i e v_j , si definisce loro distanza geodesica come la lunghezza dello shortest path da v_i a v_j . Inoltre, l'eccentricità di un nodo v_i è pari alla massima distanza geodesica tra v_i e ogni altro nodo del grafo. Il raggio di un grafo è uguale alla minima eccentricità di tutti i suoi nodi, mentre il diametro è dato dalla massima eccentricità degli stessi. Di seguito un esempio di cammino minimo tra i nodi A e F (Figura 2.4):

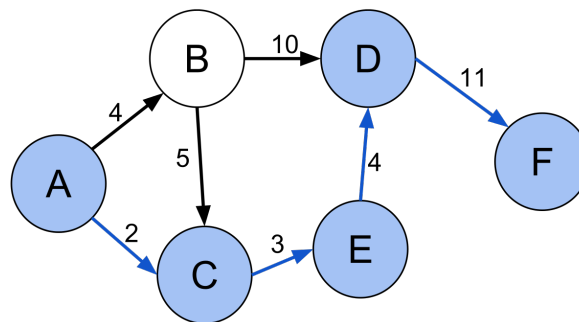


Figura 2.4: Cammino minimo tra A e F

- *Algoritmo Dijkstra*: è l'algoritmo di ricerca dei percorsi dei grafi; pubblicato da Edsger Dijkstra nel 1959, trova il percorso con costo minimo verso tutti gli altri nodi, calcolato sommando i pesi degli archi. Nei grafi dove i pesi degli archi corrispondono alla distanza (nei grafi non pesati si assume che tutti gli archi abbiano peso 1) il path trovato è il più breve.

La distanza in un grafo rappresenta un'astrazione del concetto di walk, definito il precedenza. La distanza può essere misurata nei seguenti modi:

- *Distanza euclidea*: ciascun nodo è trattato come un punto nello spazio euclideo e la distanza tra due punti in questo spazio è proporzionale al numero di vicini comuni condivisi tra i due nodi.

- *Shortest path*: corrisponde, semplicemente, al numero di archi per percorrere il grafo (necessariamente non pesato) da un punto A a un punto B .
- *Cost-based shortest path*: applicabile solo a grafi pesati, questo metodo calcola la distanza tra A a B sommando le distanze tra i punti intermedi. Il path più breve, in questo caso, è quello con la distanza minore e non necessariamente quello con meno nodi.

La distanza nei social network è un parametro fondamentale per misurare l'influenza di un utente, in quanto ci dà un'indicazione di come e quanto è probabile che si propaghi l'informazione. In un grafo non pesato la distanza è proporzionale all'abilità dell'informazione di viaggiare da Alice a Bob: maggiore è la distanza e meno probabile è che Bob possa ascoltare qualcosa detta da Alice. In un grafo pesato, invece, la distanza ci dice quanto velocemente l'informazione può diffondersi da Alice a Bob, in quanto disponiamo della frequenza della loro interazione.

Definiamo il diametro di un grafo il massimo numero di nodi che devono essere attraversati per andare da un vertice ad un altro, senza, però, considerare i cicli.

Le reti che dominano gli online social network sono gli small world network; questi sono grafi in cui la maggior parte dei nodi non sono vicini tra loro, ma la maggior parte dei nodi può essere raggiunta da ogni altro con un piccolo numero di collegamenti; in altre parole, essi mantengono una struttura con vicinato locale ma consentono a un piccolo numero di archi di andare lontano (Figura 2.5).

In particolare, si definisce small world network una rete in cui la distanza tipica (il numero di collegamenti) L tra due nodi scelti a caso cresce proporzionalmente al logaritmo del numero di nodi della rete. Ne sono esempi le mappe stradali, le catene alimentari, le reti elettriche, le reti di neuroni cerebrali, le reti di elettori, i grafi di chiamate telefoniche e le reti di regolazione genica.

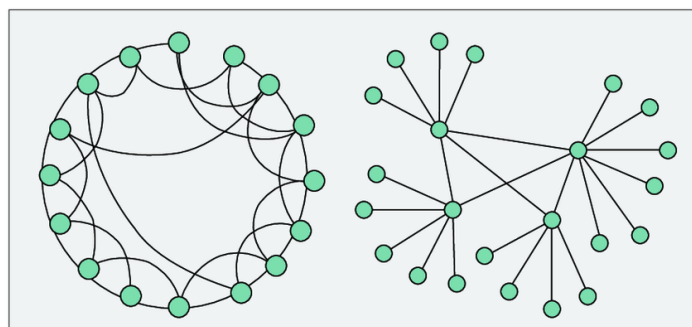


Figura 2.5: Esempio di small worlds network

2.2.2 2-mode network: definizione e applicazioni

Nonostante l'indagine sulle reti sociali si è occupata in gran parte di reti sociali unimodali (ad esempio, relazioni di amicizia tra amici e relazioni di co-autorialità tra autori), nel tempo c'è stata una crescente necessità di espandere la formulazione per adattare le relazioni tra più tipi di nodi.

Tale espansione consente di esaminare un'ampia varietà di questioni che vanno oltre quelle del contesto delle reti unimodali. Per esempio, le dinamiche che coinvolgono l'amicizia possono essere spiegate meglio considerando il modo in cui le persone partecipano a eventi o si affiliavano a organizzazioni sociali.

La co-autorialità tra gli autori può essere meglio compresa se si considerano i progetti a cui lavorano e le istituzioni in cui sono impiegati. Questi ulteriori tipi di nodi possono avere effetti sostanziali sulla formazione e sullo scioglimento dei legami sociali e, quindi, ci permettono di studiare interazioni sociali più complesse tra più tipi di entità.

La configurazione più semplice dei multimodal network (reti multimodali), ovvero reti con più di un tipo di nodi, è quella dei 2-mode network (reti a due modalità). Infatti, gran parte dei dati disponibili attualmente si trova in formato 2-mode (o bimodale, o bipartito), secondo il quale ci sono due diversi tipi (o modalità) di nodi e i link determinano le relazioni tra un insieme di nodi e l'altro; basti pensare agli online social network, in cui gli utenti (tipo 1) si collegano a pagine o gruppi (tipo 2).

Di seguito un esempio di un 2-mode network in cui i due tipi sono rappresentati dalle persone (in rosso) e dai gruppi di cui fanno parte (in verde) (Figura 2.6):

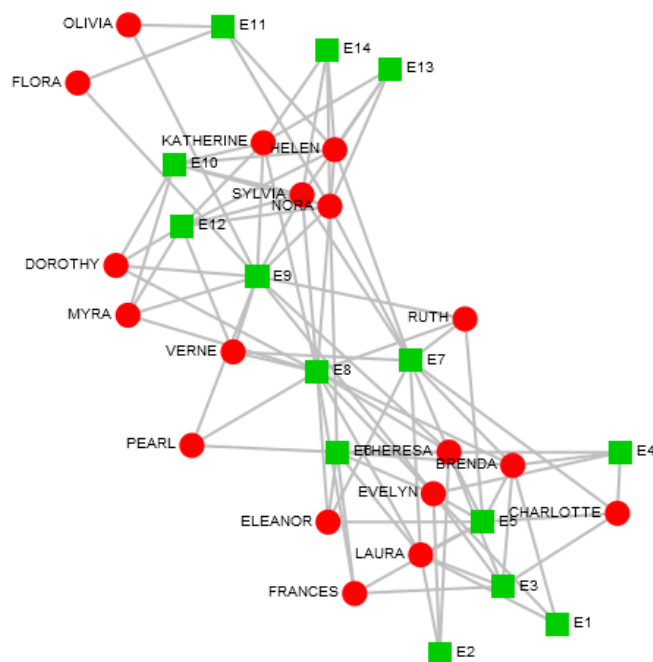


Figura 2.6: Esempio di 2-mode network

L'applicazione più rilevante di queste strutture sono le reti di affiliazione; si tratta di reti di similarità o di correlazioni tra entità appartenenti a più gruppi.

La prima applicazione di questo tipo di reti si trova in ambito sociale, dove il concetto sociologico di dualità tra persone e gruppi afferma che le idee, le attitudini e le connessioni sociali delle persone sono plasmate dall'appartenenza a gruppi e che questi ultimi sono modellati, a loro volta, dalle attitudini dei loro membri. Ne sono un esempio i gruppi nei social network, i partiti politici, i club sociali, etc.

Analizziamo, ad esempio, nella Figura 2.7, il 2-mode network che descrive la relazione tra medici A, \dots, F e i pazienti $1, 2, 3$ (Figura 2.7). La matrice di adiacenza risultante rappresenta la rete di affiliazione dei nodi A, \dots, F ; l'operazione che abbiamo effettuato non è altro che la moltiplicazione tra matrici. In particolare, abbiamo moltiplicato la matrice di adiacenza B per la sua trasposta: $A = B * B^T$.

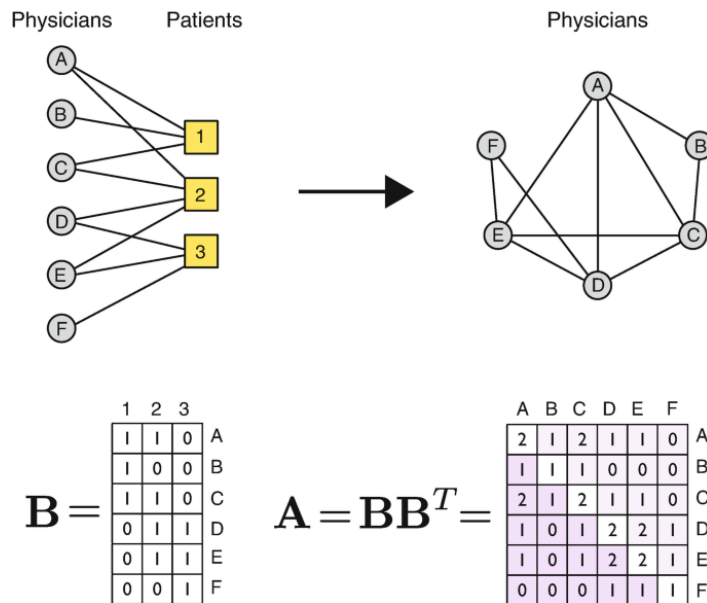


Figura 2.7: Esempio di un 2-mode network che descrive la relazione tra medici e pazienti, sia in forma grafica che con matrice di adiacenza

Un'altra applicazione dell'analisi delle reti bimodali è relativa al concetto dell'omofilia, ovvero della tendenza degli individui di associarsi, di creare legami e di comunicare con altri considerati simili per alcune peculiari caratteristiche.

La presenza di omofilia è stata descritta in una vasta gamma di studi, i quali hanno osservato il fenomeno in vari contesti e modalità e hanno stabilito una connessione di somiglianza in differenti ambiti, tra cui l'età, l'identità di genere, la classe sociale e il ruolo organizzativo.

Un esempio di applicazione dell'omofilia è la funzione "suggerisci un amico", implementata dagli online social network; ciò è possibile trattando ciascuna informazione (tag, keyword, etc.) come un nodo in una 2-mode network, calcolando da essa una rete di affiliazione persona-persona e applicando il metodo delle isole o del clustering per trovare raggruppamenti potenziali di persone. Infine, per suggerire gli amici, si prendono i link principali nella rete delle affiliazioni.

2.3 Multimodal network

Finora abbiamo considerato reti formate unicamente da uno o, al più, da due tipi di nodi. Tuttavia, le reti sociali possono contenere più di due tipi di nodi; in questo contesto, nasce la necessità di utilizzare i multimodal network, ovvero reti multimodali.

Per queste reti il modo di procedere rimane lo stesso, poichè la tecnica discussa in precedenza può essere facilmente estesa a qualunque modello di dati orientato alle reti; tuttavia, si aggiunge una maggior complessità nella loro gestione ed analisi.

Per fornire un esempio della potenza delle reti multimodali, consideriamo un diagramma che mostra le relazioni tra quattro entità distinte che rappresentano, rispettivamente, le persone, le risorse, le skill e i task di un'azienda (Figura 2.8); un certo numero di persone stanno lavorando per questa compagnia, con un qualche tipo di catena di comando, e queste persone hanno amicizie dentro e fuori l'azienda, hanno un'educazione formale in qualche campo della conoscenza e possiedono determinate risorse.

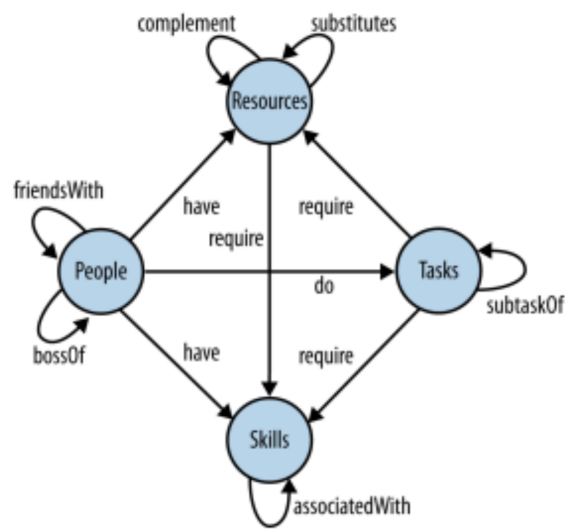


Figura 2.8: Esempio di multimodal network

Rappresentiamo, ora, in formato matriciale, il modello organizzativo attraverso le relazioni tra le entità (Figura 2.9).

	People	Skills	Resources	Tasks
People	PP: chi conosce chi?	PS: chi conosce cosa?	PR: chi ha cosa?	PT: chi fa cosa?
Skills		SS: quali skill sono affini?	SR: quali skill sono necessarie per la risorsa?	ST: quali skill sono necessarie per il task?
Resources			RR: quali risorse sono affini?	RT: quali risorse sono necessarie per il task?
Tasks				TT: quale task ha la precedenza?

Figura 2.9: Multimodal network in formato matriciale

Osserviamo come sia possibile effettuare delle inferenze a lungo respiro su questi dati. Per esempio:

- $PP * PP = PP$: "chi sono gli amici degli amici?";
- $PT * PTT = PT * TP = PP$: "chi sta lavorando insieme, e su quanti task?";
- $PT * TT = PT$: "per ciascuna persona, quanti task ci sono nei loro path critici?";
- $TT * TT = TT$: "quanti task interdipendenti?";
- $PT * TT * TP = PP$: "quante persone ci sono nel critical path di una persona?";
- $PT * TT * TT * TP = PP$: "chi sta lavorando insieme su task concorrenti?".

Infine, esistono diverse misure rappresentative per le reti multimodali:

- Misure a livello di singolo nodo:
 - *Centralità*: quali nodi sono centrali in termini di legami tra i tipi?
 - *Posizioni e ruoli*: quali nodi occupano posizioni e ruoli simili?
 - *Attributi*: quali attributi nodali hanno un impatto sui legami tra le modalità?
- Misure a livello di rete globale:

- *Densità*: quanto è densa la rete, tra i tipi e all'interno di essi?
- *Centralizzazione*: quanto è centralizzata la rete nei legami tra i tipi?
- *Sottogruppi*: ci sono sottostrutture visibili nella rete?

L'interpretazione di queste misure varia a seconda della fonte dei dati multimodali; in particolare, quando le reti multimodali sono derivate da dati tabellari, i nodi sono di diversi livelli o entità e i tipi di collegamenti dipendono dalle modalità esaminate, richiedendo, quindi, un'applicazione delle misure specifica per il contesto.

2.4 Applicazioni reali nella Social Network Analysis

Definiamo Multimodal Social Network Analysis la SNA in presenza di reti sociali multimodali (Figura 2.10).

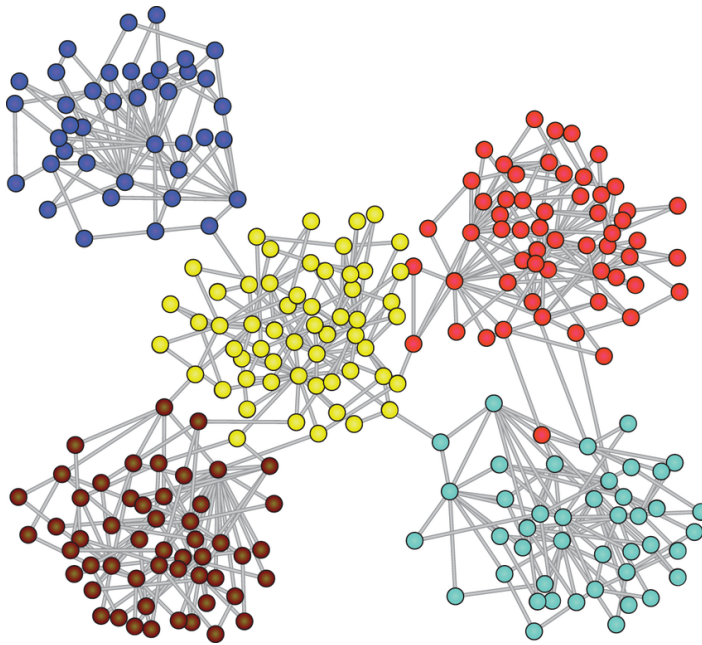


Figura 2.10: Visualizzazione grafica di un multimodal network composto da 5 tipi di nodi

Come già accennato, uno dei problemi principali della Social Network Analysis riguarda lo sforzo computazionale necessario per effettuare le analisi.

La soluzione a tale limitazione consiste nell'adozione del campionamento, il quale consiste nell'estrazione di conoscenza su una rete analizzandone soltanto una porzione. Al fine di estrarre conoscenza senza errori, è necessario affrontare i problemi della scelta degli approcci di campionamento da utilizzare, di quanto il grafo campionato debba essere più piccolo di quello originale e di come riuscire a scalare le misure del campione per l'intera rete.

2.4.1 Misure di centralità

In un social network, la prima applicazione delle reti multimodali consiste nella misurazione del potere o dell'influenza dei suoi componenti, nonostante in letteratura non ci sia unanimità né sui significati di "potere" e "influenza", né su quali tipologie di collegamenti si debbano considerare. Esistono, pertanto, diverse soluzioni, conosciute come "misure di centralità".

Degree centrality

La prima e la più semplice metrica di centralità è la degree centrality, finalizzata a individuare le celebrità, ovvero quelle persone che sono significativamente più popolari di altre; queste sono, generalmente, molto poche e di ordini di grandezza più popolari delle altre.

La degree centrality è calcolata, quindi, come il numero di connessioni che un nodo (individuo) ha, sia tra i nodi del suo stesso tipo, sia tra nodi di altri tipi; ad esempio, potremmo considerare il numero di follower di un utente, la sua appartenenza a gruppi online, il numero di interazioni sotto i suoi post, etc.

Per esempio, dato un nodo A in una posizione rilevante, ovvero in cui tutte le linee di comunicazione passano attraverso di esso, allora esso avrà una degree centrality molto elevata. Potremmo, quindi, concludere che A è forte per la natura della sua posizione; per verificare questa affermazione, però, è necessario studiare la natura delle sue relazioni.

Secondo alcune teorie, la forza di una persona in un'organizzazione è inversamente proporzionale al numero di chiavi del suo portachiavi; un bidello ha le chiavi di tutti gli uffici ma nessun potere; al contrario, un CEO non ha bisogno di chiavi, in quanto sono le persone che aprono la porta per lui. È anche vero, comunque, che un bidello (a differenza del CEO) sa ciò che avviene quotidianamente e realmente nell'organizzazione.

Closeness centrality

Una seconda misura di centralità è la closeness centrality, che rappresenta un modo per individuare quei nodi in grado di diffondere informazioni in modo molto efficiente, attraverso un grafo mono o multimodale.

La closeness centrality di un nodo misura, nell'ambito della SNA, la distanza media di un individuo da tutti gli altri nodi, sia del suo stesso tipo, sia di altri tipi; i nodi con un punteggio di closeness elevato hanno le distanze più brevi da tutti gli altri nodi.

Riguardo il suo funzionamento, l'algoritmo calcola, per ogni nodo u , la somma delle sue distanze da tutti gli altri nodi, basandosi sul calcolo dei percorsi più brevi tra tutte le coppie di nodi; la somma risultante viene poi invertita per determinare il punteggio di centralità della vicinanza per quel nodo.

Betweenness centrality

La betweenness centrality è una misura di centralità utilizzata per valutare l'importanza dei nodi in un grafo in base al loro ruolo di "intermediari" nella comunicazione tra altri nodi, siano questi del suo stesso tipo o di altri tipi.

In particolare, questa indica la frequenza con cui un nodo si trova lungo i cammini più brevi tra tutte le coppie di nodi nel grafo. Si tratta, quindi, di un modo per rilevare la quantità di influenza che un individuo ha sul flusso di informazioni.

L'algoritmo calcola i percorsi più brevi tra tutte le coppie di nodi di un grafo, poi ogni nodo riceve un punteggio basato sul numero di percorsi più brevi che passano attraverso di

esso; i nodi che si trovano più frequentemente sui percorsi più brevi tra altri nodi avranno punteggi di betweenness centrality più elevati.

Eigenvector centrality

La eigenvector centrality è un'ulteriore misura di centralità utilizzata per valutare l'importanza dei nodi in un grafo. A differenza di altre misure di centralità, come la degree centrality (che considera solo il numero di connessioni di un nodo), l'eigenvector centrality tiene conto anche dell'importanza dei nodi con cui esso è connesso.

In particolare, l'algoritmo assegna un valore di importanza a ciascun nodo in base alla sua capacità di essere connesso ad altri nodi importanti; un nodo, quindi, ha un valore elevato dell'eigenvector se è connesso a molti nodi che, a loro volta, hanno punteggi elevati.

Il calcolo dell'eigenvector centrality coinvolge l'uso degli autovettori, calcolando, nello specifico, l'autovettore associato all'autovalore assoluto più grande attraverso l'approccio dell'iterazione di potenza; all'interno di ogni iterazione, il punteggio di centralità di ciascun nodo è derivato dai punteggi dei suoi vicini entranti.

Se è vero che le misure di centralità consentono di comprendere le forze in gioco nei social network, bisogna considerare che queste sono, però, misure di tipo puntuale. Quello, quindi, che queste misure non ci dicono è come (o, meglio, perché) il visionario con un alto grado di centralità è circondato da tanti fan, cosa porta le persone a prendere una posizione in un dibattito, a mettersi insieme o ad isolarsi.

Per comprendere ciò, bisogna andare oltre le misure di centralità. Per questo esamineremo ora le triadi, che possono essere viste come il più piccolo insieme di persone che formano una società.

2.4.2 Triadi

Come già accennato in precedenza, un elemento fondamentale della teoria e dell'analisi delle reti sociali è la diade, ovvero una coppia di attori che possono condividere, o meno, una relazione sociale tra loro.

Una triade estende la singola diade a un insieme di tre attori, composto da tre diadi.

Nell'opera *The Sociology of Georg Simmel*, Georg Simmel ha sostenuto che le triadi sono un'unità fondamentale dell'analisi sociologica. Egli sosteneva che tre attori in una triade possono consentire dinamiche sociali qualitativamente diverse, che non possono essere ridotte agli individui o alle diadi.

Ad esempio, in una relazione tra tre soggetti A , B e C , il soggetto A può avere una relazione diadica con C , ma anche una relazione indiretta con C attraverso B . Il soggetto B può quindi modificare la forza o la natura della relazione tra A e C , ad esempio solidificando un'alleanza o mediando un conflitto. Se A e C non interagiscono direttamente, la parte B può mediare una transazione tra loro e può trarre potere da questa posizione di intermediazione.

Un interesse principale nello studio delle triadi è il fenomeno della transitività; una triade è transitiva se, quando esiste un legame da A a B e da B a C , esiste anche un legame da A a C . Se, invece, i legami $A - B$ e $B - C$ esistono entrambi, ma A non è legato a C , allora la triade è intransitiva.

La transitività è tipicamente studiata per le relazioni dirette, dove un legame da A a B è direzionale e non implica la stessa relazione da B ad A . Per le triadi in reti dirette, se i legami tra A e B e tra B e C esistono entrambi, ma tra A e C no, allora la triade è intransitiva (Figura 2.11). Alcune relazioni dirette hanno una tendenza pervasiva alla transitività: se A "domina" B e B "domina" C , allora è probabile che A "domini" anche C . Anche alcune relazioni non dirette tendono a essere transitive, come rappresentato dall'affermazione "un amico di un

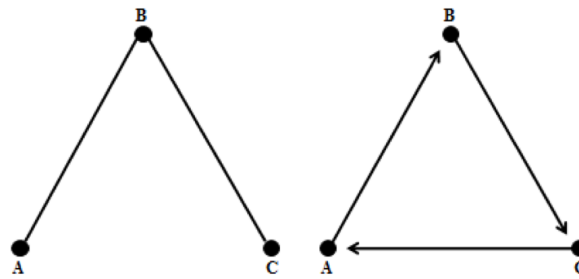


Figura 2.11: Triade indiretta (sinistra) e triade diretta (destra)

amico è un amico". Le triadi per alcune altre relazioni - in particolare i legami romantici tra amanti - sono prevalentemente intransitive.

Simmel ha anche sostenuto che le triadi possono servire come base analitica per la comprensione di gruppi sociali più ampi. I ricercatori hanno quindi descritto le reti esaminando la distribuzione delle relazioni a livello triadico, enumerando tutte le configurazioni triadiche qualitativamente distinte. Questo "censimento delle triadi" è stato utilizzato per osservare una tendenza generale alla socievolezza, alla gerarchia o alla transitività nelle reti sociali e ha costituito una prima base per l'analisi statistica delle reti.

Nei modelli di evoluzione delle reti, gli studiosi assumono spesso una propensione diretta alla chiusura delle triadi. La teoria dell'equilibrio strutturale di Cartwright e Harary offre una spiegazione motivazionale per la chiusura delle triadi nelle reti firmate, dove la relazione tra due attori può essere positiva o negativa. In particolare, il disaccordo tra amici legati positivamente (o l'accordo tra nemici legati negativamente) provoca una tensione psicologica che spinge le parti a risolvere questa dissonanza. In una triade di amicizia, quando l'amico di un amico è un nemico, allora o il nemico diventerà un amico o uno degli amici diventerà un nemico, per risolvere la tensione nelle relazioni.

Lavori recenti hanno messo in discussione l'enfasi di Simmel sulla triade come lente irriducibile per l'analisi delle dinamiche sociali, dimostrando che i modelli triadici possono essere sottoprodotti delle dinamiche sociali nelle diadi. Ad esempio, una propensione diadica all'omofilia (ovvero, scegliere amici che appartengono alle stesse categorie sociali, di cui abbiamo parlato nella sezione precedente) tenderà a favorire la transitività, anche se gli attori non hanno una propensione diretta alla chiusura delle triadi. In uno studio di ingegneri e avvocati, se gli ingegneri preferiscono altri ingegneri come amici, mentre gli avvocati preferiscono altri avvocati, questo genera una tendenza alla transitività. All'estremo, tutti gli ingegneri saranno amici di tutti gli ingegneri e tutti gli avvocati saranno amici di tutti gli avvocati e, quindi, tutte le triadi saranno chiuse senza alcuna propensione diretta a farlo. La ricerca empirica che prende in considerazione l'omofilia e la chiusura delle triadi, utilizzando di recente modelli di grafi casuali a famiglia esponenziale (exponential-family random graph models, ERGMs), mostra quanto siano intrecciate queste dinamiche. L'interdipendenza delle dinamiche sociali a livello diadico e triadico richiede una maggiore sofisticazione sia nelle teorie che nei modelli statistici di evoluzione delle reti.

Indipendentemente dai processi sociali che generano i modelli osservati nelle triadi, questi modelli hanno implicazioni sostanziali per le dinamiche dei gruppi, delle organizzazioni e dei mercati. Per esempio, James Coleman, nel suo trattato *Foundations of Social Theory* del 1990, ha sostenuto che la chiusura nelle triadi porta alla fiducia interpersonale, a una maggiore cooperazione e all'applicazione delle norme. Questo argomento può essere generalizzato a contesti più ampi, offrendo previsioni sul livello di cooperazione nei gruppi in funzione della struttura della rete.

Mark Granovetter, nell'articolo *The Strength of Weak Ties* del 1973, ha sostenuto che la

transitività è più pervasiva nelle triadi legate da legami forti (rispetto a quelle legate da legami deboli), con il risultato di avere gruppi densi di attori fortemente legati e legami deboli che vanno oltre i gruppi locali. Egli ha quindi sostenuto che i legami deboli trasmettono più informazioni nuove tra i cluster. Ronald Burt ha sviluppato una teoria dei "buchi strutturali", elaborando i vantaggi strategici di occupare una posizione di rete lungo tali percorsi. Entrambi traggono importanti conseguenze di alto livello dai modelli a livello triadico.

Related Literature

In questo capitolo analizziamo il contesto di riferimento relativo alla ricerca descritta nella presente tesi.

I multilayer network forniscono una rappresentazione davvero esaustiva dei sistemi complessi, poiché possono catturare e rappresentare molteplici tipologie di interazioni e relazioni, consentendo di analizzarne più aspetti contemporaneamente. Per questo motivo, come abbiamo visto nel Capitolo 1, queste reti hanno ricevuto un'attenzione significativa in molte aree di ricerca, tra cui la fisica [35], la biologia, la biomedicina [22] e le scienze sociali [3].

Il presente lavoro si propone di fornire un contributo alla ricerca sui multilayer network; in particolare, esso intende proporre un framework per studiare l'evoluzione temporale e la resilienza di un multilayer network associato a un multimodal social network, ovvero una rete sociale dedicata alla condivisione e alla diffusione di informazioni.

In realtà, come vedremo, il framework sviluppato può essere esteso e applicato a qualsiasi multilayer network, indipendentemente dal sistema complesso che esso rappresenta. Pertanto, per meglio posizionare questo lavoro rispetto alla letteratura correlata, ci concentriamo su due ambiti di ricerca, ovvero: (i) studi che valutano l'evoluzione, la resilienza o altri aspetti simili nei multilayer network e (ii) studi che indagano i multimodal social network e alcune delle loro proprietà attraverso la modellazione basata su multilayer network.

Da quanto emerso dallo studio della letteratura correlata, l'approccio qui descritto è il primo che interseca queste due tematiche.

In passato, lo studio di framework in grado di valutare il comportamento dei multilayer network ha riguardato soprattutto contesti applicativi; ne è un esempio l'articolo pubblicato da Shin et al. nel 2018, *The diffusion of misinformation on social media: Temporal pattern, message, and source* [37], il quale realizza un'analisi dinamica dei processi di diffusione della disinformazione online. Un altro esempio è lo studio del 2016, *Analysis of the Chinese Airline Network as multi-layer networks* [14], di Du et al., in cui si cerca di rappresentare l'intera rete di trasporto della Chinese Airline attraverso un multilayer network.

Esistono, tuttavia, anche alcuni lavori più orientati alla teoria, come *The structure and dynamics of multilayer networks* [35] di Boccaletti et al., in cui gli autori realizzano una rassegna completa sull'organizzazione strutturale e dinamica dei multilayer network, analizzando come la loro natura multilayer ne influenzi i processi e le dinamiche.

Per studiare vari fenomeni legati ai multilayer network e per proporre diverse misure al riguardo, solitamente viene utilizzato il concetto di modularità.

Inoltre, prima di analizzare questi lavori in modo più dettagliato, è anche importante notare che il termine "resilienza" viene utilizzato in modo diverso a seconda del contesto di utilizzo; tuttavia, gran parte della ricerca su questo tema è stata condotta, finora, nell'ambito dei sistemi di ingegneria generale.

Nonostante l'aspetto inedito del presente studio, di seguito vengono proposti gli articoli più rilevanti che hanno esteso e migliorato la ricerca sui multilayer network, evidenziandone gli aspetti rilevanti, le similarità e le differenze.

Un primo esempio è l'articolo *Evolution of prosocial behaviours in multilayer populations. Nature Human Behaviour* [38] pubblicato nel 2022 da Su et al; in questo articolo, gli autori utilizzano i multilayer network per studiare le relazioni nelle società umane in diversi ambiti. Essi dimostrano che l'accoppiamento tra i layer promuove un comportamento pro-sociale e che piccoli cambiamenti nelle interazioni in un dominio possono catalizzare la pro-socialità in un dominio diverso. Questo studio ci restituisce la capacità dell'analisi dei multilayer network di studiare le relazioni sociali complesse nelle società umane. L'approccio di Su et al. e il nostro framework possono essere considerati ortogonali; in particolare, diverse intuizioni scoperte nel primo possono essere utilizzate nel secondo al fine di studiare le relazioni tra pro-socialità e resilienza.

Analizzando, più nel dettaglio, l'articolo di Du et al. citato in precedenza, in questo studio gli autori analizzano la rete della compagnia aerea cinese, utilizzando il metodo della k-core decomposition e rappresentandola come un multilayer network con tre layer: Core, Bridge e Periphery. Essi dimostrano che la rete è più robusta quando vengono rimossi i nodi di basso grado o i collegamenti ad alto flusso di volo. Questo approccio presenta un'interessante analisi empirica; esso studia, infatti, la robustezza dei multilayer network rimuovendo i nodi in base a una strategia di selezione di alto o basso grado; analogamente, anche il nostro framework utilizza la tecnica di rimozione dei nodi per valutare la resilienza. Tuttavia, a parte questa somiglianza, gli obiettivi, e dunque gli approcci, dei due studi sono, in sostanza, diversi.

Gli autori Gao et al., nell'articolo *Multilayer Feature-Rich Satellite Network Analysis: An Application-Oriented and Time-Evolving Approach* [18] pubblicato nel 2020 propongono una rete multilayer di comunicazione satellitare che tiene conto di vari requisiti applicativi e bilancia le risorse limitate. L'obiettivo è quello di sviluppare una strategia di analisi della rete ricca di funzionalità e modelli di ottimizzazione multi-obiettivo per affrontare le variabili di progettazione. Questo approccio segue una procedura di sviluppo top-down ed è facilmente estendibile per includere elementi dettagliati di progettazione e di modellazione del sistema. Sia la metodologia che i risultati presentati sono rilevanti, soprattutto nel contesto delle reti di comunicazione, nonostante gli obiettivi di questo approccio e quelli del nostro framework siano, sostanzialmente, differenti.

Nell'articolo *Using multilayer network analysis to explore the temporal dynamics of collective behavior* [17] pubblicato nel 2021, gli autori D.N.Fisher e N.Pinter-Wollman analizzano le dinamiche temporali delle interazioni sociali e dei comportamenti collettivi nel ragno *Stegodyphus dumicola*. Utilizzando i multilayer network per quantificare i cambiamenti nelle reti sociali nel tempo e determinarne le conseguenze, rispettivamente, sui cambiamenti individuali e di gruppo, scoprono che le interazioni sociali cambiano nel tempo a un ritmo costante e che la stabilità della rete sociale è correlata all'audacia individuale e non al comportamento collettivo del gruppo. Il lavoro di Fisher e Pinter-Wollman presenta alcune analogie con lo studio di questa ricerca; in particolare, il primo si concentra sulla riducibilità strutturale di una rete, che è una misura della differenza nei modelli di interazione nei multilayer network; il nostro framework, invece, non limita la sua analisi ai pattern di interazione, ma considera la rete nel suo complesso, utilizzando diverse misure di analisi.

Boccaletti et al., nell'articolo *The structure and dynamics of multilayer networks* [35] men-

zionato in precedenza, passano in rassegna i recenti progressi della teoria delle reti relativi all'analisi di multilayer network, considerando le proprietà temporali e contestuali delle interazioni, che le ricerche precedenti avevano trascurato. Il loro studio sottolinea l'importanza di considerare queste proprietà nell'analisi dei sistemi complessi, dando, in aggiunta, un solido contributo ai concetti di resilienza e percolation.

La ricerca condotta da Chen e Lu e riportata nello studio *Measure of layer centrality in multilayer network* [11] propone un framework per caratterizzare la centralità dei layer nei multilayer network. Tale centralità può essere utilizzata per descrivere sistemi complessi costituiti da unità connesse da diversi tipi di relazioni. Gli autori introducono, nello specifico, due misure di centralità basate sulla connettività della rete e sull'assegnazione di valori di centralità ai layer. Validando tali misure attraverso un set di dati reali di multilayer network relative al trasporto aereo, scoprono che le misure sono in grado di estrarre nuove e utili informazioni dal set di dati considerato. Sebbene gli obiettivi di questo approccio differiscano da quelli del nostro framework, alcuni concetti e intuizioni possono essere utili anche nel nostro contesto.

Modularity in complex multilayer networks with multiple aspects: a static perspective [43] è un articolo del 2017 di Zhang et al. in cui gli autori propongono un approccio per il rilevamento di comunità in multilayer network, le cui entità hanno vari tipi di interazioni che possono cambiare nel tempo. Questa ricerca utilizza una funzione di modularità multistrato e un metodo spettrale, chiamato mSpec, per ottimizzare la funzione. Gli autori convalidano il loro approccio su multilayer network empirici e ne dimostrano l'affidabilità, fornendo un quadro di riferimento per la comprensione della struttura delle comunità nei sistemi complessi. Questo approccio presenta alcune somiglianze con quello della nostra ricerca; ad esempio, entrambi considerano il concetto di modularità, anche se questo studio si concentra sull'individuazione delle comunità, mentre il nostro framework utilizza le comunità per raggiungere i suoi obiettivi, ovvero lo studio dell'evoluzione temporale e della resilienza.

Nell'articolo *Community structure in time-dependent, multiscale, and multiplex networks* [29] del 2010, gli autori Mucha et al. presentano un framework generale che utilizza funzioni di qualità della rete per analizzare la struttura delle comunità in multislice network, i quali combinano le singole reti attraverso i collegamenti tra i nodi di diverse slice (o fette) della rete. Lo studio di Mucha et al. consente di approfondire l'analisi sulla struttura delle comunità in un contesto generale, includendo reti con diversi tipi di collegamenti e considerandone l'evoluzione temporale. È interessante notare che questo framework introduce l'indice di modularità multistrato che viene utilizzato anche nel nostro studio.

Il secondo filone di ricerca che analizziamo riguarda l'indagine dei social network attraverso l'utilizzo di multilayer network.

Nella ricerca condotta da Bonifazi et al. e riportata nel recente articolo del 2022 *Investigating the COVID-19 vaccine discussions on Twitter through a multilayer network-based approach* [7], gli autori propongono un approccio basato su multilayer network per analizzare le discussioni sui social network, concentrandosi su argomenti sensibili, come la sanità. Essi applicano l'approccio proposto a un set di dati Twitter relativi ai vaccini COVID-19 e identificano gli hashtag rilevanti per ogni linea di pensiero. I risultati mostrano che gli anti-vaxxer hanno ego-network più dense e coese, la quali portano a un maggior numero di interazioni tra di loro; da queste conclusioni ne deriva l'efficacia nel rilevare gli utenti influenti. L'approccio degli autori e il nostro framework condividono alcune similitudini, come l'utilizzo di un multilayer network per analizzare un social network. Tuttavia, mentre lo studio di Bonifazi et al. presenta un'analisi empirica, il nostro si propone di studiare l'evoluzione temporale e la resilienza, da prospettive sia teoriche che empiriche.

Logan et al., nello studio del 2023 *Social network analysis of Twitter interactions: a directed multilayer network approach* [26], hanno impiegano la social network analysis e il topic mode-

ling per modellare le interazioni degli utenti su Twitter come una rete diretta, ponderata e multilayer; l'obiettivo è quello di identificare gli utenti influenti e i gruppi altamente connessi. Gli autori dimostrano che una rete sociale focalizzata su topic che rappresenta le conversazioni restituisce risultati più solidi per quanto riguarda gli utenti influenti, soprattutto quando è associata a tweet relativi a un'ampia varietà di discussioni. Questo approccio può essere considerato ortogonale al nostro framework; infatti, pur avendo obiettivi diversi, alcune intuizioni del primo (ad esempio, l'uso della modellazione dei topic) possono essere incorporate nel secondo.

Con l'articolo *Discovering suspicious behavior in multilayer social networks* [5], gli autori Bindu et al. propongono un approccio non supervisionato, privo di parametri e basato su caratteristiche di rete, chiamato ADOMS (Anomaly Detection on Multilayer Social Networks), per cercare utenti anomali nei multilayer social network. L'approccio, basato sulle anomalie clique/near-clique e star/near-star, classifica gli utenti in base al grado di somiglianza delle loro vicinanze nei diversi layer, a strutture simili a stelle e clique. Anche in questo caso, l'approccio dei autori e il nostro framework presentano un punto di incontro; in particolare, nel nostro studio potrebbe essere integrato un task di rilevamento delle anomalie basato sulla misura di resilienza.

Detecting topic authoritative social media users: a multilayer network approach [32] è l'articolo pubblicato da Oro et al. nel 2017, in cui viene presentato un approccio per individuare gli utenti influenti sugli online social network. Questo viene fatto analizzando il contenuto dei loro post attraverso una rete a tre layer: utenti, articoli e parole chiave. L'approccio sfrutta l'algebra multilineare per individuare gli utenti più attivi nell'affermare il proprio punto di vista su articoli dominanti taggati con parole chiave dominanti.

Infine, menzioniamo l'articolo *A multi-layer approach to disinformation detection in US and Italian news spreading on Twitter* [33], con cui gli autori Pierri et al. hanno sviluppato un framework per classificare gli articoli di notizie come disinformazione o notizie mainstream, basandosi solo sui loro meccanismi di diffusione su Twitter. Nel multilayer network utilizzato a supporto della loro analisi, ogni layer descrive un diverso tipo di interazione; ciò consente di quantificare l'impatto di ciascun layer nell'attività di classificazione. Le diverse intuizioni individuate dagli autori possono essere utilizzate anche nel nostro framework, in cui la resilienza potrebbe essere definita come una misura che indica la facilità con cui il modello di condivisione di un contenuto viene effettuato nel multilayer network.

Descrizione del framework

In questo capitolo presentiamo il framework proposto finalizzato a studiare, attraverso un multilayer network, l'evoluzione temporale e la resilienza di un multimodal social network. In particolare, nella Sezione 4.1 verrà illustrato il modello del multilayer network utilizzato, mentre nella Sezione 4.2 verranno descritte le misure e gli approcci di analisi implementati.

4.1 Descrizione del modello basato su un multilayer network

Il nostro modello basato su un multilayer network può essere rappresentato dalla quadrupla:

$$\mathcal{M} = \langle A, C, T, E \rangle$$

dove: A , C e T rappresentano tre layer associati, rispettivamente, agli autori, ai commenti e ai topic (argomenti), mentre E denota l'insieme degli archi (edge) inter-layer, i quali collegano, cioè, nodi appartenenti a layer diversi. Più precisamente:

- $A = \langle N^A, E^A \rangle$ è il layer che rappresenta gli autori. In esso è presente un nodo $a_i \in N^A$ per ogni autore che ha scritto un commento; inoltre, esiste un arco $(a_i, a_j, w_{ij}^A, \tau_{ij}^A) \in E^A$ se si è verificata almeno un'interazione tra a_i e a_j , ovvero se a_i ha risposto a un commento di a_j , o viceversa. Il peso w_{ij}^A rappresenta il numero di interazioni tra a_i e a_j , mentre τ_{ij}^A indica il timestamp dell'ultima interazione.
- $C = \langle N^C, E^C \rangle$ è il layer che rappresenta i commenti. È presente un nodo $c_i \in N^C$ per ogni commento pubblicato nel multimodal social network; inoltre, esiste un arco $(c_i, c_j, w_{ij}^C, \tau_{ij}^C) \in E^C$ se c_i si riferisce a c_j , o viceversa. Il peso w_{ij}^C rappresenta il grado di somiglianza tra il contenuto espresso in c_i e c_j , mentre τ_{ij}^C indica il timestamp in cui è stato pubblicato l'ultimo dei commenti tra c_i e c_j .
- $T = \langle N^T, E^T \rangle$ è il layer che rappresenta i topic. Esiste un nodo specifico $t_i \in N^T$ per ogni argomento discusso; inoltre, esiste un arco $(t_i, t_j, w_{ij}^T, \tau_{ij}^T) \in E^T$ se t_i e t_j sono in relazione tra loro. In particolare, t_i potrebbe essere un argomento della stessa categoria di t_j (ad esempio, entrambi potrebbero riferirsi al cibo), oppure potrebbe essere derivato da t_j (ad esempio, t_i potrebbe essere "Inflazione" e t_j potrebbe essere "Economia"). Il peso w_{ij}^T rappresenta, in questo caso, la forza della relazione tra t_i e t_j , mentre τ_{ij}^T indica

il momento in cui l'ultimo degli argomenti tra t_i e t_j è apparso per la prima volta nel multimodal social network.

- E è l'insieme degli archi inter-layer, i quali collegano, cioè, nodi appartenenti a layer diversi. Nel nostro modello possiamo avere archi tra i nodi di T e quelli di A e tra i nodi di C e quelli di C stesso. In particolare, esiste un arco $(t_i, c_j, \tau_{ij}^{TC}) \in E$ se l'argomento t_i è discusso nel commento c_j , dove τ_{ij}^{TC} è il timestamp di quando c_j è stato pubblicato. Infine, esiste l'arco $(c_i, a_j, \tau_{ij}^{CA}) \in E$ se il commento c_i è stato pubblicato dall'autore a_j , dove τ_{ij}^{CA} è il timestamp di quando c_i è stato pubblicato.

In Figura 4.1, forniamo una rappresentazione grafica del modello appena descritto.

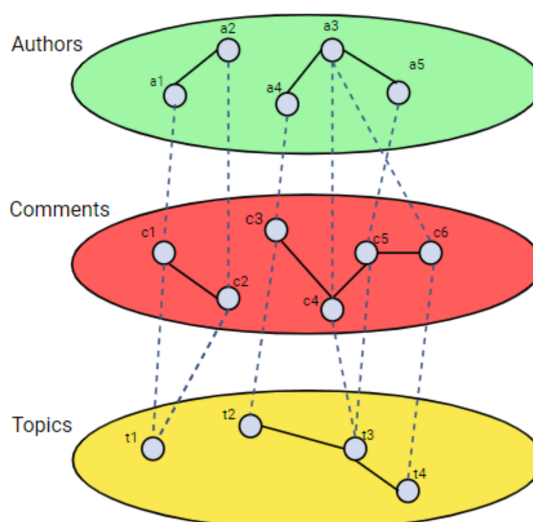


Figura 4.1: Rappresentazione del multilayer network proposto

4.1.1 Definizione dell'istanza temporale di un multilayer network

Per studiare l'evoluzione temporale e la resilienza di un multimodal social network, abbiamo bisogno di un meccanismo che ci consenta di rappresentare il fattore tempo all'interno del nostro modello. In particolare, dobbiamo esaminare le reti in diversi intervalli di tempo, per capire come esse cambiano e rispondono alle perturbazioni.

A tal fine, introduciamo il concetto di istanza temporale di un multilayer network, la quale cattura la struttura della rete in uno specifico intervallo di tempo. Attraverso di essa, possiamo studiare come la struttura della rete e le interazioni che essa modella, sia in assenza che in presenza di perturbazioni, si evolvono nel tempo.

Un'istanza temporale del modello \mathcal{M} , basato sul multilayer network proposto, può essere rappresentata come:

$$\mathcal{M}[\tau_1, \tau_2] = \langle A[\tau_1, \tau_2], C[\tau_1, \tau_2], T[\tau_1, \tau_2], E[\tau_1, \tau_2] \rangle$$

dove: τ_1 e τ_2 sono, rispettivamente, il timestamp iniziale e quello finale dell'intervallo di tempo di interesse. I layer $A[\tau_1, \tau_2]$, $C[\tau_1, \tau_2]$, $T[\tau_1, \tau_2]$ e l'insieme degli archi $E[\tau_1, \tau_2]$ hanno la stessa struttura, rispettivamente, dei layer A , C , T e dell'insieme degli archi E di \mathcal{M} definiti sopra; tuttavia, essi contengono solo i nodi e gli archi presenti nell'intervallo di tempo $[\tau_1, \tau_2]$.

Chiaramente, se indichiamo con 0 l'istante di origine del multimodal social network e con τ_l l'ultimo istante temporale a cui si riferiscono i nostri dati, abbiamo che $\mathcal{M} = \mathcal{M}[0, \tau_l]$.

Ad esempio, un'istanza del modello potrebbe rappresentare il multimodal social network durante un mese o una settimana specifici. In questo caso, il layer A include solo gli autori che hanno pubblicato commenti in quel periodo, il layer C solo i commenti pubblicati in quel periodo e il layer T solo gli argomenti discussi in quel periodo. Infine, l'insieme E contiene unicamente gli archi che rappresentano le interazioni tra i nodi di A e C , o tra i nodi di C e T , che si sono verificate durante quel periodo.

4.2 Misure associate al framework

Il nostro framework supporta sia l'analisi dei singoli layer, che quella del multilayer network nel suo complesso; inoltre, esso include l'analisi statica del multilayer network per tutto l'intervallo di tempo di interesse e l'analisi dinamica che tiene conto della sua evoluzione nel tempo.

Per garantire questo studio, dobbiamo necessariamente introdurre alcune misure. Nello specifico, in questa sezione esaminiamo le misure impiegate dal nostro framework in base a tre aspetti del multilayer network che vogliamo affrontare; ovvero la struttura, le comunità e la resilienza.

4.2.1 Misure relative alla struttura di rete

Per studiare la struttura di un multilayer network, è necessario eseguire analisi sia intra-layer che inter-layer.

Inoltre, per analizzare l'evoluzione della rete nel tempo, dobbiamo definire una lunghezza costante per gli intervalli di tempo da associare alle varie istanze temporali del modello. Questa lunghezza può essere scelta in base al set di dati di riferimento, allo scenario applicativo e al tipo di analisi che vogliamo condurre.

In particolare, indichiamo con $\mathcal{T} = [0, \tau_l]$ il periodo di tempo totale a cui si riferiscono i dati del multimodal social network che vogliamo analizzare. Dobbiamo, poi, suddividere \mathcal{T} in sotto-intervalli contigui e non sovrapposti $[\tau_1, \tau_2]$, in modo che $\tau_2 - \tau_1$ sia uguale alla lunghezza scelta.

Ad esempio, qualora \mathcal{T} fosse pari a 4 settimane, potremmo dividerlo in 4 sotto-intervalli, ognuno dei quali ha una lunghezza di 1 settimana; in questo caso, il multilayer network complessivo \mathcal{M} darà origine a quattro istanze temporali, ciascuna associata, rispettivamente, a una settimana specifica.

Siano $\mathcal{M}[\tau_1, \tau_2]$ un'istanza temporale di \mathcal{M} e L uno dei tre layer di $\mathcal{M}[\tau_1, \tau_2]$ (cioè, $L \in A[\tau_1, \tau_2], C[\tau_1, \tau_2], T[\tau_1, \tau_2]$). Introduciamo, ora, il degree-degree cross-correlation coefficient r_L tra i nodi del layer L . Questo è definito come:

$$r_L = \frac{\sum_{n_i \in N^L} \sum_{n_j \in (N^L \setminus \{n_i\})} [(\delta_i - \bar{\delta}^L)(\delta_j - \bar{\delta}^L)]}{\sqrt{\sum_{n_i \in N^L} (\delta_i - \bar{\delta}^L)^2 \sum_{n_j \in (N^L \setminus \{n_i\})} (\delta_j - \bar{\delta}^L)^2}}$$

Nello specifico: (i) N^L rappresenta l'insieme dei nodi del layer L ; (ii) δ_i (resp., δ_j) indica il grado del nodo n_i (resp., n_j) prendendo in considerazione solo gli archi intra-layer; (iii) $\bar{\delta}^L$ indica il grado medio dei nodi di L , prendendo in considerazione solo gli archi intra-layer.

Il degree-degree cross-correlation coefficient r_L misura, quindi, l'assortatività dei nodi di L rispetto al loro grado. In altre parole, esso è un indicatore della tendenza dei nodi di L a essere collegati ad altri nodi di L con un grado simile. I suoi valori sono compresi nell'intervallo

reale $[-1, 1]$, dove 1 indica l'assortatività totale, -1 rappresenta la disassortatività totale e 0 denota la mancanza di assortatività.

Introduciamo, ora, il degree-degree cross-correlation coefficient r_{L_1, L_2} tra i nodi di due diversi layer: L_1 e L_2 . Esso può essere definito come segue:

$$r_{L_1, L_2} = \frac{\sum_{n_i \in N^{L_1}} \sum_{n_j \in N^{L_2}} [(\delta_i - \overline{\delta^{L_1}})(\delta_j - \overline{\delta^{L_2}})]}{\sqrt{\sum_{n_i \in N^{L_1}} (\delta_i - \overline{\delta^{L_1}})^2 \sum_{n_j \in N^{L_2}} (\delta_j - \overline{\delta^{L_2}})^2}}$$

In questo caso: (i) N^{L_1} (risp., N^{L_2}) rappresenta l'insieme dei nodi di L_1 (risp., L_2); (ii) δ_i (risp., δ_j) indica il grado del nodo $n_i \in N^{L_1}$ (risp., $n_j \in N^{L_2}$); (iii) $\overline{\delta^{L_1}}$ (risp., $\overline{\delta^{L_2}}$) indica il grado medio dei nodi di L_1 (risp., L_2).

Dunque, il degree-degree cross-correlation coefficient r_{L_1, L_2} misura la correlazione tra i gradi dei nodi dei diversi layer, analizzando, nello specifico, l'assortatività dei nodi di L_1 e L_2 rispetto al loro grado. Anche r_{L_1, L_2} varia nell'intervallo reale $[-1, 1]$, dove -1 indica la massima disassortatività, 0 la mancanza di assortatività e disassortatività e 1 l'assortatività massima.

Un'altra misura utile per studiare la struttura della rete è il clustering coefficient, il quale rappresenta il grado di connessione tra i nodi dello stesso vicinato. In particolare, il coefficiente di clustering γ_L di un layer L di $\mathcal{M}_{[\tau_1, \tau_2]}$ può essere definito come:

$$\gamma_L = \frac{3N_{closed}}{N_{connected}}$$

dove: (i) N_{closed} indica il numero di triadi chiuse in L ; (ii) $N_{connected}$ indica il numero di triadi collegate.

Infine, un'ultima misura che il nostro framework adotta per valutare la struttura di rete è l'average path length, il quale calcola la lunghezza media dei percorsi più brevi tra coppie di nodi. In particolare, l'average path length del percorso di un layer L di un multilayer network $\mathcal{M}_{[\tau_1, \tau_2]}$ può essere definito come:

$$\mathcal{L} = \frac{1}{|N^L|(|N^L| - 1)} \sum_{n_i \in N^L} \sum_{n_j \in (N^L \setminus \{n_i\})} \langle n_i, n_j \rangle$$

In questo caso: (i) $|N^L|$ è la cardinalità di N^L , ovvero il numero dei suoi nodi; (ii) $\langle n_i, n_j \rangle$ è la lunghezza del percorso più breve tra i nodi n_i e n_j nel layer L .

4.2.2 Misure relative alle comunità

L'analisi delle comunità è un metodo ampiamente utilizzato per studiare l'organizzazione e la coesione di reti complesse. Nel contesto dei multimodal social network, le comunità possono rappresentare, ad esempio, gruppi di utenti interessati ad argomenti simili e che, quindi, si impegnano in interazioni affini.

Per eseguire l'analisi delle comunità su un multilayer network, dobbiamo, innanzitutto, identificare le comunità all'interno di ogni layer (comunità intra-layer), così come le comunità che coinvolgono nodi di layer diversi (comunità inter-layer).

Analogamente a quanto abbiamo fatto per la struttura della rete, dato un multilayer network \mathcal{M} e una delle sue istanze temporali $\mathcal{M}_{[\tau_1, \tau_2]}$, sia L uno dei suoi layer (cioè, $L \in \{A[\tau_1, \tau_2], C[\tau_1, \tau_2], T[\tau_1, \tau_2]\}$).

Al fine di identificare le comunità intra-layer, il nostro framework impiega l'algoritmo di Louvain [6], ampiamente utilizzato in letteratura. Questo algoritmo massimizza la modularità della rete (cioè la sua tendenza a essere divisa in comunità coese e non strettamente

accoppiate), unendo o dividendo iterativamente le comunità in base al miglioramento della modularità.

Invece, per identificare le comunità inter-layer, il nostro framework adotta Infomap [34], un algoritmo utilizzato per individuare le comunità inter-layer e che opera ottimizzando una funzione obiettivo che bilancia la densità dei collegamenti all'interno delle comunità e la somiglianza tra i layer. Infomap è in grado di rilevare comunità sia sovrapposte che non sovrapposte.

Per misurare la modularità di ogni layer, adottiamo l'indice di modularità classico. In particolare, l'indice di modularità Q_L di L è definito come:

$$Q_L = \frac{1}{2|E^L|} \sum_{n_i \in N^L} \left[\sum_{n_j \in N^L} \left(w_{ij} - \frac{\sigma_i \sigma_j}{2|E^L|} \right) \Gamma(C_i, C_j) \right]$$

dove: (i) $|E^L|$ indica il numero di archi di L ; (ii) N^L indica l'insieme dei nodi di L ; (iii) w_{ij} è il peso dell'arco che collega n_i a n_j ; (iv) σ_i (risp., σ_j) è la forza del nodo n_i (risp., n_j) in L ed è definito come $\sigma_i = \sum_{n_j \in N^L} w_{ij}$; (v) C_i (risp., C_j) indica la comunità a cui appartiene n_i (risp., n_j); (vi) $\Gamma(C_i, C_j)$ è la funzione delta di Kronecker, la quale è uguale a 1 se $C_i = C_j$, altrimenti è uguale a 0.

Q_L è compreso nell'intervallo reale $[0, 1]$; più alto è Q_L , più forte è la suddivisione di L in comunità distinte.

In modo analogo, per quantificare la modularità dell'intero multilayer network \mathcal{M} , il nostro framework impiega un indice di modularità multilayer finalizzato a misurare il grado di suddivisione di \mathcal{M} in comunità distinte che comprendono più layer. Se denotiamo con $\mathcal{L}\mathcal{S}$ l'insieme dei layer di \mathcal{M} (nel nostro caso $\mathcal{L}\mathcal{S} = \{A, C, T\}$), l'indice di modularità multilayer utilizzato nel nostro framework è definito come segue:

$$Q_{\mathcal{M}} = \frac{1}{2 \sum_{L \in \mathcal{L}\mathcal{S}} |E^L|} \sum_{n_i \in N^{L_1}, L_1 \in \mathcal{L}\mathcal{S}} \left[\sum_{n_j \in N^{L_2}, L_2 \in \mathcal{L}\mathcal{S}, L_2 \neq L_1} \left(w_{ij} - \rho \frac{\sigma_i \sigma_j}{2 \sum_{L \in \mathcal{L}\mathcal{S}} |E^L|} \right) \Gamma(C_i, C_j) \Gamma(L_i, L_j) \right]$$

In questo caso:

- $|E^L|$, N^L , w_{ij} , C_i e C_j sono stati già definiti in precedenza.
- ρ è un parametro che tiene conto dell'eterogeneità del numero di nodi nei vari layer del multilayer network e influisce in modo sostanziale sull'importanza da attribuire a una comunità di una certa dimensione. Ad esempio, una comunità di 10 nodi è molto grande e importante in un layer che contiene un totale di 20 nodi, mentre, in un layer che contiene un totale di 1000 nodi, essa è trascurabile. ρ varia nell'intervallo reale $[0, 1]$; maggiore è ρ e più i layer sono omogenei tra loro in termini di numero di nodi.
- Poiché gli archi inter-layer non sono pesati, assumiamo che tutti abbiano un peso pari a 1; pertanto, $\sigma_i = \delta_i$ (risp., $\sigma_j = \delta_j$), dove δ_i (risp., δ_j) rappresenta il numero di archi inter-layer incidenti su n_i (risp., n_j).

$Q_{\mathcal{M}}$ varia nell'intervallo reale $[-1, 1]$; i valori maggiori (risp., minori) di 0 indicano che le comunità in cui \mathcal{M} è diviso sono più (risp., meno) coese di quelle che si otterrebbero inserendo lo stesso numero di archi intra-layer e inter-layer in modo casuale.

Nella definizione dell'indice di modularità multilayer $Q_{\mathcal{M}}$ è stato preso come modello il cross-modularity index, introdotto nell'articolo *Community structure in time-dependent, multiscale, and multiplex networks* [29], pubblicato da Mucha et al. e già citato nel Capitolo 3.

Sebbene le due formulazioni condividano alcune somiglianze, esse presentano anche rilevanti differenze; infatti, nonostante entrambe vengano utilizzate per misurare la modularità di multilayer network, l'approccio di [29] è stato progettato per le reti multislice, ovvero multilayer network senza archi inter-layer, mentre il nostro framework gestisce multilayer network con archi sia intra-layer che inter-layer.

Analizzando le comunità all'interno di ogni layer e del multilayer network, possiamo identificare i gruppi di utenti interessati ad argomenti simili o che interagiscono tra loro; inoltre, confrontando le comunità relative a diverse istanze temporali del multilayer network, possiamo identificare i cambiamenti nella sua organizzazione e modularità nel tempo.

4.2.3 Misure relative alla resilienza

Al fine di studiare la resilienza di un multilayer network, è necessario verificare cosa succede a quest'ultimo in presenza di perturbazioni sia intra-layer che inter-layer. Per valutarla quantitativamente, il nostro studio adotta diverse metriche che definiamo nelle prossime sottosezioni.

Perturbazioni intra-layer

Le perturbazioni intra-layer si riferiscono ai cambiamenti nei pattern di connettività all'interno di un singolo layer, i quali possono essere causati da vari fattori, come il comportamento dell'utente o i cambiamenti negli argomenti trattati.

Le perturbazioni intra-layer possono essere di sei tipi, ossia: (i) aggiunta di nuovi nodi e dei relativi archi; (ii) rimozione di nodi esistenti e degli archi corrispondenti; (iii) aggiunta di nuovi archi a nodi esistenti; (iv) rimozione di archi esistenti; (v) aumento dei pesi degli archi; (vi) diminuzione dei pesi degli archi.

Le perturbazioni intra-layer possono rappresentare cambiamenti nelle interazioni tra gli utenti o negli argomenti di cui questi ultimi discutono; ad esempio, un cambiamento nel comportamento di un particolare utente o di gruppi di utenti può portare alla rimozione o all'aggiunta di connessioni. Siano:

- \mathcal{M} un multilayer network e L un livello di \mathcal{M} (cioè $L \in \mathcal{L}\mathcal{S}$);
- $L[0, \tau_1]$ (risp., $L[0, \tau_2]$) l'occorrenza di L nell'istanza temporale $\mathcal{M}[0, \tau_1]$ (risp., $\mathcal{M}[0, \tau_2]$) di \mathcal{M} ;
- $\delta_M^L[0, \tau_1]$ (risp., $\delta_M^L[0, \tau_2]$) il grado massimo di un nodo in $L[0, \tau_1]$ (risp., $L[0, \tau_2]$);
- δ_M il massimo tra $\delta_M^L[0, \tau_1]$ e $\delta_M^L[0, \tau_2]$.

Per studiare la resilienza di \mathcal{M} alle perturbazioni intra-layer che si verificano nell'intervallo di tempo $[\tau_1, \tau_2]$, definiamo le seguenti metriche che abbiamo incluso nel nostro framework:

- *Clustering Coefficient Variation* $\Delta\gamma[L_1, L_2]$: questa metrica misura l'entità della variazione del clustering coefficient γ_L di L causata dalle perturbazioni intra-layer che si verificano nell'intervallo di tempo $[\tau_1, \tau_2]$. In altre parole, essa misura il grado in cui le perturbazioni intra-layer, verificatesi nell'intervallo di tempo $[\tau_1, \tau_2]$, hanno influenzato la struttura della rete associata a L . La definizione è la seguente:

$$\Delta\gamma_L[\tau_1, \tau_2] = \frac{\gamma_L[0, \tau_2] - \gamma_L[0, \tau_1]}{\gamma_L[0, \tau_1]}$$

dove $\gamma_L[0, \tau_2]$ (risp., $\gamma_L[0, \tau_1]$) denota il clustering coefficient di $L[0, \tau_2]$ (risp., $L[0, \tau_1]$).

- *Level-wise Degree Distribution Variation* $\Delta D_L[\tau_1, \tau_2]$: essa misura l'entità della variazione del numero di nodi $D_L(k)[0, \tau_1]$ di grado k in $L[0, \tau_1]$, con $0 \leq k \leq \delta_M$, causata dalle perturbazioni intra-layer verificatesi nell'intervallo di tempo $[\tau_1, \tau_2]$. In altre parole, questa metrica misura il grado in cui le perturbazioni intra-layer verificatesi nell'intervallo di tempo $[\tau_1, \tau_2]$ hanno influenzato complessivamente i pattern di connettività all'interno di L . Essa è definita nel seguente modo:

$$\Delta D_L[\tau_1, \tau_2] = \frac{\sum_{k=0}^{\delta_M} |\Delta D_L(k)[\tau_1, \tau_2]|}{\delta_M}$$

dove:

$$\Delta D_L(k)[\tau_1, \tau_2] = \frac{\Delta D_L(k)[0, \tau_2] - \Delta D_L(k)[0, \tau_1]}{\Delta D_L(k)[0, \tau_1]}$$

Le due metriche appena definite ci permettono di analizzare la resilienza del multilayer network contestualmente ai cambiamenti nelle interazioni tra gli utenti (risp., commenti, argomenti) che discutono (risp., che vengono postati, che vengono discussi) nel social network.

Ad esempio, una variazione significativa $\Delta \gamma_L[\tau_1, \tau_2]$ del clustering coefficient γ_L dopo alcune perturbazioni intra-layer che si verificano nell'intervallo di tempo $[\tau_1, \tau_2]$ può indicare che il comportamento dei nodi coinvolti sta avendo un impatto sproporzionato sulla struttura della rete locale. D'altra parte, un cambiamento significativo nella level-wise degree distribution $\Delta D_L(k)[\tau_1, \tau_2]$ di L significa che le interazioni degli utenti (risp., i commenti, gli argomenti) si stanno spostando da uno o più utenti (risp., commenti, argomenti) ad altri. Tale spostamento può portare a cambiamenti significativi nella struttura generale del multilayer network.

Perturbazioni inter-layer

Le perturbazioni inter-layer sono cambiamenti nella struttura delle connessioni che collegano i nodi di diversi layer. Questo tipo di perturbazioni richiede l'aggiunta o la rimozione di nodi in almeno due layer, così come l'aggiunta o la rimozione di archi inter-layer.

Ad esempio, i cambiamenti nella popolarità degli argomenti o negli interessi di un utente o di un gruppo di utenti possono influenzare i commenti corrispondenti e i modelli di connettività tra i layer.

Le perturbazioni inter-layer possono essere di quattro tipi, ovvero: (i) aggiunta di archi inter-layer, che comporta anche l'aggiunta di nodi in almeno due layer diversi; (ii) rimozione di archi inter-layer, che comporta anche la rimozione di nodi in almeno due layer diversi; (iii) aggiunta di nodi, ciascuno dei quali ha almeno un arco inter-layer che lo collega con un nodo esistente o un nuovo nodo in un altro layer; (iv) rimozione di nodi, ciascuno dei quali ha almeno un arco inter-layer incidente su di esso.

Analizzando la resilienza di un multilayer network contestualmente a perturbazioni inter-layer possiamo valutare la capacità della rete di adattarsi ai cambiamenti nel modo in cui gli argomenti vengono discussi da diversi gruppi di utenti. Siano:

- \mathcal{M} un multilayer network e siano $\mathcal{M}_1 = \mathcal{M}[0, \tau_1]$ e $\mathcal{M}_2 = \mathcal{M}[0, \tau_2]$ due istanze temporali di \mathcal{M} ;
- L_1 e L_2 due layer di \mathcal{M} , tali che $L_1 \neq L_2$;
- $L_1[0, \tau_1]$ (risp., $L_2[0, \tau_2]$) l'occorrenza del layer L_1 (risp., L_2) corrispondente all'istanza temporale \mathcal{M}_1 ;

- $L_1[0, \tau_2]$ (risp., $L_2[0, \tau_2]$) l'occorrenza del layer L_1 (risp., L_2) corrispondente all'istanza temporale \mathcal{M}_2 ;
- $E[0, \tau_1]$ (risp., $E[0, \tau_2]$) l'occorrenza dell'insieme E di archi inter-layer corrispondenti all'istanza temporale \mathcal{M}_1 (risp., \mathcal{M}_2);
- $\delta_M^E[0, \tau_1]$ (risp., $\delta_M^E[0, \tau_2]$) il grado massimo di un nodo connesso ad almeno un arco di $E[0, \tau_1]$ (risp., $E[0, \tau_2]$);
- δ il massimo tra $\delta_M^E[0, \tau_1]$ e $\delta_M^E[0, \tau_2]$.

Per studiare la resilienza di \mathcal{M} alle perturbazioni inter-layer che si verificano nell'intervallo di tempo $[\tau_1, \tau_2]$ definiamo le seguenti metriche:

- *Modularity Variation* $\Delta Q_{\mathcal{M}}[\tau_1, \tau_2]$: questa metrica misura l'entità della variazione dell'indice di modularità $Q_{\mathcal{M}}$ del multilayer network causata dalle perturbazioni inter-layer che si verificano nell'intervallo di tempo $[\tau_1, \tau_2]$. In altre parole, essa misura il grado in cui le perturbazioni inter-layer, verificatesi nell'intervallo di tempo $[\tau_1, \tau_2]$, hanno alterato la struttura delle comunità della rete. La definizione è la seguente:

$$\Delta Q_{\mathcal{M}}[\tau_1, \tau_2] = \frac{Q_{\mathcal{M}}[0, \tau_2] - Q_{\mathcal{M}}[0, \tau_1]}{Q_{\mathcal{M}}[0, \tau_1]}$$

In particolare, $Q_{\mathcal{M}}[0, \tau_2]$ (risp., $Q_{\mathcal{M}}[0, \tau_1]$) denota l'indice di modularità multilayer rispetto a $\mathcal{M}[0, \tau_2]$ (risp., $\mathcal{M}[0, \tau_1]$).

- *Inter-layer Correlation Variation* $\Delta r_{L_1, L_2}[\tau_1, \tau_2]$: essa misura l'entità della variazione del degree-degree cross-correlation coefficient r_{L_1, L_2} tra i layer L_1 e L_2 di \mathcal{M} , causata dalle perturbazioni inter-layer che si verificano nell'intervallo di tempo $[\tau_1, \tau_2]$. In altre parole, questa metrica misura il grado in cui le perturbazioni inter-layer, che si verificano nell'intervallo di tempo $[\tau_1, \tau_2]$, hanno disturbato la correlazione tra le modalità di connessioni dei layer L_1 e L_2 . Essa è definita come:

$$\Delta r_{L_1, L_2}[\tau_1, \tau_2] = \frac{r_{L_1, L_2}[0, \tau_2] - r_{L_1, L_2}[0, \tau_1]}{r_{L_1, L_2}[0, \tau_1]}$$

dove $r_{L_1, L_2}[0, \tau_2]$ (risp., $r_{L_1, L_2}[0, \tau_1]$) indica il degree-degree cross-correlation coefficient tra i layer L_1 e L_2 di $\mathcal{M}[0, \tau_2]$ (risp., $\mathcal{M}[0, \tau_1]$).

- *Multilayer Degree Distribution Variation* $\Delta D_E[\tau_1, \tau_2]$: questa metrica misura l'entità della variazione del numero di nodi $D_E(k)[0, \tau_1]$ di grado k in E , con $0 \leq k \leq \delta_M$, causata dalle perturbazioni inter-layer verificatesi nell'intervallo di tempo $[\tau_1, \tau_2]$. In altre parole, essa misura il grado in cui le perturbazioni inter-layer, verificatesi nell'intervallo di tempo $[\tau_1, \tau_2]$, hanno influenzato i modelli di connettività complessiva all'interno di E . La definizione è la seguente:

$$\Delta D_E[\tau_1, \tau_2] = \frac{\sum_{k=0}^{\delta_M} |\Delta D_E(k)[\tau_1, \tau_2]|}{\delta_M}$$

dove:

$$\Delta D_E(k)[\tau_1, \tau_2] = \frac{\Delta D_E(k)[0, \tau_2] - \Delta D_E(k)[0, \tau_1]}{\Delta D_E(k)[0, \tau_1]}$$

Grazie alle metriche appena definite, possiamo valutare la resilienza del multilayer network \mathcal{M} rispetto ai cambiamenti nel modo in cui gli argomenti vengono discussi dai diversi gruppi di utenti.

Ad esempio, se l'indice di modularità $\Delta Q_M[\tau_1, \tau_2]$ rimane relativamente stabile, possiamo affermare che la struttura della comunità della rete è resistente ai cambiamenti degli argomenti discussi. D'altra parte, se il degree-degree cross-correlation coefficient diminuisce in modo significativo (cioè, il valore di $\Delta r_{L_1, L_2}[\tau_1, \tau_2]$ è negativo), possiamo sostenere che i modelli di connettività dei diversi layer stanno diventando meno correlati, il che può potenzialmente portare alla formazione di sottocomunità isolate.

In questo capitolo presentiamo gli esperimenti condotti e i risultati ottenuti attraverso l'applicazione del nostro framework. In particolare: nella Sezione 5.1 descriviamo le attività di raccolta e di pre-elaborazione dei dati; nella Sezione 5.2 illustriamo alcune analisi esplorative condotte sul nostro set di dati; nella Sezione 5.3 presentiamo i nostri esperimenti relativi alla struttura della rete; nella Sezione 5.4 descriviamo gli esperimenti relativi all'analisi delle comunità; nella Sezione 5.5 illustriamo l'analisi della resilienza; infine, nella Sezione 5.6 discutiamo i risultati ottenuti.

5.1 Raccolta e pre-elaborazione dei dati

Per svolgere l'analisi sperimentale, abbiamo costruito un set di dati ottenuti da Reddit, un social network popolare dove gli utenti possono condividere notizie, opinioni e contenuti su innumerevoli argomenti [2]; in particolare, abbiamo ricavato i dati di interesse dal subreddit *r/worldcup*. Quest'ultimo è dedicato alle discussioni relative al torneo calcistico FIFA World Cup, in cui gli utenti dialogano, principalmente, in merito a squadre, giocatori e partite.

Per estrarre i dati relativi a questo subreddit, abbiamo utilizzato *pushshift.io* [4], uno dei più grandi repository di dati di Reddit; nello specifico, sono stati raccolti i dati per i mesi di novembre e dicembre 2022, i quali corrispondono al periodo in cui il torneo si è svolto in Qatar.

Reddit offre un'opportunità unica per studiare le interazioni sociali e le dinamiche comunitarie in un ambiente online, in quanto consente di analizzare i contenuti generati dagli utenti e dalle loro interazioni [10].

Il subreddit *r/worldcup* fornisce, infatti, una ricca fonte di dati per studiare le dinamiche di una specifica comunità relative a un'importante manifestazione. FIFA World Cup è un evento popolare e molto seguito, con un pubblico ampio e diversificato; per questo, rappresenta un caso ideale per studiare come le comunità nei social network online si formano, si evolvono, interagiscono e scompaiono [21].

Il set di dati iniziale, ottenuto da *r/worldcup* attraverso *pushshift.io*, consisteva in una serie di commenti, a ciascuno dei quali erano associati i seguenti campi:

- *id*, che rappresenta il suo identificatore univoco;
- *parent_id*, che indica l'*id* del commento genitore, se esso esiste;
- *author*, che indica lo username anonimizzato del suo autore;

- *created_utc*, che rappresenta la data e l'ora in cui il commento è stato pubblicato;
- *text*, che denota il testo del commento stesso.

Abbiamo pre-elaborato il set di dati eliminando i commenti non in inglese, i commenti vuoti, i commenti contenenti solo link o immagini e i commenti scritti da account "bannati". Per automatizzare le attività di NLP ci siamo serviti della libreria Python *nlTK*.

A seguito di tutte queste attività, il set di dati è stato ridotto del 10%, raggiungendo un numero finale di commenti di 473.229.

Come passo successivo, sono stati estratti i topic (argomenti) associati ai commenti, sotto forma di parole chiave. Per ottenerli, abbiamo applicato *BERTopic* [19] sul nostro dataset; esso ha restituito un elenco di 247 topic univoci discussi nei commenti. Successivamente, i topic sono stati assegnati ai commenti in base al contenuto di questi ultimi.

Dopo aver completato questo task, abbiamo aggiunto, nel nostro set di dati, un nuovo campo, chiamato *topics*, il quale memorizza, per ogni commento, un elenco dei topic in esso discussi.

5.2 Analisi esplorativa

Per comprendere meglio il nostro dataset, abbiamo eseguito un'analisi esplorativa sui dati pre-elaborati, di cui forniamo, in questa sezione, i risultati ottenuti.

In primo luogo, sono state ricavate alcune statistiche di base relative al nostro set di dati; queste sono riportate nella Tabella 5.1.

<i>Proprietà</i>	<i>Valore</i>
Numero di commenti	473,229
Numero di autori	77,655
Numero di topic	247
Media del numero di parole per commento	11.95 parole
Mediana del numero di parole per commento	7 parole

Tabella 5.1: Statistiche di base del dataset proposto

Da questa tabella, risulta che il numero di commenti è molto maggiore di quello degli autori, il quale, a sua volta, è molto più alto del numero di argomenti. In particolare, il numero medio di commenti per autore è di 6.09 e il rapporto autori/argomenti è di circa 314. Ne consegue che la rete multilayer associata al set di dati presenta layer molto sbilanciati rispetto al numero di nodi, e questo deve essere preso in considerazione nelle prossime analisi (ad esempio, impostando opportunamente il parametro ρ nella definizione dell'indice di modularità multilayer Q_M).

Dopo aver ricavato le statistiche di base del nostro set di dati, abbiamo iniziato a svolgere l'analisi esplorativa dello stesso. Come primo task, è stato ricavato il trend temporale dei commenti. La Figura 5.1 mostra il numero di commenti pubblicati ogni giorno durante il periodo di raccolta dei dati; come ci si potrebbe aspettare, il numero di commenti raggiunge i valori più alti nei giorni in cui si disputano le partite, con il numero massimo di commenti raggiunto il giorno della finale del torneo. Chiaramente, dopo la finale, si registra una diminuzione significativa del numero di commenti.

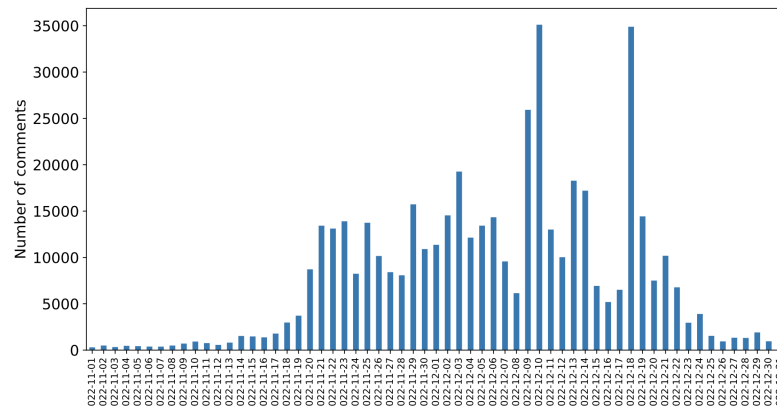


Figura 5.1: Trend temporale dei commenti nel dataset

L'analisi successiva riguarda l'attività degli utenti. La Figura 5.2 mostra la distribuzione dei commenti rispetto agli utenti in scala semi-logaritmica; dall'analisi di questa figura, possiamo vedere che la distribuzione corrispondente segue una power law. Nello specifico, la maggioranza degli utenti ha pubblicato un solo commento, mentre un piccolo numero di utenti è stato estremamente attivo e ha pubblicato un gran numero di commenti.

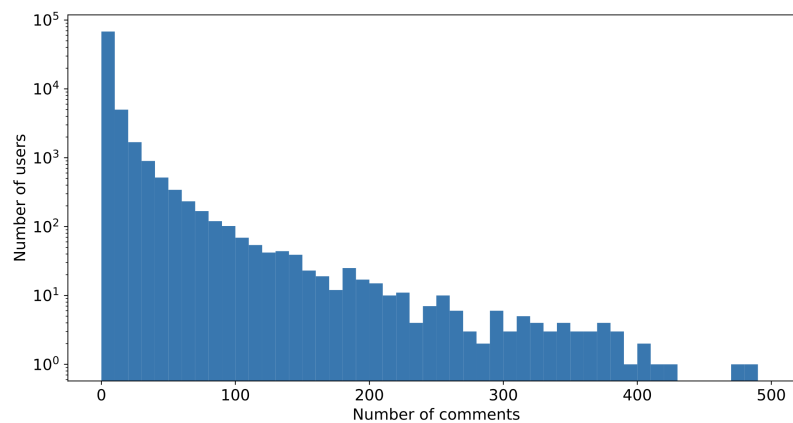


Figure 3: Distribution of comments against users (semi-log scale)

Figura 5.2: Distribuzione dei commenti rispetto agli utenti in scala semi-logaritmica

Successivamente, abbiamo anche condotto un'analisi relativa ai topic sul set di dati attraverso *BERTopic*. La Figura 5.3 mostra la distribuzione dei commenti rispetto agli argomenti in scala semi-logaritmica; anche in questo caso, si può notare che la distribuzione segue una power law. Abbiamo scoperto che l'argomento più discusso riguarda i risultati delle partite, seguito dalle squadre e dai giocatori. Inoltre, è da menzionare la presenza di argomenti relativi agli aspetti sociali e culturali della World Cup, come quelli riguardanti il paese ospitante e i tifosi.

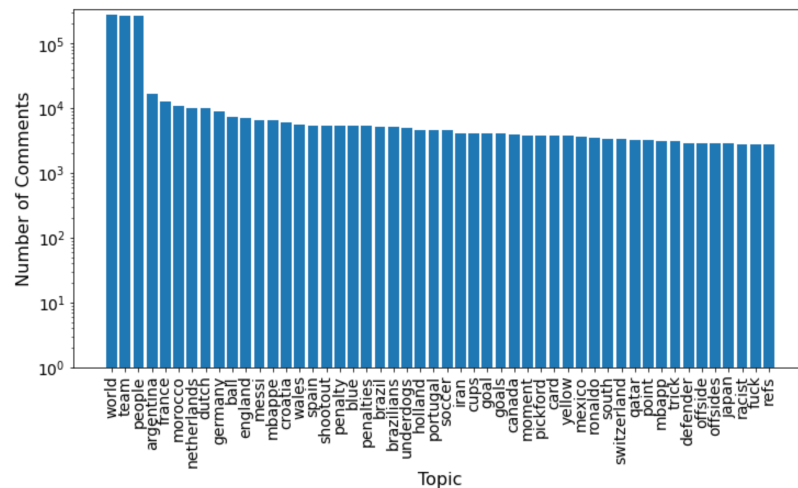


Figura 5.3: Distribuzione dei commenti rispetto agli argomenti in scala semi-logaritmica

5.3 Analisi strutturale

In questa sezione presentiamo gli esperimenti condotti in merito all'analisi strutturale del multilayer network associato al nostro set di dati. Nello specifico, consideriamo sia l'analisi intra-layer che quella inter-layer, calcolando i valori di assortatività di ogni layer e dell'intero multilayer network nel tempo.

Queste informazioni potrebbero essere utili per diverse applicazioni, come, ad esempio, lo studio della diffusione delle informazioni o la progettazione di interventi mirati nella rete.

5.3.1 Analisi intra-layer

L'analisi intra-layer riguarda lo studio delle proprietà strutturali di ogni layer del multilayer network. Come primo task, abbiamo calcolato il valore del degree-degree cross-correlation coefficient r_L per ogni layer del multilayer network nel tempo. I valori ottenuti sono mostrati nella Figura 5.4.

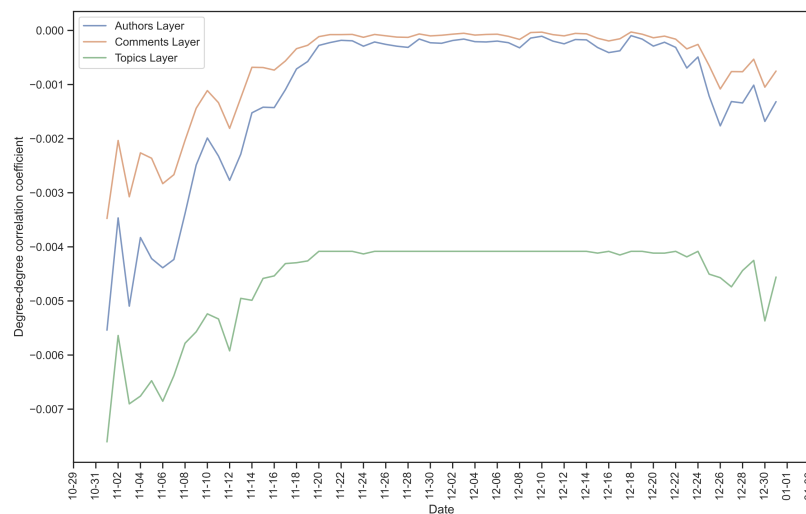


Figura 5.4: Valori di r_L per ogni layer nel tempo

In questa figura, l'andamento di r_L per il layer *Authors* è evidenziato in blu, quello per *Comments* in arancione e, infine, quello per *Topics* in verde.

Dall'analisi di questa figura osserviamo che i valori di r_L sono negativi e in aumento nel tempo; ciò implica che l'assortatività dei nodi di ogni layer della rete è negativa (e, quindi, siamo in presenza di disassortatività) e il suo valore assoluto diminuisce nel tempo. A sua volta, ciò implica che i nodi tendono ad essere collegati a nodi con gradi significativamente diversi.

Possiamo, inoltre, vedere che i valori di assortatività per i layer *Authors* e *Comments* sono relativamente vicini tra loro, il che indica che i nodi in questi layer mostrano modelli di connessione simili.

La presenza di disassortatività è più evidente nel layer *Topics* rispetto agli altri. Una possibile spiegazione di questo fenomeno può risiedere nel fatto che esistono archi tra due topic se uno appartiene alla stessa categoria o a una sottocategoria dell'altro. Ciò implica, quindi, che esistono archi tra topic generali (i quali hanno molti archi e un grado elevato) e topic molto specifici legati a sottocategorie (i quali, invece, hanno pochi archi).

Questo comportamento non si osserva in modo così marcato nei layer *Authors* e *Comments*, dove ogni nodo può interagire sia con nodi che hanno un grado simile, sia con nodi che hanno un grado molto diverso dal suo.

La tendenza alla diminuzione dei valori di disassortatività nel tempo suggerisce che la rete sta diventando meno modulare, cioè si sta evolvendo verso una struttura più casuale. Ciò potrebbe essere dovuto alla crescita della rete; infatti, con il progredire del torneo, sono apparsi nuovi autori, commenti e topic su Reddit e sono stati aggiunti nuovi nodi al multilayer network. L'aggiunta di nuovi autori, commenti e topic è avvenuta secondo schemi diversi da quelli originali. In particolare, le nuove modalità di crescita si sono rivelate più casuali e meno strutturate di quelli originali.

Infine, abbiamo visto che r_L si è dimostrato una misura efficace per distinguere i comportamenti dei diversi layer rispetto all'assortatività e per estrarre spunti di riflessione su questo tema.

Un'altra misura strutturale rilevante è il clustering coefficient γ_L , il quale, in un contesto intra-layer, misura la tendenza con cui i nodi dello stesso layer tendono a formare gruppi strettamente interconnessi. La Figura 5.5 mostra i valori del coefficiente γ_L per i livelli *Authors*, *Comments* e *Topics* nel tempo.

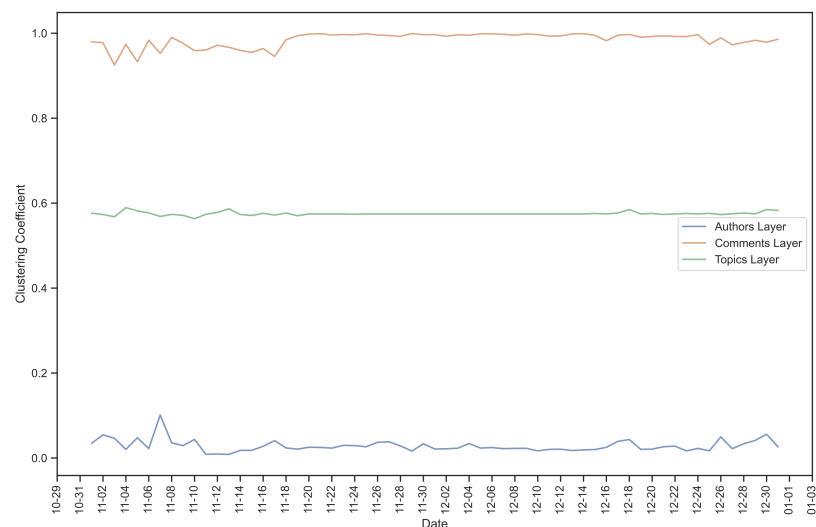


Figura 5.5: Valori di γ_L per ogni layer nel tempo

Dall'analisi di questa figura possiamo osservare che γ_L è più basso nel layer *Authors* rispetto agli altri due, il che suggerisce una minore tendenza degli autori a formare gruppi coesi e, quindi, a non formare comunità strettamente connesse. Ciò può essere attribuito all'ampia gamma di argomenti trattati dai vari autori, che porta ad una struttura di rete più disgiunta. Questa tendenza di γ_L riflette le diverse prospettive degli autori in questo subreddit.

Al contrario, il layer *Comments* ha il valore di γ_L più alto; ciò suggerisce che i commenti tendono a formare dei cluster interconnessi, probabilmente legati alla catena di commenti in cui uno viene pubblicato come risposta ad un altro.

Infine, il layer dei topic mostra un valore di γ_L intermedio rispetto a quello dei due layer precedenti. Questo può essere spiegato dal fatto che, come abbiamo visto, i commenti sono coesi e tendono, chiaramente, a indurre coesione tra gli argomenti; d'altra parte, gli autori mostrano interessi diversi facendo parte di più comunità ciascuna delle quali pubblica commenti che sono coesi ma diversi da quelli pubblicati da un'altra comunità. Quindi, le influenze contrastanti degli autori e dei commenti sui topic fanno sì che il valore di γ_L del layer *Topics* sia intermedio rispetto a quello degli altri due.

Anche in questo caso, il clustering coefficient si è rivelato una misura efficace per distinguere i comportamenti dei vari layer ed estrarre informazioni ad essi correlate.

Infine, consideriamo l'average path length \mathcal{L}_L , il quale può fornirci importanti indicazioni sulle interconnessioni dei nodi nei vari layer della rete. I valori di \mathcal{L}_L per ogni layer nel corso del tempo sono mostrati nella Figura 5.6.

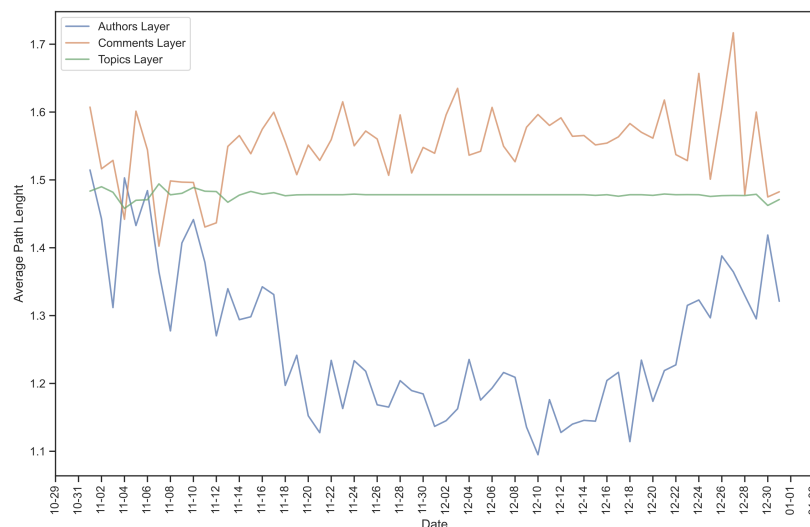


Figura 5.6: Valori di \mathcal{L}_L per ogni layer nel tempo

Dall'analisi di questa figura possiamo vedere che i valori di \mathcal{L}_L sono molto bassi in ogni layer; infatti, essi sono compresi nell'intervallo $[1.12, 1.51]$ per il layer *Authors*, nell'intervallo $[1.40, 1.72]$ per il layer *Comments* e nell'intervallo $[1.46, 1.58]$ per il layer *Topics*. Questo ci dice che gli autori, i commenti e i topic sono ben collegati tra loro nei layer corrispondenti, o direttamente o attraverso pochissimi intermediari. Di conseguenza, possiamo concludere che, nel subreddit da cui è stato ricavato il nostro dataset, la comunicazione è diretta e immediata e le informazioni fluiscono in modo efficiente.

Un secondo aspetto che emerge da questa figura è che il valore di \mathcal{L}_L è abbastanza costante nel tempo per i topic, mentre varia in modo più marcato per gli autori e i commenti. Ciò può essere spiegato considerando che i topic rimangono sostanzialmente invariati per tutto il periodo di interesse, subendo pochissime variazioni. Al contrario, gli autori e i commenti

variano notevolmente nel tempo, in quanto nuovi autori partecipano alle discussioni, altri non vi partecipano più e ogni giorno vengono pubblicati sempre nuovi commenti.

L'average path length, come gli altri due coefficienti, si è dimostrato efficace nell'evidenziare le somiglianze e le differenze nei comportamenti dei layer nonché nello scoprire approfondimenti di interesse.

5.3.2 Analisi inter-layer

L'analisi inter-layer riguarda lo studio delle proprietà strutturali del multilayer network nel suo complesso, considerando anche i nodi appartenenti a layer diversi.

Per questa analisi, abbiamo calcolato il valore del degree-degree cross-correlation coefficient r_{L_1, L_2} per le diverse coppie di layer del nostro multilayer network nel tempo. I valori ottenuti sono riportati nella Figura 5.7. In particolare, mostriamo i valori calcolati per la coppia *Authors-Topics* in verde, quelli per la coppia *Comments-Topics* in arancione e quelli per la coppia *Authors-Comments* in blu.

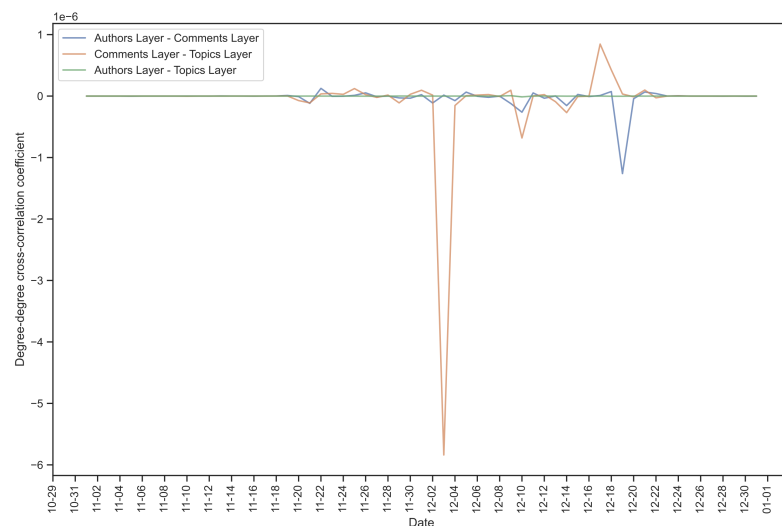


Figura 5.7: Valori di r_{L_1, L_2} per coppie di layer nel tempo

Da questa analisi possiamo osservare che r_{L_1, L_2} è essenzialmente uguale a 0 per tutte le coppie di layer e rimane costante nel tempo. Ciò significa che non c'è alcuna forma di assortatività o di disassortatività di grado tra autori e topic, tra commenti e topic e tra autori e commenti.

Ciò implica che, ad esempio, gli autori centrali di una comunità possono pubblicare sia commenti molto discussi che commenti poco discussi, e viceversa. Allo stesso modo, i commenti molto discussi possono riguardare sia topic generali che topic molto specifici, e viceversa. Infine, gli utenti centrali possono essere interessati sia a topic altamente correlati che a topic isolati, e viceversa.

In questo caso, r_{L_1, L_2} ci ha permesso di individuare la similarità di comportamento tra i tre layer della nostra rete.

5.4 Analisi delle comunità

In questa sezione presentiamo i nostri esperimenti sull'analisi delle comunità del multi-layer network associato al nostro dataset. Anche in questo caso, consideriamo sia l'analisi intra-layer che quella inter-layer.

5.4.1 Analisi intra-layer

L'analisi intra-layer si occupa dello studio delle proprietà delle comunità esistenti in ogni layer del multilayer network; a tale scopo, si utilizza l'indice di modularità Q_L . Un valore elevato di Q_L indica che le comunità sono ben definite all'interno del layer a cui appartengono; al contrario, un valore basso di Q_L indica che è difficile identificare comunità ben definite nei layer corrispondenti.

Abbiamo calcolato il valore di Q_L per i layer *Authors*, *Comments* e *Topics*; i risultati ottenuti sono mostrati nella Figura 5.8.

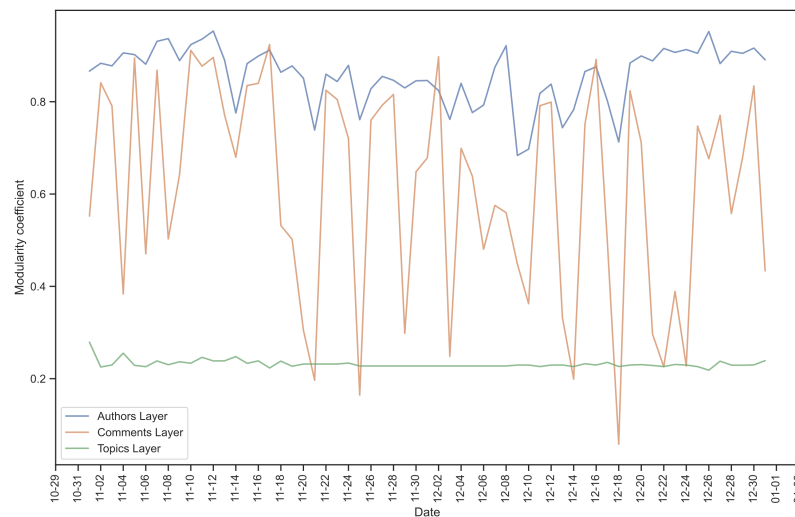


Figura 5.8: Valori di Q_L per ogni layer nel tempo

Analizzando la figura precedente possiamo ricavare alcune informazioni rilevanti. In primo luogo, osserviamo che il layer *Authors* ha un valore elevato di Q_L , con fluttuazioni limitate; ciò significa che gli autori tendono ad organizzarsi in comunità che sono piuttosto coese internamente e debolmente accoppiate tra loro.

All'altro estremo, abbiamo il layer *Topics*, che presenta valori bassi di Q_L , i quali rimangono abbastanza costanti nel tempo. Ciò implica che i topic sono difficili da raggruppare in comunità coese e la loro organizzazione rimane costante nel tempo. La ragione di ciò è la stessa che spiega l'andamento di \mathcal{L}_L per il layer *Topics* nella Figura 5.6: i topic sono stati generalmente definiti all'inizio del periodo di osservazione ed è improbabile che cambino nel tempo.

Infine, possiamo osservare che il valore di Q_L per il layer *Comments* è intermedio rispetto a quello associato agli altri due, con fluttuazioni molto ampie; ciò è conseguenza del fatto che la misura in cui i commenti possono essere organizzati in cluster coesi e debolmente accoppiati varia nel tempo.

In alcuni giorni osserviamo la presenza di picchi molto elevati di Q_L , paragonabili ai valori di Q_L relativi agli autori. Al contrario, in altri giorni, notiamo la presenza di picchi molto bassi di Q_L , paragonabili, o addirittura inferiori, ai valori di Q_L per i topic.

Concludiamo questa descrizione notando come l'indice di modularità Q_L sia stato efficace nel delineare il diverso comportamento dei layer rispetto all'esistenza di gruppi coesi di nodi al loro interno. Infatti, grazie ad esso, siamo stati in grado di ricavare utili approfondimenti su questo tema.

5.4.2 Analisi inter-layer

L'analisi inter-layer riguarda lo studio delle caratteristiche delle comunità formate da nodi appartenenti a layer diversi; per condurre questo studio, abbiamo utilizzato l'indice di modularità multilayer Q_M . Come nel caso di Q_L visto in precedenza, valori elevati di Q_M indicano la presenza di cluster di nodi appartenenti a layer diversi, i quali risultano essere fortemente coesi e debolmente accoppiati; l'inverso è vero per valori bassi di Q_M . Successivamente, abbiamo calcolato i valori di Q_M sul multilayer network associato al nostro set di dati; i risultati ottenuti sono mostrati nella Figura 5.9.

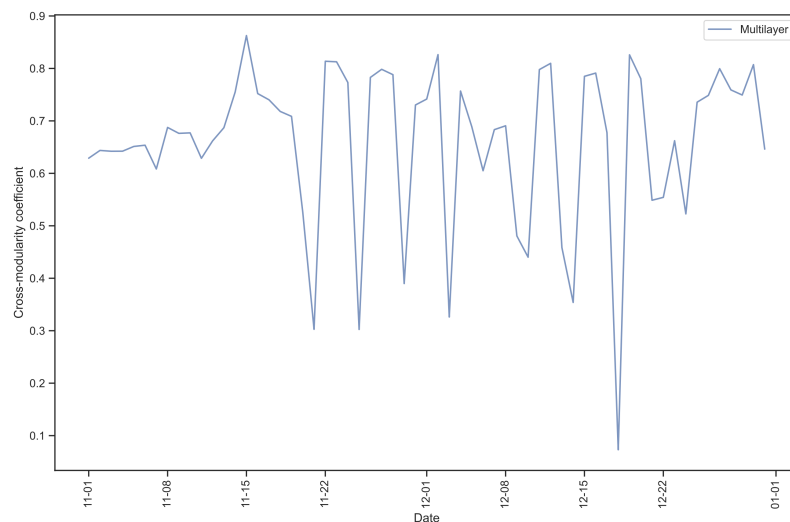


Figura 5.9: Valori di Q_M per ogni layer nel tempo

Dall'analisi di questa figura possiamo osservare che i valori di Q_M variano notevolmente e frequentemente nel tempo, da un massimo di 0.8626 ad un minimo di 0.0731; tuttavia, per la maggior parte del tempo, i valori di Q_M si assestano in un intervallo compreso tra 0.4 e 0.8. Ciò denota una certa tendenza, piuttosto debole, dei nodi di diversi layer a organizzarsi in cluster omogenei.

Possiamo ipotizzare che, in molti casi, gli utenti, insieme ai commenti che essi pubblicano e ai topic corrispondenti, siano propensi a organizzarsi nella stessa comunità. Tuttavia, questa tendenza non è stabile o forte. Ciò può essere abbastanza comprensibile se si considera che ci sono alcuni topic specifici (legati, ad esempio, a un giocatore o a una squadra), che possono essere racchiusi in un cluster inter-layer. Tuttavia, allo stesso tempo, ci sono altrettanti topic che sono così generici che è praticamente impossibile confinarli in un unico cluster inter-layer, dal momento che essi vengono discussi in molti commenti o sono postati da utenti appartenenti a comunità completamente diverse.

In ogni caso, anche l'indice di modularità multilayer Q_M si è rivelato utile per supportare l'analisi dell'evoluzione temporale del multilayer network, in particolare la capacità dei suoi nodi di organizzarsi in cluster fortemente coesi e debolmente accoppiati, comprendenti nodi appartenenti a layer diversi.

5.5 Analisi della resilienza

In questa sezione ci concentriamo sull'analisi della resilienza, effettuando, anche in questo caso, sia un'analisi intra-layer che un'analisi inter-layer.

Nonostante le possibili modifiche della rete su cui analizzare la resilienza siano numerose, per mancanza di spazio, ci concentriamo solo su alcune di esse, al fine di vedere se le nostre misure sono efficaci, cioè in grado di aiutarci ad estrarre informazioni rilevanti relative alla resilienza del multilayer network in esame.

5.5.1 Analisi intra-layer

L'analisi intra-layer si occupa dello studio della resilienza in ogni layer del multilayer network. In particolare, in questa sezione, valutiamo le misure di resilienza prese in esame, applicandole alla nostra rete solo in presenza di rimozioni di nodi e di archi.

Rimozione di nodi esistenti e dei rispettivi archi

Per quanto riguarda la rimozione di nodi esistenti e dei rispettivi archi, dobbiamo considerare sia la Clustering Coefficient Variation $\Delta\gamma_L[\tau_1, \tau_2]$ che la Level-wise Degree Distribution Variation $\Delta D_L[\tau_1, \tau_2]$. In questa sottosezione ci concentriamo unicamente su quest'ultima misura, mentre utilizzeremo la prima negli esperimenti della prossima sottosezione.

Durante gli esperimenti abbiamo rimosso in modo casuale il 5%, il 10%, il 20%, il 30%, il 40% e il 50% dei nodi e dei rispettivi archi da ogni layer, per poi calcolare $\Delta D_L[\tau_1, \tau_2]$. Abbiamo ripetuto questo task cinque volte e abbiamo calcolato i valori medi. Nelle Figure 5.10 - 5.12 mostriamo i risultati ottenuti per i layer *Authors*, *Comments* e *Topics* nel tempo.

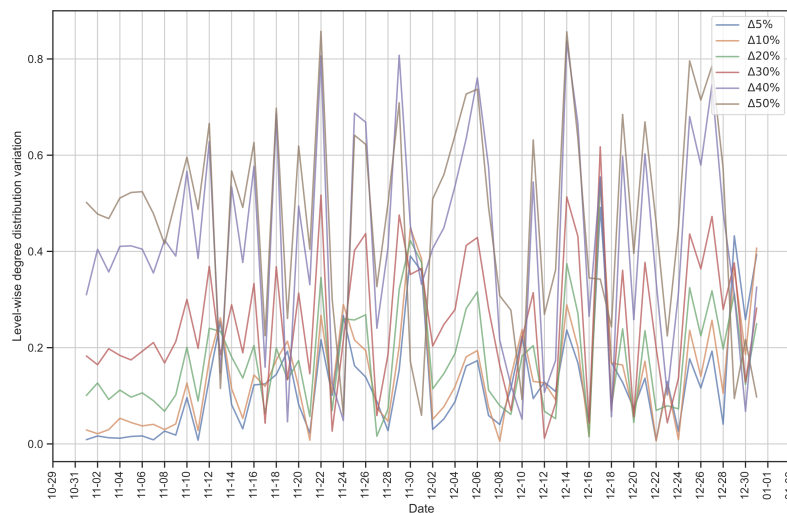


Figura 5.10: Valori di $\Delta D_L[\tau_1, \tau_2]$ per il layer *Authors* nel tempo

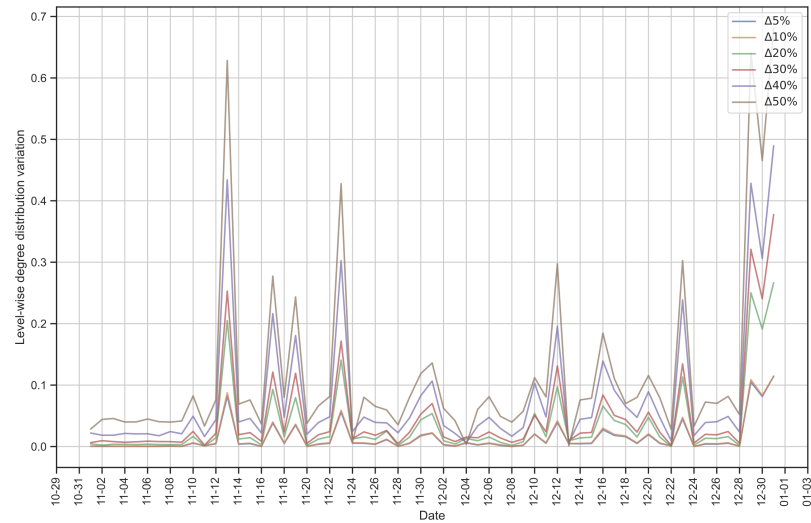


Figura 5.11: Valori di $\Delta D_L[\tau_1, \tau_2]$ per il layer *Comments* nel tempo

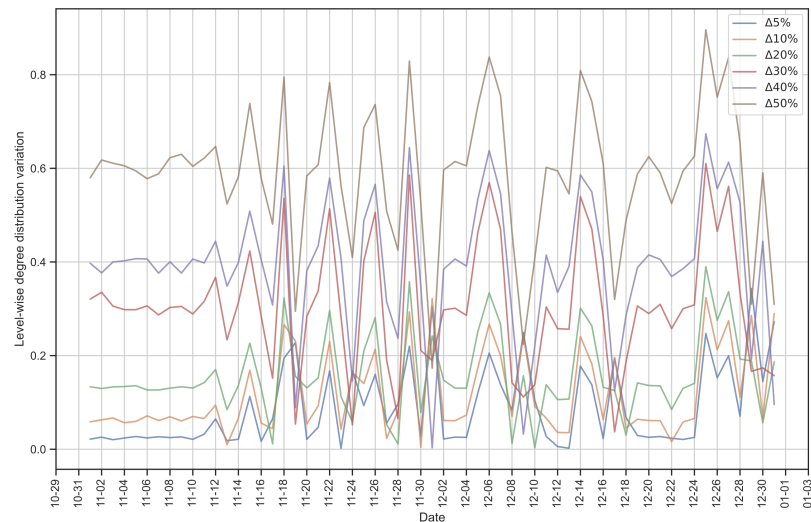


Figura 5.12: Valori di $\Delta D_L[\tau_1, \tau_2]$ per il layer *Topics* nel tempo

Dall'analisi di queste figure, possiamo trarre alcuni spunti interessanti in merito alla resilienza dei tre layer. In primo luogo, osserviamo che, in tutti e tre i casi, è presente un andamento molto irregolare, con la presenza di variazioni e di picchi molto forti.

Nello specifico, per quanto riguarda i valori di resilienza, questi sono generalmente più bassi per il layer *Comments* rispetto agli altri due. Inoltre, in tutti e tre i layer, possiamo osservare che, finché la percentuale di nodi rimossi è bassa (cioè fino al 20%), le variazioni di $\Delta D_L[\tau_1, \tau_2]$ sono piccole, il che indica una resilienza soddisfacente in tutti e tre i casi. Quando la percentuale di nodi rimossi aumenta ulteriormente, notiamo che le variazioni di $\Delta D_L[\tau_1, \tau_2]$ crescono in modo significativo, raggiungendo picchi di 0.8 (ricordiamo che il valore massimo possibile è 1).

La validità di questo risultato è dovuta al fatto che ogni layer dovrebbe essere capace di riconoscere la presenza di perturbazioni in grado di sconvolgere la struttura e il comportamento della rete sottostante. Da questo punto di vista, i layer *Authors* e *Topics* mostrano le migliori prestazioni complessive. Le prestazioni di *Comments* sono soddisfacenti, ma denotano una certa difficoltà nel riconoscere perturbazioni anche sostanziali della rete sottostante il

layer.

Da questi risultati emerge che la Level-wise Degree Distribution Variation $\Delta D_L[\tau_1, \tau_2]$ è stata in grado di darci un'idea molto chiara della resilienza dei vari layer, in presenza di rimozioni di nodi e degli archi a essi associati.

Rimozione di nodi esistenti

Per quanto riguarda la rimozione degli archi esistenti, per motivi di spazio, mostriamo solo l'analisi relativa alla Clustering Coefficient Variation $\Delta\gamma_L[\tau_1, \tau_2]$. Nello specifico, durante gli esperimenti, sono stati rimossi, in modo casuale, il 5%, il 10%, il 20%, il 30%, il 40% e il 50% degli archi da ogni layer, per poi calcolare $\Delta\gamma_L[\tau_1, \tau_2]$. Abbiamo ripetuto questo task cinque volte, per poi calcolare i valori medi. Nelle Figure 13-15 mostriamo i risultati ottenuti per i layer *Authors*, *Comments* e *Topics* nel tempo.

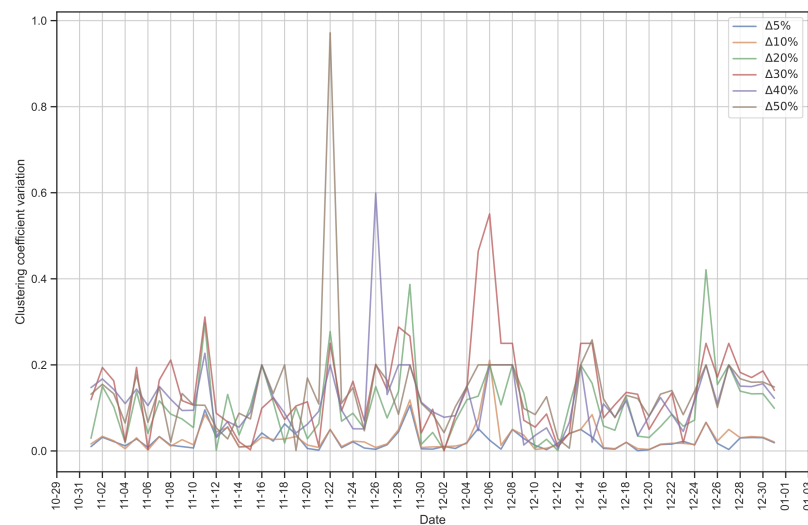


Figura 5.13: Valori di $\Delta\gamma_L[\tau_1, \tau_2]$ per il layer *Authors* nel tempo

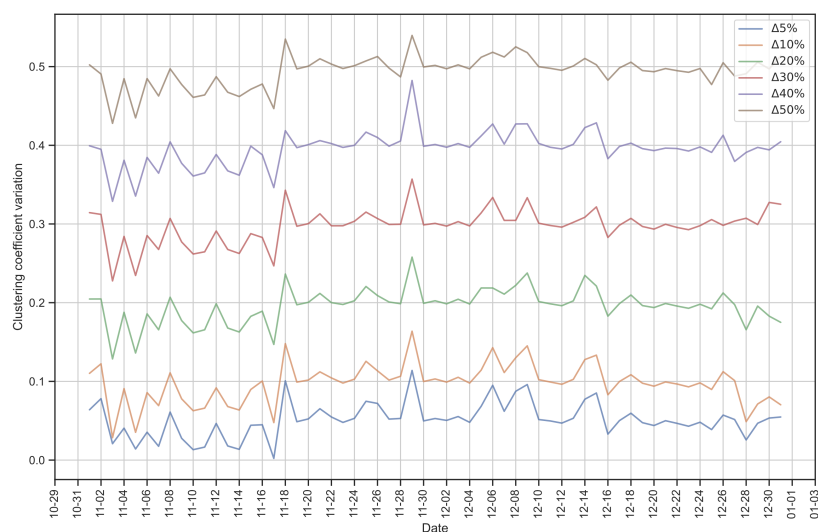


Figura 5.14: Valori di $\Delta\gamma_L[\tau_1, \tau_2]$ per il layer *Comments* nel tempo

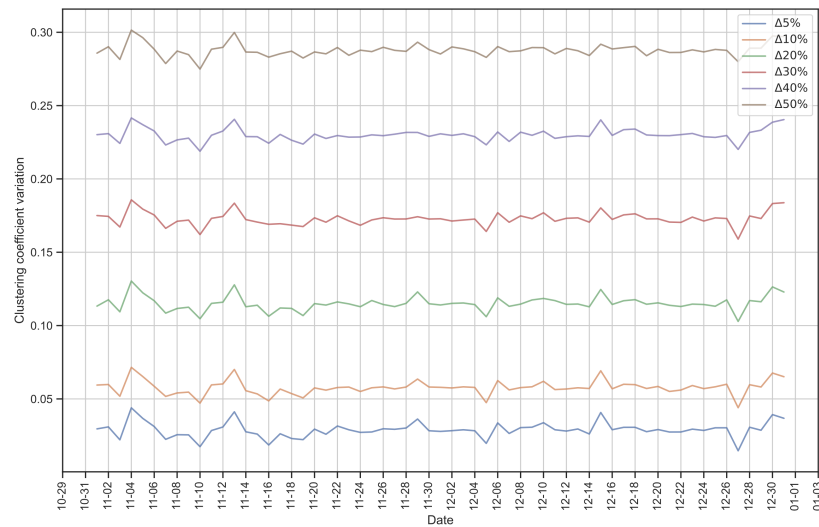


Figura 5.15: Valori di $\Delta\gamma_L[\tau_1, \tau_2]$ per il layer *Topics* nel tempo

Dall'analisi di queste figure, possiamo vedere che il layer *Authors* ha un comportamento completamente diverso rispetto ai layer *Comments* e *Topics*; in particolare, è molto meno resiliente degli altri due quando la frazione di archi rimossi è bassa. Quando tale frazione diventa molto alta, invece, il layer *Authors* si dimostra molto più resiliente.

Inoltre, l'andamento di $\Delta\gamma_L[\tau_1, \tau_2]$ per *Authors* è molto più irregolare rispetto agli altri due layer; infatti, possiamo notare la presenza di variazioni brusche e di alcuni picchi pronunciati in corrispondenza delle riduzioni più elevate del numero di archi.

Gli altri due layer mostrano tendenze molto più regolari e sono molto resilienti quando la percentuale di archi rimossi è bassa; tuttavia, essi risultano molto meno resilienti quando tale percentuale è alta. A nostro avviso, questo è il comportamento ideale che un layer dovrebbe avere, ovvero non dovrebbe subire variazioni enormi per la perdita di pochi archi, mentre dovrebbe mostrare campanelli d'allarme quando tale perdita inizia a diventare sostanziale.

Per quanto riguarda l'efficacia della Clustering Coefficient Variation $\Delta\gamma_L[\tau_1, \tau_2]$, osserviamo che anche questa metrica è stata in grado di darci un'idea molto chiara in merito alla differenza nel comportamento della resilienza dei vari layer, in presenza di rimozioni di archi.

5.5.2 Analisi inter-layer

L'analisi inter-layer si occupa dello studio della resilienza dell'intero multilayer network. In questa sezione, valutiamo le nostre misure di resilienza applicate al multilayer network derivato dal nostro set di dati, considerando solo le rimozioni di nodi e archi.

Rimozione di nodi esistenti e dei rispettivi archi

Per quanto riguarda la rimozione di nodi e degli archi corrispondenti, dobbiamo considerare i tre coefficienti previsti per lo studio della resilienza in questo ambito, ossia la Modularity Variation $\Delta Q_M[\tau_1, \tau_2]$, l'Inter-layer Correlation Variation $\Delta r_{L_1, L_2}[\tau_1, \tau_2]$ e la Multi-layer Degree Distribution Variation $\Delta D_E[\tau_1, \tau_2]$. Per motivi di spazio, ci concentriamo solo su $\Delta r_{L_1, L_2}[\tau_1, \tau_2]$ e $\Delta D_E[\tau_1, \tau_2]$, mentre utilizzeremo $\Delta Q_M[\tau_1, \tau_2]$ nell'analisi descritta nella prossima sezione.

Durante gli esperimenti abbiamo rimosso, in modo casuale, il 5%, il 10%, il 20%, il 30%, il 40% e il 50% dei nodi e dei rispettivi inter- e intra-archi, per poi calcolare il valore di

$\Delta r_{L_1, L_2}[\tau_1, \tau_2]$ per i layer *Authors* e *Comments* nel tempo. Abbiamo ripetuto queste misurazioni cinque volte e abbiamo calcolato i valori medi. Nella Figura 5.16 mostriamo i risultati ottenuti.

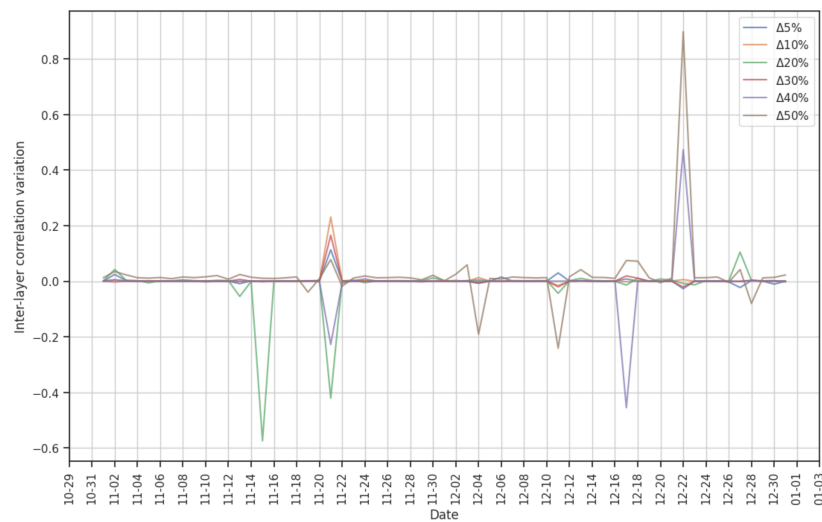


Figura 5.16: Valori di $\Delta r_{L_1, L_2}[\tau_1, \tau_2]$ per i layer *Authors* e *Comments* nel tempo

Dall’analisi di questa figura, possiamo osservare che $\Delta r_{L_1, L_2}[\tau_1, \tau_2]$ è quasi sempre nullo, anche in presenza di perturbazioni sostanziali, e che, quindi, il multilayer network e il multimodal social network corrispondente sono resilienti alla rimozione dei nodi e dei corrispondenti archi intra- e inter-layer. Ciò significa che, generalmente, non c’è assortatività o disassortatività di grado tra i nodi dei vari layer. Per questo motivo, ad esempio, un utente molto famoso potrebbe aver pubblicato sia commenti molto dibattuti che commenti che non hanno suscitato alcun interesse, e viceversa.

In presenza di perturbazioni molto grandi, è possibile osservare l’esistenza di alcuni picchi di assortatività o disassortatività, i quali durano per un tempo molto breve. I picchi di assortatività sono più frequenti, ma in media più piccoli, di quelli di disassortatività. L’unica eccezione è il picco di assortatività molto elevato verificatosi il 22 dicembre 2022.

Infine, vale la pena sottolineare il fatto che, nonostante i risultati mostrati nella Figura 5.16 sono più frequenti, ma in media più piccoli, di quelli di disassortatività, esiste una forte somiglianza tra i risultati dell’assortatività inter-layer mostrati nella Figura 5.16 e quelli dell’assortatività intra-layer riportati nella Figura 5.7.

Vogliamo ancora una volta evidenziare che, al di là dello specifico dataset e dei dati in esso memorizzati, l’indice $\Delta r_{L_1, L_2}[\tau_1, \tau_2]$ introdotto nel nostro framework si è dimostrato efficace nell’estrarre intuizioni potenzialmente interessanti sulla resilienza di un multilayer network in presenza di perturbazioni inter-layer.

Procedendo allo stesso modo e considerando le stesse perturbazioni del caso precedente, abbiamo calcolato il valore di $\Delta D_E[\tau_1, \tau_2]$ nel tempo. I risultati ottenuti sono mostrati nella Figura 5.17.

Dall’analisi di questa figura possiamo osservare che il multilayer network associato al nostro set di dati presenta un comportamento ottimale per quanto riguarda la resilienza. Infatti, in presenza di un basso numero di perturbazioni, il valore di $\Delta D_E[\tau_1, \tau_2]$ è piccolo. Al contrario, quando il numero di perturbazioni diventa elevato, il valore di $\Delta D_E[\tau_1, \tau_2]$ diventa molto grande, agendo come un campanello d’allarme.

Quindi, possiamo concludere che anche $\Delta D_E[\tau_1, \tau_2]$ è notevolmente efficace nell’estrarre intuizioni potenzialmente interessanti sulla resilienza di un multilayer network in presenza di perturbazioni inter-layer.

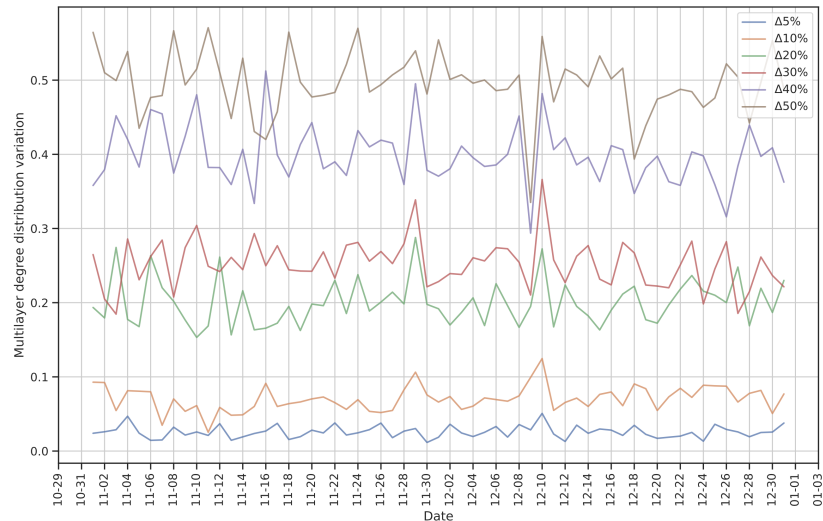


Figura 5.17: Valori di $\Delta D_E[\tau_1, \tau_2]$ nel tempo

Rimozione di nodi esistenti

Per quanto riguarda la rimozione degli archi inter-layer, per motivi di spazio, in questa sezione mostriamo unicamente le analisi eseguite per la Modularity Variation $\Delta Q_M[\tau_1, \tau_2]$ e la Multilayer Degree Distribution Variation $\Delta D_E[\tau_1, \tau_2]$.

Durante gli esperimenti abbiamo rimosso in modo casuale il 5%, il 10%, il 20%, il 30%, il 40% e il 50% degli archi inter-layer e abbiamo calcolato il valore di $\Delta Q_M[\tau_1, \tau_2]$ e $\Delta D_E[\tau_1, \tau_2]$ nel tempo. L'andamento di $\Delta Q_M[\tau_1, \tau_2]$ è mostrato nella Figura 5.18, mentre quello di $\Delta D_E[\tau_1, \tau_2]$ è riportato nella Figura 5.19.

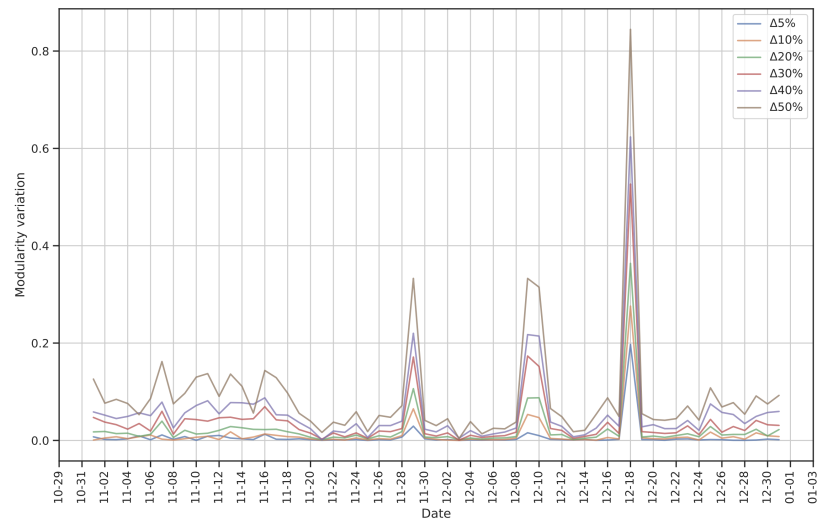


Figura 5.18: Valori di $\Delta Q_M[\tau_1, \tau_2]$ nel tempo

Dall'analisi della Figura 5.18 possiamo osservare che il multilayer network associato al nostro set di dati si dimostra resiliente rispetto alle perturbazioni legate alla rimozione degli archi inter-layer. In un certo senso, anche la resilienza evidenziata da $\Delta Q_M[\tau_1, \tau_2]$ appare eccessiva, poiché anche le perturbazioni significative causano cambiamenti limitati in questa misura. Possiamo osservare alcuni picchi che, tuttavia, scompaiono entro il giorno successivo.

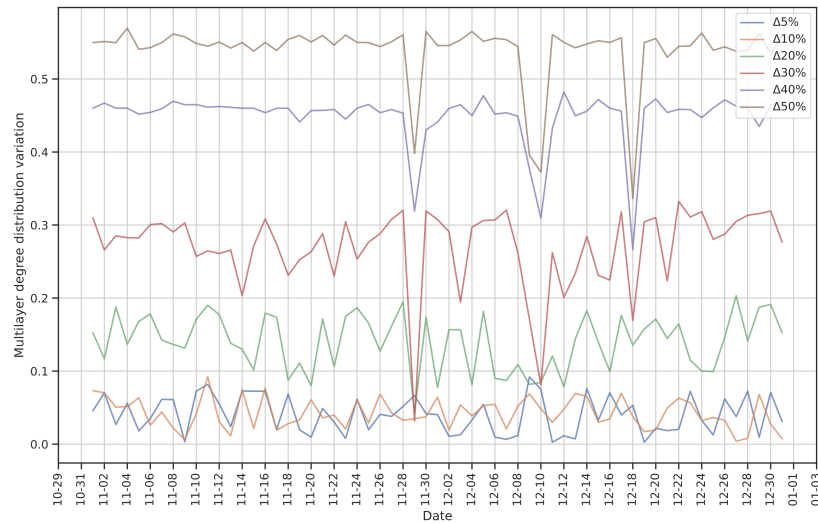


Figura 5.19: Valori di $\Delta D_E[\tau_1, \tau_2]$ nel tempo

D'altra parte, l'analisi della Figura 5.19 mostra un comportamento ottimale del nostro multilayer network rispetto alla resilienza, poiché una piccola riduzione del numero di archi inter-layer porta a una variazione limitata del valore di $\Delta D_E[\tau_1, \tau_2]$. Al contrario, quando il numero di archi inter-layer rimossi inizia a diventare elevato, il valore di $\Delta D_E[\tau_1, \tau_2]$ tende a variare in modo sostanziale rispetto al suo valore originale. Anche in questo caso, tale variazione può essere utilizzata come un campanello d'allarme per segnalare la presenza di perturbazioni eccessive che potrebbero disarticolare la rete originale.

Infine, anche per la rimozione degli archi esistenti, i due coefficienti $\Delta Q_M[\tau_1, \tau_2]$ e $\Delta D_E[\tau_1, \tau_2]$ si sono dimostrati efficaci nel supportare l'analisi della resilienza di un multilayer network e nel ricavare approfondimenti ad essa correlati.

5.6 Discussione sui risultati ottenuti

Giunti alla fine di questi esperimenti, sono necessarie alcune considerazioni sul lavoro svolto e sui risultati ottenuti.

Innanzitutto, sottolineiamo nuovamente l'obiettivo principale di questo studio, ovvero la definizione di un framework per analizzare l'evoluzione temporale e la resilienza di un multilayer network e di un multimodal social network modellato attraverso di esso. In questo contesto, il multilayer network e il multimodal social network associato, costruiti a partire da un dataset sul torneo calcistico della FIFA World Cup, devono essere intesi come strumenti per i nostri esperimenti e non come il nostro obiettivo finale. In altre parole, l'individuazione di approfondimenti sull'evoluzione temporale e sulla resilienza di queste due reti specifiche rappresenta un risultato collaterale, anche se molto interessante.

Il nostro obiettivo principale è quello di capire se il framework proposto e le sue misure di analisi siano adeguate per valutare l'evoluzione temporale e la resilienza di un multilayer network e, quindi, del multimodal social network a esso associato.

Da questo punto di vista, i risultati ottenuti sono promettenti. Infatti, le misure di analisi associate al nostro framework si sono dimostrate in grado di: (i) evidenziare la presenza di comportamenti diversi da parte dei vari layer; (ii) individuare la presenza di assortatività, disassortatività o l'assenza di entrambe, sia nei singoli layer che nel multilayer network nel suo insieme; (iii) capire se le informazioni fluiscono bene all'interno di ogni layer e nel multilayer network complessivo; (iv) verificare se sono presenti o meno comunità ben

definite all'interno di ogni layer e nella totalità del multilayer network; (v) capire se un layer e il multilayer network nel suo complesso sono resilienti a varie forme di perturbazioni; (vi) capire se il multilayer network, nonostante la sua resilienza, è comunque in grado di riconoscere la presenza di perturbazioni eccessive che possono disturbare uno dei suoi layer, una coppia di essi o il multilayer network nel suo complesso.

Per quanto riguarda l'evoluzione temporale e, soprattutto, la resilienza, alcune delle misure di analisi incluse nel nostro framework si sono rivelate più appropriate in alcuni casi, mentre altre si sono rivelate più adeguate in altri.

Possiamo, dunque, concludere che il framework, con il suo insieme di misure, si è dimostrato molto efficace nel gestire sia l'evoluzione temporale che la resilienza di un multilayer network e del multimodal social network a esso associato; non abbiamo, infatti, incontrato un solo caso in cui almeno una (e quasi sempre più di una) delle misure di analisi proposte nel nostro studio non si sia dimostrata totalmente efficace per gli obiettivi preposti.

Nella presente tesi è stato sviluppato e testato un framework basato su un multilayer network per studiare l'evoluzione temporale e la resilienza dei multimodal social network. In particolare, abbiamo dimostrato che i multilayer network sono uno strumento potente ed efficace per modellare tutte le caratteristiche di un multimodal social network, ovvero un social network con diverse tipologie di nodi e archi.

Inizialmente, abbiamo definito il modello per la mappatura di un multimodal social network in un multilayer network, con un layer per ogni tipo di nodo e dove archi inter-layer modellano i collegamenti tra i nodi di tipi diversi. Successivamente, abbiamo descritto i dettagli tecnici del nostro framework, il quale è basato su tre tipi di misure; le prime due riguardano la struttura della rete e delle comunità e sono utilizzate per studiare l'evoluzione temporale; il terzo tipo, invece, riguarda la resilienza della rete. Infine, abbiamo presentato una campagna sperimentale con l'obiettivo di dimostrare l'efficacia del framework proposto nel raggiungere gli obiettivi per cui è stato progettato.

In sintesi, i principali contributi di questo studio sono: (i) la definizione di un framework e di una serie di misure per studiare l'evoluzione temporale e la resilienza di un multilayer network; (ii) la definizione di un metodo per "mappare" un multimodal social network in un multilayer network; (iii) la definizione di un framework e di una serie di metriche per studiare l'evoluzione temporale e la resilienza di un multimodal social network.

Questo studio non deve essere considerato come un punto di arrivo, ma piuttosto come un punto di partenza per ulteriori ricerche su questo tema. In particolare, sarebbe possibile estendere il nostro framework per rappresentare e gestire le relazioni non diadiche, ossia le interazioni che coinvolgono più di due nodi, fornendo un supporto per layer che rappresentano hypernetwork. Un secondo aspetto da affrontare è il dinamismo degli utenti, ossia capire quanto spesso gli utenti entrano ed escono dalla rete sottostante, lasciano le comunità a cui appartenevano, si uniscono alle comunità esistenti o cercano di costruirne di nuove. Infine, il nostro framework potrebbe essere utilizzato per indagare le dinamiche di coinvolgimento legate ai vari topic; in altre parole, vorremmo capire come un topic può diventare attraente per gli utenti, o, al contrario, cessare di esserlo, o cessare di essere attraente per una categoria di utenti ma iniziare ad esserlo per un'altra categoria.

- [1] A. Amelio, G. Bonifazi, F. Cauteruccio, E. Corradini, M. Marchetti, D. Ursino, and L. Virgili. Representation and Compression of Residual Neural Networks through a Multilayer Network Based Approach. *Expert Systems With Applications*, 215(119391), 2023. Elsevier.
- [2] K.E. Anderson. Ask me anything: what is Reddit? *Library Hi Tech News*, 32(5):8–11, 2015. Emerald.
- [3] O. Artime, B. Benigni, G. Bertagnolli, V. d’Andrea, R. Gallotti, A. Ghavasieh, S. Raimondo, and M. De Domenico. Multilayer Network Science: From Cells to Societies. *Elements in Structure and Dynamics of Complex Networks*, 2022. Cambridge University Press.
- [4] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn. The pushshift Reddit dataset. In *Proc. of the International AAAI Conference on Web and Social Media (ICWSM’20)*, volume 14, pages 830–839, Atlanta, GA, USA, 2020. AAAI Press.
- [5] P.V. Bindu, P.S. Thilagam, and D. Ahuja. Discovering suspicious behavior in multilayer social networks. *Computers in Human Behavior*, 73:568–582, 2017. Elsevier.
- [6] V.D. Blondel, J.L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008. IOP Publishing.
- [7] G. Bonifazi, B. Breve, S. Cirillo, E. Corradini, and L. Virgili. Investigating the COVID-19 vaccine discussions on Twitter through a multilayer network-based approach. *Information Processing & Management*, 59(6):103095, 2022. Elsevier.
- [8] G. Bonifazi, F. Cauteruccio, E. Corradini, M. Marchetti, D. Ursino, and L. Virgili. Applying Social Network Analysis to Model and Handle a Cross-Blockchain Ecosystem. *Electronics*, 12(1086), 2023. MDPI.
- [9] M. Bujak and R. Kucharski. Network structures of urban ride-pooling problems and their properties. *Social Network Analysis and Mining*, 13(1):89, 2023. Springer.
- [10] C. Buntain and J. Golbeck. Identifying Social Roles in Reddit Using Network Structure. In *Proc. of the International Conference on World Wide Web (WWW’14)*, page 615–620, Seoul, Korea, 2014. ACM.
- [11] X. Chen and Z.M. Lu. Measure of layer centrality in multilayer network. *International Journal of Modern Physics C*, 29(06):1850051, 2018. World Scientific.

-
- [12] M.M. Danziger, I. Bonamassa, S. Boccaletti, and S. Havlin. Dynamic interdependence and competition in multilayer networks. *Nature Physics*, 15(2):178–185, 2019. Nature Publishing Group UK London.
- [13] F. Gamgne Domgue, N. Tsopze, and R. Ndoundam. Correlation and dimension relevance in multidimensional networks: a systematic taxonomy. *Social Network Analysis and Mining*, 11:1–19, 2021. Springer.
- [14] W.B. Du, X.L. Zhou, O. Lordan, Z. Wang, C. Zhao, and Y.-B. Zhu. Analysis of the Chinese Airline Network as multi-layer networks. *Transportation Research Part E: Logistics and Transportation Review*, 89:108–116, 2016. Elsevier.
- [15] G. Dudek. Multilayer perceptron for short-term load forecasting: from global to local approach. *Neural Computing and Applications*, 32(8):3695–3707, 2020. Springer.
- [16] S. El-Sappagh, J.M. Alonso, S.M. Islam, A.M. Sultan, and K.S. Kwak. A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer’s disease. *1*, 11(1):1–26, 2021. Springer.
- [17] D.N. Fisher and N.Pinter-Wollman. Using multilayer network analysis to explore the temporal dynamics of collective behavior. *Current zoology*, 67(1):71–80, 2021. Oxford University Press.
- [18] S. Gao and Z. Fang. Multilayer feature-rich satellite network analysis: An application-oriented and time-evolving approach. *IEEE Transactions on Network Science and Engineering*, 8(2):1316–1327, 2020. IEEE.
- [19] M. Grootendorst. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*, 2022. arXiv.
- [20] Y. Gu, X. Fu, Z. Liu, X. Xu, and A. Chen. Performance of transportation network under perturbations: Reliability, vulnerability, and resilience. *Transportation Research Part E: Logistics and Transportation Review*, 133:101809, 2020. Elsevier.
- [21] C. Gunaratne, S.K. Ray, A.C. Lourencço Alves, and M. Waldl. Exogenous Shocks Lead to Increased Responsiveness and Shifts in Sentimental Resilience in Online Discussions. In *Proc. of the International Conference of The Computational Social Science Society of the Americas (CSS’19)*, pages 57–71, Santa Fe, New Mexico, USA, 2021. Springer.
- [22] Z. Hammoud and F. Kramer. Multilayer networks: aspects, implementations, and application in biomedicine. *Big Data Analytics*, 5(1):2, 2020. Springer.
- [23] R. Interdonato, M. Atzmueller, S. Gaito, R. Kanawati, C. Largeron, and A. Sala. Feature-rich networks: going beyond complex network topologies. *Applied Network Science*, 4:1–13, 2019. Springer.
- [24] X. Kong, Y. Shi, S. Yu, J. Liu, and F. Xia. Academic social networks: Modeling, analysis, mining and applications. *Journal of Network and Computer Applications*, 132:86–103, 2019. Elsevier.
- [25] Q. Li, S. Dong, and A. Mostafavi. Modeling of inter-organizational coordination dynamics in resilience planning of infrastructure systems: A multilayer network simulation framework. *PloS one*, 14(11):e0224522, 2019. Public Library of Science San Francisco.

-
- [26] A.P. Logan, P.M. LaCasse, and B.J. Lunday. Social network analysis of Twitter interactions: a directed multilayer network approach. *Social Network Analysis and Mining*, 13(1):65, 2023. Springer.
- [27] F. Long, N. Ning, Y. Zhang, C. Song, P. Zhou, and B. Wu. Mining latent academic social relationships by network fusion of multi-type data. *Social Network Analysis and Mining*, 10:1–16, 2020. Springer.
- [28] D. Mohapatra. A hybrid approach for pair-wise layer similarity in a multiplex network. *Social Network Analysis and Mining*, 11(1):88, 2021. Springer.
- [29] P.J. Mucha, T. Richardson, K. Macon, M.A. Porter, and J. Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878, 2010. AAAS.
- [30] M.E. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006. National Academy of Sciences.
- [31] M.E.J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20):208701, 2002. APS.
- [32] E. Oro, C. Pizzuti, N. Procopio, and M. Ruffolo. Detecting topic authoritative social media users: a multilayer network approach. *IEEE Transactions on Multimedia*, 20(5):1195–1208, 2017. IEEE.
- [33] F. Pierri, C. Piccardi, and S. Ceri. A multi-layer approach to disinformation detection in US and Italian news spreading on Twitter. *EPJ Data Science*, 9(1):35, 2020. Springer.
- [34] M. Rosvall, D. Axelsson, and C. T. Bergstrom. The map equation. *The European Physical Journal Special Topics*, 178(1):13–23, 2009. Springer.
- [35] S. Boccaletti and G. Bianconi and R. Criado and C. I. Del Genio and J. Gómez-Gardenes and M. Romance and I. Sendina-Nadal and Z. Wang and M. Zanin. The structure and dynamics of multilayer networks. *Physics reports*, 544(1):1–122, 2014. Elsevier.
- [36] L. Scabini, L.C. Ribas, M.B. Neiva, A.G.B. Junior, A.J.F. Farfan, and O.M. Bruno. Social interaction layers in complex networks for the dynamical epidemic modeling of COVID-19 in Brazil. *Physica A: Statistical Mechanics and its Applications*, 564:125498, 2021. Elsevier.
- [37] J. Shin, L. Jian, K. Driscoll, and F. Bar. The diffusion of misinformation on social media: Temporal pattern, message, and source. *Computers in Human Behavior*, 83:278–287, 2018. Elsevier.
- [38] Q. Su, A. McAvoy, Y. Mori, and J.B. Plotkin. Evolution of prosocial behaviours in multilayer populations. *Nature Human Behaviour*, 6(3):338–348, 2022. Nature Publishing Group UK London.
- [39] F. Touvet and D. Harle. Network resilience in multilayer networks: A critical review and open issues. In *Proc. of the International Conference on Networking (Networking—ICN 2001)*, pages 829–837, Colmar, France, 2001. Springer.
- [40] K. S. Trivedi, D. S. Kim, and R. Ghosh. Resilience in computer systems and networks. In *Proc. of the International Conference on Computer-Aided Design (ICCAD’09)*, pages 74–77, San Jose, CA, USA, 2009. ACM.

- [41] M. Tsvetovat and A. Kouznetsov. *Social Network Analysis for Startups: Finding connections on the social web*. Sebastopol, CA, USA, 2011. O'Reilly Media, Inc.
- [42] M. Vaiana and S.F. Muldoon. Multilayer brain networks. *Journal of Nonlinear Science*, 30(5):2147–2169, 2020. Springer.
- [43] H. Zhang, C.D. Wang, J.H. Lai, and P.S. Yu. Modularity in complex multilayer networks with multiple aspects: a static perspective. *Applied Informatics*, 4(1):1–29, 2017. Springer.

Ringraziamenti

Al professor Domenico Ursino, a cui sono grato per avermi reso partecipe di questo progetto di ricerca e, ancora di più, per avermi dimostrato che la gentilezza, la disponibilità e il rispetto sono requisiti imprescindibili di un bravo ingegnere.

A Michele Marchetti, per avermi accompagnato in ogni fase del lavoro, con la massima attenzione e cortesia.

Alla mia famiglia, ai miei parenti e a Benedetta, per aver condiviso con me tutte le fatiche e le gioie di questo cammino, sostenendomi dal primo all'ultimo momento.

Ai miei amici e colleghi di Università, per l'aiuto reciproco e per aver portato un po' di leggerezza in questo percorso.

Grazie a tutti voi.