Università degli Studi di Padova

Dipartimento di Scienze Statistiche

Corso di Laurea Magistrale in

Scienze Statistiche

ANALISI DI UN'ORGANIZZAZIONE CRIMINALE ATTRAVERSO MODELLI DI RETE TEMPORALI

Relatore: Prof. Emanuele Aliverti

Dipartimento di Scienze Statistiche

Laureanda: Chiara Bellio

Matricola: 1239136

Anno Accademico: 2021/2022

"L'inchiostro dei sapienti è più prezioso del sangue dei martiri."

Indice

Elenco delle figure

Elenco delle tabelle

Introduzione

"L'uomo è un animale sociale". Così già nel IV secolo a.C Aristotele definiva l'uomo evidenziando quanto le relazioni con gli altri individui fossero parte della sua essenza. Per rispondere a questa esigenza l'uomo ha quindi sviluppato la capacità di comunicare con parole e con gesti. Questa abilità non è però una prerogativa esclusiva dell'essere umano ma le sue finalità lo sono certamente. Nel regno animale la comunicazione e la relazione con l'altro sono finalizzate alla sopravvivenza limitandosi a esprimere dolore o piacere. La parola, invece, è in grado di mostrare l'utile e il dannoso, il giusto e l'ingiusto. Permette agli individui di esprimere emozioni, di scherzare, di comunicare le proprie conoscenze e di chiedere per scoprire: questo, al contrario di tutti gli altri animali, è proprio degli uomini.

Emerge, poi, come la capacità di relazionarsi non è solo un mezzo ma una vera e propria necessità umana. Il COVID e la quarantena hanno mostrato quanto ancora le relazioni interpersonali siano centrali nella vita di tutti, di quanto se ne sentisse la mancanza e il bisogno anche se attorniati da tutte le comodità che la vita moderna assicura.

Proprio per l'importanza che le relazioni sociali ricoprono nella vita di tutti, è fondamentale studiare dati che non si concentrano solo sul singolo individuo ma su come esso si relazioni con gli altri. A porsi questo obiettivo è la *Social Network Analysis*.

In questo elaborato si vuole quindi studiare il comportamento di una struttura sociale, un'organizzazione criminale, e il suo sviluppo nel tempo. Si sono utilizzati i dati relazionali che fotografano le connessioni tra i membri dell'organizzazione in 11 diversi istanti temporali e, tramite l'a-

nalisi dei gruppi e l'applicazione di un modello per dati di reti a tempo discreto, si vogliono studiare le caratteristiche e l'evoluzione della struttura relazionale.

Nel primo capitolo si presenta la *Social Network Analysis* e le sue caratteristiche; partendo dalla definizione del concetto di rete, si presenteranno i principali indici descrittivi per reti a livello singolo e multi-livello per poi passare alla presentazione di un algoritmo per la determinazione delle comunità per reti multi-livello che è una generalizzazione del metodo di Louvain.

Nel secondo capitolo si descrive il modello TERGM (*Temporal Exponential Random Graph Models*): iniziando dalla presentazione del modello ERGM, di cui TERGM è un'estensione per dati di rete a tempo discreto, si vuole introdurre il lettore ai modelli per dati di rete passando poi alla specificazione del modello TERGM, alle sue caratteristiche e metodi di stima. Si prosegue con la descrizione dei metodi della *Markov chain Monte Carlo maximum likelihood estimation* (MCMC-MLE) edi stima della massima pseudo-verosimiglianza (MPLE) evidenziandone limiti e problematicità. L'esposizione del metodo di stima della massima pseudo-verosimiglianza con intervalli di confidenza di tipo *bootstrap* concluderà il capitolo e sarà poi utilizzato nelle analisi dei dati.

Infine, nel terzo capitolo, le procedure definite nei capitoli precedenti saranno applicate ai dati relativi a un'organizzazione criminale attiva in Canada negli anni '90 e su cui la polizia di Montreal ha svolto un'attività investigativa unica nel suo genere; questa operazione ha permesso di valutare l'evoluzione della struttura di rete attraverso l'analisi preliminare mediante gli indici descrittivi, l'analisi dei gruppi e l'applicazione del modello TERGM.

Capitolo 1

Social Network Analysis

"Viviamo in un mondo fatto di connessioni". Tale espressione cattura bene il perchè le reti e l'analisi di rete abbiano riscosso un così forte interesse negli ultimi decenni. A partire dai *social network*, come *Facebook*, fino allo stesso Internet, si è sommersi da esempi di come svariati tipi di connessioni ci colleghino l'uno all'altro. Esistono, inoltre, collegamenti anche a livelli differenti, legati alle istituzioni (es. governi), processi (es. economici) e infrastrutture (es. rete aerea globale).
Negli anni '60, lo psicologo statunitense S. Milgram teorizzarò il concetto di *Small Word* secondo cui i nodi di una rete complessa, nella maggior parte dei casi, sono collegati direttamente ad un esiguo numero di altri nodi ma, nella quasi totalità dei casi, un nodo qualsiasi può essere raggiunto partendo da un qualsiasi altro attraverso un piccolo numero di passaggi (Travers and Milgram, 1977). A sostegno di ciò, un grande passo avanti avvenne nel 1967 quando Milgram stesso decise di condurre l'esperimento che confermò la teoria. Grazie a questo esperimento è nato il concetto dei "Sei gradi di separazione"; tali risultati, poi riconfermati da un altro esperimento del 2001 svolto attraverso l'utilizzo della posta elettronica, dimostravano che mediamente con 6 passaggi è possibile consegnare un messaggio al destinatario sconosciuto.

Certamente, gli umani non solo gli unici ad instaurare strutture di relazioni complesse. Nel mondo naturale si trovano un'infinità di possibili sistemi di rete come eco-sistemi, reti alimentari biologiche, sistema

metabolico degli esseri viventi sino alle comunicazioni neuronali (Kolaczyk, 2014). Per questo l'analisi di rete copre svariate discipline e tra i principali macro-campi troviamo:

- Reti tecnologiche (*Technological Networks*)
- Reti biologiche (*Biological Networks*)
- Reti dell'informazione (*Information Networks*)
- Reti sociali (*Social Networks*)

In questo elaborato ci si concentrerà sullo studio delle reti sociali.
L'Analisi delle reti sociali (*Social Network Analysis* - SNA), sviluppatasi negli anni '30, è una tecnica che consente di misurare e visualizzare le relazioni sociali tra soggetti, gruppi, organizzazioni o altre entità coinvolte in processi di scambio di informazioni.
Una rete sociale è costituita da un qualsiasi gruppo di individui (nodi) connessi tra loro da diversi legami (archi) che vanno dalla conoscenza casuale ai vincoli parentali.
L'obiettivo della SNA è quello di studiare intere strutture sociali (reti complete) o strutture locali (reti ego-centrate) individuando e analizzando i legami tra le unità o i gruppi che rappresentano i nodi della rete.
L'applicazione delle reti complesse nell'ambito delle Scienze Sociali, si sviluppa in particolare con la metodologia della *network analisys* e del *clustering* dei dati.

1.1 Elementi costitutivi di una rete

Come detto in precedenza e come si può vedere in Figura 1.1, con il termine rete, in ambito sociale, si intende un insieme di nodi (individui, gruppi, istituzioni e luoghi) collegati tra di loro tramite archi che rappresentano una qualche relazione che si instaura tra i nodi. Il concetto di rete sociale rimane generico fino quando non viene specificato il motivo, il luogo e il contesto di formazione della stessa.

Figura 1.1: Esempio di rete sociale.

Le tipologie di reti possono essere definite attraverso la direzionalità degli archi e il valore numerico a essi associato. Come si vede in Figura 1.2, si riportano delle esemplificazioni delle due tipologie di reti che guardano alla direzionalità degli archi (indirette e dirette). Nelle reti indirette gli archi non sono caratterizzati da una direzione, ma entrambi i nodi che vengono connessi condividono lo stesso tipo di relazione; un esempio comune di questo tipo di reti indirette è la rete delle amicizie su Facebook, una volta stretta la relazione entrambe i soggetti fanno parte della stessa relazione.
Nelle reti dirette, invece, gli archi sono direzionati e i nodi coinvolti non condividono necessariamente lo stesso tipo di relazione. A livello grafico, l'arco sarà rappresentato come una freccia che congiunge un nodo di partenza e un nodo di arrivo. Facendo ancora riferimento alla Figura 1.2 e guardando il grafo a destra, si può immaginare che rappresenti la rete dei seguiti su Instagram. Tale relazione non è reciproca e, infatti, si osserva che solo i nodi A e C si seguono vicendevolmente mentre tutti gli altri collegamenti non sono reciproci, come nel caso di B che segue E ma non accade il contrario.

Per quanto riguarda invece la seconda distinzione di tipologie, l'interesse si concentra attorno al valore associato all'arco, come si vede in Figura 1.3. Si definiscono reti binarie quelle reti in cui il valore associato a

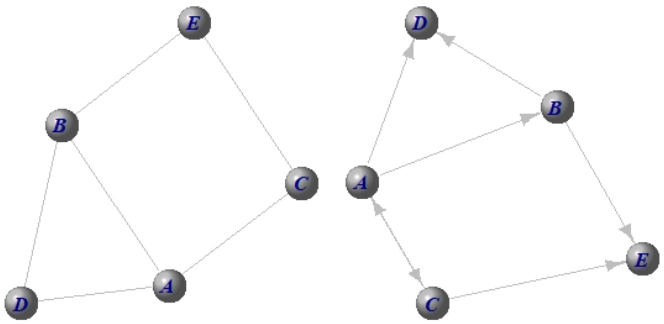

Figura 1.2: Tipi di reti secondo la direzionalità degli archi: rete indiretta e rete diretta.

ogni arco ne caratterizza solo la presenza o assenza (assumendo rispettivamente valore 1 e 0). L'altro caso è quello delle reti pesate nelle quali ogni arco è caratterizzato da un valore numerico che può indicare l'intensità, la forza o la durate della relazione che tale connessione indica. Si possono avere, quindi, quattro tipologie di reti combinando queste caratteristiche di direzionalità e valore dell'arco.
In Figura 1.3 il grafo a destra può rappresentare la rete delle amicizie tra un ristretto numero di persone e il peso associato ad ogni arco può rappresentare il numero di anni di rapporto tra i due individui o altri valori indice della relazione.

1.2 Rappresentazione matematica della rete

Non vi è un modo univoco per rappresentare una rete, infatti è possibile utilizzare grafi o matrici di adiacenza.
Formalmente, si dice grafo una coppia ordinata di insiemi $G = (N, A)$,

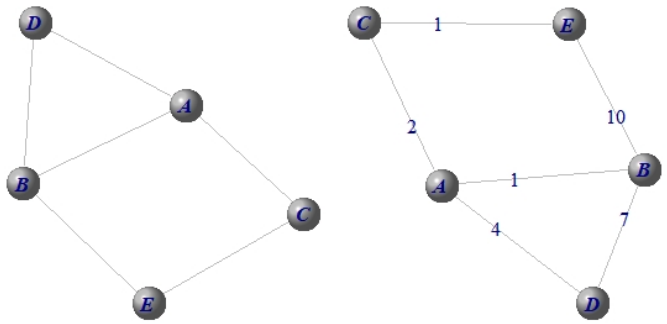

Figura 1.3: Tipi di reti secondo il valore associato agli archi: rete binaria e rete pesata.

dove $N = \{1, 2, ..., V\}$ è l'insieme dei nodi e l'insieme degli archi è $A \subseteq \{\{i, j\} : i, j \in N\}$, tale che un arco è definito come una coppia $\{i, j\} : i, j \in N$. Nel caso di rete pesata, l'insieme degli archi deve riportare anche il valore del peso e pertanto $A \subseteq \{\{i, j, w_{ij}\} : i, j \in N, w_{ij} \in \mathbb{R}\}$. Per rappresentare un rete si può ricorrere anche alla matrice di adiacenza. Tale matrice quadrata, Y, con dimensione $V \times V$ è simmetrica nel caso di rete indiretta e, se è anche binaria, si definiscono i suoi elementi come $Y_{ij} = Y_{ji} = 1$ se $\{i, j\} \in A$ (i e j sono connessi), altrimenti 0. Nel caso di rete diretta non avremo più la necessaria uguaglianza tra $Y_{ij} = Y_{ji}$ e se invece fosse pesata gli elementi sarebbero identificati da $Y_{ij} = w_{ij}$ se $\{i, j, w_{ij}\} \in A$.

1.3 Tipologie di reti

Esistono differenti tipologie di *social network* che possono essere studiate. Come largamente presentato da Wasserman e Faust (1994), si possono categorizzare le reti a seconda della natura degli individui e per le proprietà

delle connessioni che li legano.

- *One-mode Networks*:

  Sono il più diffuso e semplice tipo di rete che studiano un definito set di individui. Tali individui (fisici o virtuali), o gruppi di essi, rappresentano i nodi della rete e gli archi vengono identificati da uno specifico legame sostanziale, un ben definito tipo di relazione, che viene rilevato a livello di coppie di nodi.

- *Two-mode Networks*:

  Si tratta di un particolare tipo di rete per il quale si possono avere due set di attori o un set di attori che viene osservato nel corso di due eventi. Nel primo caso, anche noto come *dyadic two-mode network*, si hanno due insiemi di nodi e i collegamenti vengono stabiliti solo tra nodi appartenenti a insiemi diversi. Il secondo caso, invece, definito *affiliation network*, si ha quando si guarda ad un singolo set di nodi per quando riguarda la partecipazione a diversi eventi o attività. Avremo, quindi, un set di individui e un set di eventi o attività ai quali i soggetti possono aver aderito.

- *Ego-Networks*:

  Si definiscono le *ego-networks* come reti costituite da un singolo attore (*ego*) scelto arbitrariamente, dall'insieme degli attori a cui è collegato (*alter*) e tutti i collegamenti tra questi alter. Ciò che fa emergere l'attrattività verso questa tipologia di rete è la facilità di raccolta dei dati rispetto alle altre tipologie di reti. Le informazioni sugli alter, incluso il modo in cui sono collegati, sono generalmente ottenute interamente dall'ego.

- Reti multi-livello:

  Tali reti, anche dette *multilayer*, sono costituite da un insieme di soggetti connessi tra loro su diversi livelli, tipicamente sovrapposti l'uno all'altro, in cui il dato relazionale si manifesta (esempio in Figura 1.4).

  Solitamente, ma non necessariamente, i nodi sono i medesimi, e se

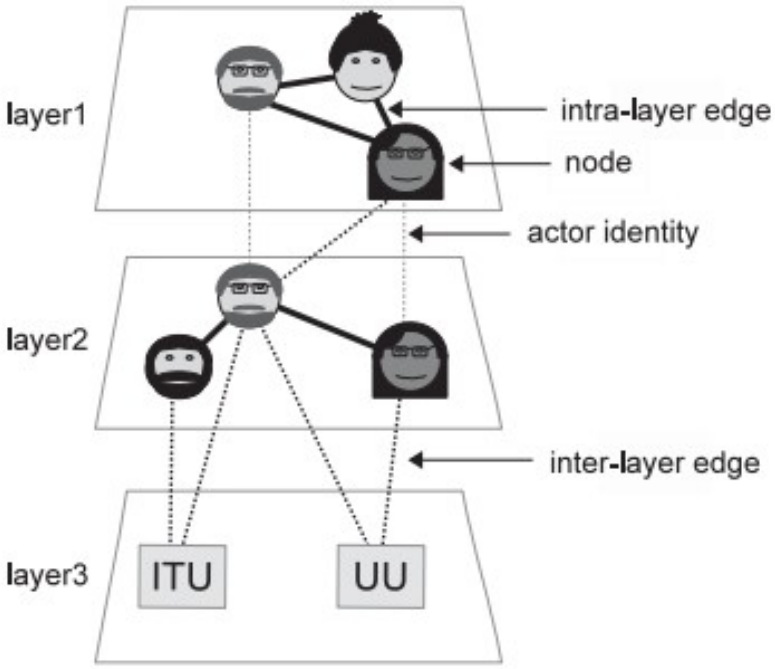

Figura 1.4: Rappresentazione di una rete multi-livello.

manca un nodo in un livello è perchè non ha legami. I vari livelli possono anche essere considerati come diverse tipologie di *edge* (come se ogni livello fosse un attributo del legame).
Una peculiarità importante delle reti multi-livello è quella di presentare due tipologie di archi: quelli tra nodi dello stesso livello, detti *intralayer edges*, e quelli tra nodi di livelli diversi, detti *interlayer edges*. Gli *intralayer edges* dei diversi livelli possono anche avere natura diversa, per esempio, un livello può contenere archi che rappresentano relazioni di lavoro, un livello relazioni di amicizia e un livello può indicare le connessioni tra account *social* o, in alternativa, ogni livello può rappresentare lo stesso tipo di relazione in diversi momenti temporali.

Definizione 1.1 (Rete multi-livello)**.** *Una rete multi-livello può essere definita come* $G = (N, A, L)$ *dove* L *è l'insieme dei livelli,* N *l'insieme dei nodi e* A *è l'insieme degli archi, del tipo* $\{u, v, l_u, l_v\}$ *con* $u, v \in N$ *e* $l_u, l_v \in L$*.*

1.4 Indici descrittivi di rete

Un primo modo per poter caratterizzare la struttura di rete è quello di associarle degli indici descrittivi sintetici che permettano di avere risposte a molteplici domande: come sono connessi i nodi nella rete? Ci sono nodi che hanno un'importanza maggiore? Quanto sono vicini tra loro i nodi? Si possono riconoscere delle comunità all'intero della rete?...
Alcuni indici descrittivi sono riferiti al singolo nodo della rete mentre altri si riferiscono all'intera rete e altri ancora a gruppi di nodi, ma è prima utile introdurre delle quantità che caratterizzano la rete che si basano sul concetto di distanza.

1.4.1 Matrice delle distanza geodetiche

La distanza geodetica è in matematica, o più precisamente in geometria differenziale, la curva più breve che congiunge due punti di uno spazio. Nella teoria dei grafi, tale distanza è, per ogni coppia di nodi i e j, il cammino [1] più corto tra nodi interconnessi che uniscono i e j, ovvero *shortest paths*.
Una matrice utile per la descrizione e lo studio delle reti è la matrice delle distanze geodetiche, S, di dimensione $V \times V$ che permette di misurare la distanza tra due nodi e la lunghezza dello *shortest path*. Si noti che potrebbe non esserci un unico *shortest paths* tra due nodi e che potrebbe non esistere alcun cammino che collega i due nodi; in questo ultimo caso, convenzionalmente, la distanza è definita infinita.
Nel caso di reti pesate, la matrice delle distanze geodetiche, S, è calcolata differentemente. Infatti, se abbiamo ad esempio la rete indiretta pesata $G = (N, A)$, con $N = \{A, B, C\}$ e $A = \{\{A, B, 2\}; \{A, C, 4\}; \{B, C, 1\}\}$, la connessione tra A e C è due volte più forte della connessone tra A e B. Ciò potrebbe significare che il nodo A ha contatti più frequenti con il nodo C che con il nodo B. Dando un'occhiata al percorso più breve tra

[1] Un cammino, nella teoria dei grafi, è definito come una sequenza finita di archi adiacenti che permettono di collegare due nodi del grafo. Affinché due archi siano adiacenti devono condividere uno dei due nodi; se abbiamo due archi $\{i, z\}$ e $\{j, z\}$ appartenenti ad una rete indiretta questi sono adiacenti in quanto condividono il nodo z.

il nodo C e il nodo B, la connessione diretta ha un peso di 1; tuttavia, la connessione indiretta attraverso il nodo A è composta da legami più forti. Pertanto, un'informazione potrebbe essere trasmessa più rapidamente attraverso il nodo A che direttamente.

Dijkstra (2022) ha proposto un algoritmo che somma il costo delle connessioni e trova il percorso di minor resistenza. Un esempio applicativo di tale algoritmo possono essere i dispositivi GPS. Questi utilizzano tale algoritmo assegnando un costo temporale a ciascun tratto di strada; quindi, trovano il percorso che costa meno in termini di tempo. Questo algoritmo può essere utilizzato anche nell'analisi delle *social network*. L'algoritmo di Dijkstra prevede che per ogni arco venga considerato il reciproco del proprio peso, $\frac{1}{w_{ij}}$, e gli *shortest path* vengono calcolati come la somma di tali valori per ogni arco che appartiene al percorso che connette i a j. Si vede quindi che la connessione diretta tra il nodo C e il nodo B ha un costo di 1, mentre la connessione indiretta tramite il nodo A ha un costo di 0,75 ($\frac{1}{2} + \frac{1}{4}$). Pertanto, secondo questo algoritmo, le informazioni viaggeranno più velocemente attraverso la connessione indiretta.

Si consideri la rete, illustrata in Figura 1.5, diretta e pesata come esempio esplicativo delle quantità sopra descritte.

Il grafo $G = (N, A)$ è composto dall'insieme dei nodi $N = \{A, B, C, D\}$ e l'insieme degli archi $A = \{\{A, B, 2\}; \{B, A, 1\}; \{B, D, 1\}; \{C, B, 5\}\}$. La matrice di adiacenza Y è:

$$Y = \begin{bmatrix} . & 2 & 0 & 0 \\ 1 & . & 0 & 1 \\ 0 & 5 & . & 0 \\ 0 & 0 & 0 & . \end{bmatrix}$$

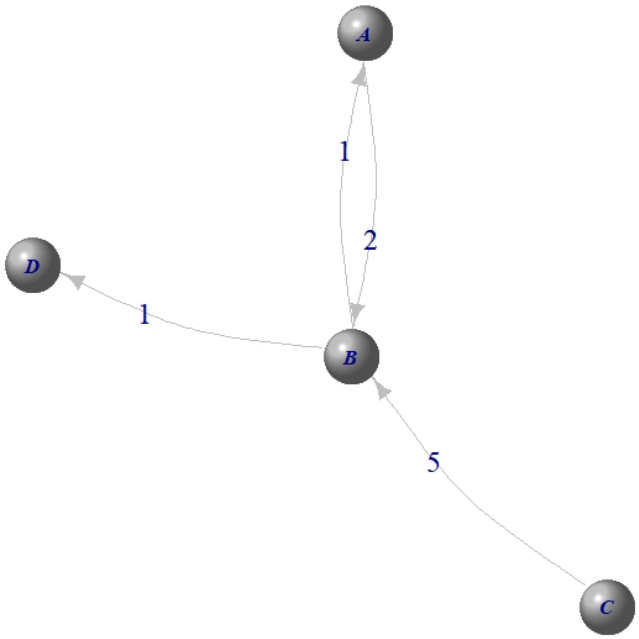

Figura 1.5: Rete di esempio diretta-pesata.

La matrice delle distanze geodetiche S è:

$$S = \begin{bmatrix} 0 & 0.5 & \infty & 1.5 \\ 1 & 0 & \infty & 1 \\ 1.2 & 1 & 0 & 1.2 \\ \infty & \infty & \infty & \infty \end{bmatrix}$$

1.4.2 Indici descrittivi a livello di nodo

- Grado del nodo i

  Tale indice indica il numero di nodi con cui i è connesso:

  $$n_i = \sum_{j=1}^{V} Y_{ij}. \tag{1.1}$$

  La sua versione normalizzata si ha dividendo per il numero delle possibili connessioni, ovvero $\frac{n_i}{V-1}$. Nel caso di rete diretta quanto detto in precedenze vale ma si possono considerare anche gli *out-degree* e *in-degree*. In questi ultimi due casi, il calcolo del grado si

concentra, rispettivamente, solo sugli archi che partono da i e solo gli archi che entrano in i.

Tipicamente, il grado del nodo è elevato solo per pochi nodi della rete che sono pertanto coinvolti in molte relazioni occupando una posizione di rilievo. Questi pochi nodi con molte connessioni si possono definire *hub*.

- *Closeness centrality* del nodo i

  L'indice evidenzia i nodi che sono in grado di diffondere informazioni in modo efficiente attraverso un grafico. Si calcola come il reciproco della somma delle distanze geodesiche:

$$c_i = \frac{1}{\sum_{j=1}^{V} s_{ij}}. \tag{1.2}$$

  Tale misura è normalizzata se moltiplicata per il numero di possibili connessioni totali, $c_i(V-1)$. I nodi con un punteggio di *closeness* elevato hanno le distanze più brevi da tutti gli altri nodi. Tale indice evidenzia come una caratteristica importante per un nodo non sia solo quante connessioni ha con gli altri nodi ma quanto facile sia, per il nodo i, scambiare informazioni con gli altri nodi.

- *Betweenness* del nodo i

  Si può definire la *betweenness* come la somma del *shortest paths* che passano per uno stesso nodo. Questo indice permette di tenere conto dell'interazione tra soggetti non adiacenti e fa emergere se il nodo i ha la funzione di ponte dell'informazione. Un valore elevato di *betweenness*, infatti, comporta un ruolo decisivo nel controllo delle comunicazioni all'interno della struttura della rete.

  Si ottiene come la somma, fatta su tutte le coppie di nodi u e v diversi da i, del rapporto tra il numero degli *shortest paths* tra u e v che passano per i ($n_{uv}(i)$) e il totale degli *shortest paths* tra u e v,

(n_{uv}):

$$g_i = \sum_{u \neq v \neq i}^{V} \frac{n_{uv}(i)}{n_{uv}}. \tag{1.3}$$

La forma normalizzata della *betweenness* si ottiene dividendo il suo valore per il numero di coppie che non includono il nodo i: $\frac{g_i}{[(V-1)(V-2)/2]}$

1.4.3 Indici descrittivi a livello di rete

- Densità di Y

  La densità rappresenta la proporzione di possibili relazioni nella rete che sono effettivamente presenti, sul totale delle relazioni potenziali:

$$D = \frac{1}{V(V-1)} \sum Y_{ij}. \tag{1.4}$$

  Il valore va da 0 a 1, con il limite inferiore che corrispondente alle reti senza relazioni e il limite superiore che rappresenta le reti con i nodi connessi a tutti gli altri nodi presenti nella rete. Più il valore è vicino a 1, più densa è la rete e più coesi sono i nodi della rete. Le informazioni nelle reti dense possono fluire più facilmente delle informazioni nelle reti sparse.

- Distribuzione esponenziale (*power low*)

  È la distribuzione empirica di $[n_1, ..., n_V]$. Tale indice spiega poco delle caratteristiche della rete nel suo complesso ma fornisce indizi importanti sulla struttura della rete. Ad esempio, nelle reti più piccole, si troverebbe che la maggior parte dei nodi della rete ha gradi simili. Tuttavia, le reti del mondo reale di solito hanno distribuzioni del grado molto diverse. In una qualsiasi rete reale, è molto probabile riscontrare che la maggior parte dei nodi ha un grado relativamente piccolo e che solo alcuni nodi hanno grado elevato, essendo collegati a molti altri nodi. Un nuovo nodo, infatti, tenderà a connettersi con nodi che hanno più collegamenti nella rete: il ricco diventa sempre più ricco mentre il povero sempre più povero (in

proporzione), (Barabási, 2003). La relazione tra numero di nodi e numero di connessioni è esponenziale negativa, e quindi invariante per cambiamenti di scala: $nodi \approx e^{-\gamma connessioni}$. Una delle caratteristiche salienti di una rete scale-free è infatti che la sua topologia non differisce in maniera significativa se la rete viene presa in considerazione nella sua totalità o in un suo sottoinsieme.

Questa invarianza di scala permette che sia possibile confrontare il numero di nodi con diverso numero di connessioni; si vede che la proporzione è $e^{-\gamma(N_i - N_j)}$, dove N_i e N_j sono le numerosità dei nodi con due determinati gradi mentre γ è un parametro del tipo di rete considerato. Ciò si ricollega alla tipica forma che prende l'istogramma della distribuzione del grado, detta *power laws*, di cui γ è il parametro. Proprio questa caratteristica della rete fa emergere i già definiti *hub*, una delle principali caratteristiche che distingue una *scale-free network* da un *random network*.

- Diametro di Y

  Può essere definito come il più lungo di tutti gli *shortest paths*, $max\{s_{ij}\}$. È la distanza più breve tra la coppia di nodi più distante della rete. In altre parole, una volta calcolata la lunghezza del percorso più breve tra tutte le coppie di nodi, il diametro è la più lunga di tutte le lunghezze minime del percorso calcolate.

- Lunghezza media *shortest paths*

  È definito come il numero medio degli archi attraversati per congiungere due nodi:

$$L = \frac{1}{V(V-1)} \sum s_{ij}. \tag{1.5}$$

  È una misura dell'efficienza del trasporto delle informazioni sulla rete. La lunghezza media del percorso distingue una rete facilmente trattabile da una che è complicata e inefficiente, essendo desiderabile una lunghezza media del percorso più breve.

- Transitività

  La transitività indica la probabilità che vertici adiacenti (allo stesso nodo) siano a loro volta connessi (*clustering coefficient*); rappresenta la capacità di formare strutture triangolari all'interno della rete.

  La transitività all'interno di una rete sociale è basata sul concetto di triade: sono infatti coinvolti 3 attori per poterne parlare. Esiste transitività nella triade quando, dati tre attori indicizzati da i, j e h, si verifica che i è connesso a j e h (triadi potenziale) e anche j e h sono connessi tra loro (triadi reali); se solo i nodi j e h non fossero connessi, saremo in una condizione di intransitività.

  Tale indice è pari alla probabilità che vertici adiacenti (allo stesso nodo) siano a loro volta connessi (*clustering coefficient*),

$$\text{Transitività} = \frac{\text{num. di triadi transitive reali}}{\text{num. triadi transitive potenziali}}. \tag{1.6}$$

  L'indice varia da 0 a 1, dove 1 sta ad indicare che il grafo è completamente transitivo. Nelle reti sociali solitamente l'indice di transitività varia tra 0.3 e 0.6 (Orman et al., 2013).

- Reciprocità

  La reciprocità è la proporzione dei legami corrispondenti esistenti tra i nodi della rete sociale,

$$\text{Reciprocità} = \frac{\text{num. di archi reciproci}}{\text{num. di archi totali}}; \tag{1.7}$$

  ricopre un ruolo fondamentale nelle reti in cui i legami non sono reciproci per natura dei dati (reti dirette). In determinate condizioni è utile quindi verificare il grado di reciprocità.

1.4.4 Indici descrittivi a livello di gruppo

Con questi indici si vuole valutare la coesione dei gruppi. Nelle reti i soggetti tendono a formare gruppi; tali gruppi sono caratterizzati da un numero di connessioni interne elevato e poche connessioni tra gruppi. Il concetto fondamentale attorno a cui ruotano i gruppi è quello dell'omo-

	Gruppo 1	Gruppo 2	Marginale
Gruppo 1	e_{11}	e_{21}	a_1
Gruppo 2	e_{12}	e_{22}	a_2
Marginale	a_1	a_2	

Tabella 1.1: Esempio di divisione in gruppi per calcolo modularità e assortatività.

filia. Per omofilia si intende la tendenza degli individui a formare legami con persone demograficamente (età, sesso, razza) o psicologicamente (intelligenza, attitudini, aspirazioni, educazione, passioni) simili.

- Modularità

  Tale indice vuole indagare la qualità della divisione in gruppi della rete; ad alti valori della modularità si collega una buona suddivisione in gruppi. All'interno delle comunità la densità sarà elevata ma fra le comunità ci saranno pochi collegamenti e quindi una densità inferiore.

  Tale indice si può definire come la frazione di archi che connettono nodi dello stesso gruppo meno il valore atteso di connessioni che ci si aspetterebbe se gli archi fossero distribuiti casualmente, si veda Tabella 1.1.

$$Q = \sum_{k}^{K} e_{kk} - \sum_{k}^{K} a_k^2 \quad , \quad \text{dove K è il numero di gruppi} \tag{1.8}$$

- Assortatività

  È la versione normalizzata della modularità e permette non solo di valutare la bontà della divisione in comunità della rete, ma anche di confrontare quanto le comunità sono coese tra diverse reti.

$$R = \frac{\sum_k^K e_{kk} - \sum_k^K a_k^2}{1 - \sum_k^K a_k^2} \tag{1.9}$$

  Tale indice varia tra 0 e 1 ma a livello empirico, per definire che i gruppi hanno un buon livello di assortatività non è necessario avere valori prossimo a 1 ma sono sufficienti valori di R superiori a 0.3; tali valori suggeriscono che le comunità trovate sono non banali.

1.4.5 Indici descrittivi per reti multi-livello

In questo paragrafo si presentano una serie di indici descrittivi per le reti multi-livello facendo riferimento alla trattazione fatta da Dickison (2016). Alcune di queste misure hanno una controparte nell'analisi di rete a livello singolo, mentre altre si concentrano sui diversi livelli o sulle loro interazione e non hanno equivalenti specifici tra le misure tradizionali.

Si indagheranno gli indici degli attori (nodi) utilizzati per descriverne le caratteristiche rispetto alle loro connessioni sui diversi strati. Alcune sono versioni estese di misure SNA esistenti, per esempio, il grado, la *betweenness* e il coefficiente di *clustering*, mentre altre sono specifiche per le reti multi-livello e possono essere utilizzate per quantificare la rilevanza di uno o più strati per un attore. Le misure dei livelli, inoltre, si concentrano sulle relazioni tra gli strati, ad esempio la loro somiglianza. È importante sottolineare come e perchè le reti multi-livello presentino un insieme unico di caratteristiche e di problemi che richiedono approcci specifici per essere affrontate.

Quando si studiano le reti multi-livello, la complessità aggiuntiva introdotta dalle relazioni esistenti tra gli strati può essere gestita in diversi modi, a seconda dell'interpretazione dei dati della rete e degli obiettivi dell'analisi. A livello generale si possono identificare quattro approcci diversi.

Il primo approccio consiste nell'unire gli strati per ottenere una rete *one-mode*. Questo processo, spesso chiamato *flattening* (appiattimento) può essere eseguito in diversi modi. Un modo semplice è quello di creare un nuovo grafo con un nodo per ogni attore e un arco tra due nodi se gli attori corrispondenti sono connessi in uno qualsiasi degli strati. Per conservare più informazioni, possiamo anche aggiungere un peso a ogni arco della *flattened network*, che rappresenta il numero di strati in cui gli attori sono connessi. Ottenuta la rete *one-mode* è possibile calcolare le misure tradizionali di SNA.

Il secondo approccio consiste nell'applicare le misure esistenti a ciascun

livello separatamente e poi confrontarne i risultati. Questo approccio, complementare al primo, cerca di preservare le informazioni che potrebbero essere contenute in alcuni strati ma che vengono nascoste dal processo di appiattimento.

Diversamente da questi primi due approcci, il terzo e il quarto rappresentano le reti multistrato con modelli ad hoc. Il terzo approccio considera più strati allo stesso tempo, ma senza trattarli come ontologicamente diversi. Le misure basate su questo approccio considerano esplicitamente la differenza tra archi *intra-layer* e *inter-layer* e fanno anche delle distinzioni numeriche tra diversi strati, ad esempio attraverso i pesi, ma alla fine producono tipicamente singoli valori numerici che uniscono i contributi dei diversi tipi di archi. Diverse misure tradizionali possono essere estese alle reti multi-livello utilizzando questo approccio. Ad esempio, si possono eseguire più *random walk* estesi e contare quante volte un nodo viene attraversato quando si passa da un attore all'altro, richiamando l'idea di centralità. Questa classe di approcci considera la molteplicità dei livelli come una caratteristica inevitabile delle reti, che richiede una nuova serie di strumenti per essere gestita correttamente. In questo caso, l'attenzione non si concentra solo sull'ottenimento di informazioni aggiuntive dall'adozione di una prospettiva di rete multi-livello, ma anche sull'evitare possibili distorsioni dovute a un processo di appiattimento e a una sottostima delle correlazioni tra gli strati. Tuttavia, i *random walk* introducono un'ipotesi di fondo secondo cui archi di diversi strati siano di natura comparabile. È ovviamente possibile determinare parametri che affermano come è più probabile rimanere all'interno di determinati livelli piuttosto che altri, ma quando si devono calcolare delle misure, i percorsi che attraversano diversi livelli sono trattati in modo indistinto. Questo potrebbe limitare la nostra capacità di gestire la molteplicità di livelli, in cui gli attori sono collegati da legami qualitativamente diversi. Queste considerazioni portano al quarto approccio, in cui le relazioni che coinvolgono livelli diversi sono analizzate insieme ma non vengono mescolate tra loro. Pertanto, in questa prospettiva, è necessario introdurre misure

che descrivano queste informazioni aggiuntive senza sintetizzarle in un unico valore ma mantenendo, per quanto possibile, la distinzione tra i diversi tipi di legami relazionali, come si vedrà per il concetto di distanza multi-livello. Diversi autori, come Magnani e Rossi (2013), e Solé-Ribalta et a (2014), hanno cercato di adottare questo approccio complementare, definendo misure basate su un modello di rete multistrato ma, allo stesso tempo, mantenendo i diversi strati sostanzialmente distinti. Tutti i seguenti indici fanno riferimento a quanto esposto nel lavoro di Berlingerio et al. (2011); si deve sottolineare che parlando di dimensioni della rete multi-livello si farà riferimento agli strati.

Indici a livello di nodo

- Grado del nodo v

  Per far fronte all'impostazione multidimensionale, si può definire il grado di un nodo rispetto a una singola dimensione (*layer*) o a un insieme di dimensioni. A tal fine, dobbiamo ridefinire il dominio della funzione di grado vista in precedenza, includendo anche le dimensioni.

  Definizione 1.2 (Grado)**.** *Sia $v \in N$ un nodo di una rete $G = (N, A, L)$. La funzione $Degree : V \times P(L) \to \mathbb{N}$ definita come*

  $$Degree(v, D) = |\{(u, v, d) \in E \quad s.t. \quad u \in N \wedge d \in D\}| \tag{1.10}$$

  calcola il numero di archi, appartenenti ad una delle dimensioni in D, tra v e qualsiasi altro nodo u.

  Possiamo considerare due casi particolari: quando $D = L$ abbiamo il grado del nodo v all'interno dell'intera rete, mentre quando l'insieme di dimensioni D contiene solo d, abbiamo il grado di v nella dimensione d, che è il grado di un nodo in una rete monodimensionale. Dato che possiamo calcolare per il nodo v il valore del grado per ogni dimensione è possibile calcolarne la deviazione standard.

Definizione 1.3 (Deviazione standard del grado)**.** *Dato un nodo $v \in N$ e $D \in L$ l'insieme delle dimensioni della rete multi-livello $G = (N, A, L)$. La deviazione standard del grado v sulle dimensioni L è*

$$\sqrt{\frac{\sum_{l \in L}(Degree(v,l) - \frac{Degree(v,L)}{|L|})^2}{|L|}} \tag{1.11}$$

Un attore con lo stesso grado su tutti i livelli avrà una deviazione pari a 0, mentre un attore con grado elevato su un livello e solo pochi su un altro avrà una deviazione di grado elevata.

- Vicini del nodo v

  In una rete a livello singolo indiretta, ogni nodo può avere un solo arco che lo connetta con ogni altro nodo presente nella rete. Nelle reti multidimensionali il grado di un nodo ed il numero di nodi ad esso adiacenti, i vicini, non sono più corrispondenti, poiché può esserci più di un arco tra due nodi qualsiasi. Definiamo una misura relativa ai vicini di un nodo.

  Definizione 1.4 (*Neighbors*)**.** *Sia $v \in N$ e $D \subseteq L$ un nodo e un insieme di dimensioni di una rete $G = (N, A, L)$. La funzione* Neighbors*: $V \times P(L) \to \mathbb{N}$ è definita come*

$$Neighbors(v, D) = |NeighborSet(v, D)| \tag{1.12}$$

dove $NeighborSet(v, D) = \{u \in N | \exists (u, v, d) \in A \wedge d \in D\}$. Questa funzione calcola il numero di tutti i nodi direttamente raggiungibili dal nodo v tramite archi etichettati con dimensioni appartenenti a D.

Si noti che, nel caso monodimensionale, il valore di questa misura corrisponde al grado. È facile vedere che $Neighbors(v, D) \leq Degree(v)$, ma possiamo anche facilmente dire qualcosa sul rapporto $\frac{Neighbors(v,D)}{Degree(v)}$. Quando il numero di vicini è piccolo, ma ognuno di essi è connesso attraverso molti archi a v, si avrà un basso valore

del rapporto, ciò significa che l'insieme delle dimensioni è in qualche modo ridondante rispetto alla connettività di quel nodo. All'estremo opposto, se le due misure coincidono questo rapporto è uguale a 1, il che significa che ogni dimensione è necessaria (e non ridondante) per la connettività di quel nodo: rimuovendo una qualsiasi dimensione, si taglierebbero le connessioni nodo da alcuni dei suoi vicini.

Definiamo anche una variante della funzione *Neighbors*, che tiene in considerazione solo i nodi adiacenti che sono collegati da archi appartenenti solo a un determinato insieme di dimensioni.

Definizione 1.5 ($Neighbors_{XOR}$)**.** *Siano $v \in N$ e $D \subseteq L$ un nodo e un insieme di dimensioni di una rete $G = (V, E, L)$. La funzione $Neighbors_{XOR} : V \times P(L) \to \mathbb{N}$ è definita come: $Neighbors_{XOR}(v, D) = |\{u \in N | \exists d \in D : (u, v, d) \in A \wedge \nexists d' \notin D : (u, v, d') \in A\}|$. Calcola il numero di nodi vicini connessi da archi appartenenti solo a dimensioni in D.*

- *Dimension Relevance*

  Un aspetto fondamentale dell'analisi di rete multidimensionale è capire quanto sia importante una particolare dimensione rispetto alle altre per la connettività di un nodo; vale a dire cosa succede alla connettività del nodo se rimuoviamo quella dimensione.

  Definizione 1.6 (*Dimension Relevance*)**.** *Sia $v \in N$ e $D \subseteq L$ siano un nodo e un insieme di dimensioni di una rete $G = (N, A, L)$. La funzione $DR : N \times P(L) \to [0, 1]$ è definita come*

$$DR(v, D) = \frac{Neighbors(v, D)}{Neighbors(v, L)} \tag{1.13}$$

  e calcola il rapporto tra i vicini di un nodo v connessi da spigoli appartenenti a uno specifico insieme di dimensioni in D e il numero totale dei suoi vicini.

  Chiaramente, l'insieme D potrebbe anche contenere una sola dimensione d, di cui l'analista potrebbe voler studiare il ruolo speci-

fico all'interno della rete. Tuttavia, in un contesto multidimensionale, questa misura può non cogliere informazioni importanti sulla connettività di un nodo.

Definizione 1.7 (Dimension Relevance XOR)**.** *Sia $v \in N$ e $D \subseteq L$ siano rispettivamente un nodo e un insieme di dimensioni di una rete $G = (N, A, L)$. $DR_{XOR} : V \times P(L) \to [0, 1]$ è definito come*

$$DR_{XOR}(v, D) = \frac{Neighbors_{XOR}(v, D)}{Neighbors(v, L)} \tag{1.14}$$

e calcola la frazione di vicini direttamente raggiungibili dal nodo v seguendo archi appartenenti solo alle dimensioni D.

Indici a livello di *layer*

Interessante nell'ambito delle reti multi-livello è discutere le misure di somiglianza tra gli strati. Le relazioni tra i livelli possono essere studiate da due prospettive diverse: da un lato, possiamo descrivere le differenze tra i diversi livelli come segno di comportamenti diversi degli attori, che scelgono strategicamente quali tipi di connessioni vogliono instaurare su ogni strato (definiamo questa prospettiva centrata sull'attore); dall'altro lato, possiamo descrivere queste differenze in termini di influenze *inter-layer* (definiamo questa prospettiva come una prospettiva centrata sullo strato). Il modo più semplice per indagare su questo aspetto è quello di applicare concetti esistenti, come quello di correlazione, per indagare le somiglianze delle reti multi-livello; l'idea di somiglianza si concretizza nella presenza di archi tra gli stessi attori su strati diversi.
Questo è stato l'approccio seguito da Berlingerio et al. Berlingerio et al. (2011) quando hanno sviluppato l'idea di correlazione di strato come una versione, per reti multi-livello, del classico coefficiente di correlazione di Jaccard [2] in grado di gestire più di due strati contemporaneamente.

[2] L'indice di Jaccard è un indice statistico utilizzato per confrontare la similarità e la diversità di insiemi campionari. Il coefficiente di Jaccard misura la similarità tra insiemi campionari, ed è definito come la dimensione dell'intersezione divisa per la dimensione dell'unione degli insiemi campionari:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Questo coefficiente è chiamato *D-Correlation* ed è definito come:

Definizione 1.8 (Pair D-Correlation)**.** *Dato $D \subseteq L$ insieme degli strati della rete $G = (N, A, L)$. La Pair D-Correlation è la funzione:*

$$\rho_{pairs}(D) = \frac{|\bigcap_{d \in D} P_d|}{|\bigcup_{d \in D} P_d|} \tag{1.15}$$

dove P_d è l'insieme delle coppie di nodi (u, v) connessi nello strato d. Si calcola il rapporto tra le coppie di nodi connessi in tutti gli strati di D e il numero totale di coppie connesse in almeno uno degli strati di D.

La correlazione *inter-layer* può fornire molti spunti interessanti sulle dinamiche sociali in corso. Come per ogni correlazione, sarebbe sbagliato ipotizzare una relazione causale tra i due livelli osservati, ma è comunque importante essere in grado di individuare i livelli di somiglianza all'interno delle reti multi-livello.

1.5 *Community Detection*

Dopo la presentazione degli indici descrittivi per le reti singole e multi-livello, che permettono di avere le prime informazioni sulla struttura della rete e sulle sue peculiarità, è utile guardare alla possibile presenza di comunità all'interno della rete. Una comunità si identifica come una insieme di nodi densamente collegati tra loro e scarsamente connesso con i nodi appartenenti agli altri gruppi. L'obiettivo dell'identificazione delle comunità è comune nell'analisi di rete e da la possibilità di delineare una plausibile struttura all'interno della rete analizzata facendo emergere, ad esempio, quali soggetti collaborano maggiormente o se i soggetti tendono a formare gruppi di grandi o piccole dimensioni. Nel caso di reti multi-livello che fotografano lo stesso tipo di relazione, sugli stessi soggetti in diversi istanti temporali, è interessante guardare alle comunità per comprendere come l'organizzazione della rete si modifichi nel corso del tempo. È pertanto necessario presentare un algoritmo di *Community Detection* per le reti multi-livello.

Il metodo che si utilizza è la generalizzazione al caso di reti multi-livello

del metodo di Louvain, basato sull'ottimizzazione della modularità; cioè cerca di trovare un'assegnazione dei nodi alle comunità tale per cui la modularità sia il più alto possibile. Tale algoritmo viene descritto da Mucha, Richardson et al. (2010) per ricercare comunità tra i vari livelli; tale metodo prevede che lo stesso attore in layer diversi possa appartenere a comunità diverse. La funzione per la modularità multi-livello è tanto più alta quanto più i nodi di una stessa comunità hanno connessioni tra loro e quanto più gli stessi nodi su *layer* diversi appartengono alla stessa comunità. Modularità multi-livello è:

$$Q_m = \frac{1}{2\mu} \sum_{i,j,s,r} [(a_{ijs} - \frac{k_{is}k_{js}}{2m_s}\delta(s,r) + \omega\delta(i,j))]\delta(\gamma_{is}, \gamma_{jr}), \tag{1.16}$$

dove i, j sono attori (nodi), s, r con *layers*, a_{ijs} è 1 se i, j sono adiacenti sul *layer* s, k_{is} è il grado del nodo i sul *layer* s, μ è il numero di coppie di nodi adiacenti o corrispondenti allo stesso nodo, m_s è il numero di archi sul *layer* s, γ_{is} è la comunità alla quale il nodo i è assegnato nel *layer* s, δ è il Delta di Kronecker e ω è il peso; quando lo stesso attore appartiene alla stessa comunità su livelli diversi allora Q_m aumenta grazie ad ω. Si presti attenzione alla scelta parametro ω; infatti, valori alti di ω porteranno a comunità che si estenderanno per più livelli, perchè include lo stesso nodo nei diversi livelli nella stessa comunità, così da incrementare il valore della modularità.

Capitolo 2

Modello TERGM

Si passi ora a considerare un modello statistico per le reti multi-livello con dipendenza temporale. Sono stati presentati, nel capitolo precedente, numerosi indici descrittivi che rappresentano le caratteristiche strutturali di una rete. Tali misure, però, descrivono solo una delle numerose configurazioni che la rete può assumere, perciò si approfondirà il modello TERGM che è un estensione del modello ERGM per reti dinamiche multi-livello.
Si presentano, in primis, la classe di modelli ERGM, il cui obiettivo è quello di identificare il processo che influenza la creazione delle relazioni (archi) tra i nodi, valutando la sua formalizzazione per reti a livello singolo. Si guarderà poi al modello TERGM specificandone struttura e metodi di stima, presentandone le caratteristiche e le problematicità.

2.1 Introduzione ai modelli ERGM

I modelli ERGM (*Exponential Random Graph Models*) sono una famiglia di modelli per l'analisi dei dati di rete basati sulla teoria delle famiglie esponenziali. L'idea alla base di questa famiglia di modelli è quella di esprimere la probabilità di osservare una determinata rete in funzione di alcune statistiche sufficienti di rete (ad esempio gli archi reciproci, per reti dirette, o la transitività). Queste statistiche esprimono la dipendenza locale tra le variabili (reciprocità e raggruppamento transitivo, rispetti-

vamente). È un modello lineare generalizzato per grafi proposto da Frank (1991) e da Wasserman e Pattison (1996). Il modello ERGM considera la rete come una singola osservazione multivariata in cui le connessioni tra i nodi possono dipendere dalle covariate o da processi endogeni alla rete.

Osservata una rete $Y = \{Y_{21}, ..., Y_{V1}, ..., Y_{ij}, ..., Y_{VV-1}\} \in \mathcal{Y}$ dove $\mathcal{Y}$ è spazio campionario di dimensione $\{0,1\}^{\frac{V(V-1)}{2}}$ per una rete binaria e indiretta, altrimenti $\mathbb{R}^{\frac{V(V-1)}{2}}$; il modello è definito dalla funzione di probabilità:

$$P(\mathbf{Y} = Y; \theta) = \frac{\exp\left\{\theta^T g(Y)\right\}}{\kappa(\theta)} = \exp\left\{\theta^T g(Y) - \log \kappa(\theta)\right\}, \qquad (2.1)$$

dove $g(Y)$ è un vettore di statistiche di rete p-dimensionale, θ il vettore dei parametri p-dimensionale e $\kappa(\theta)$ è la costante di normalizzazione che assicura che la probabilità totale sommi a 1. Il modello rappresenta la distribuzione di probabilità ogni possibile rete binaria e indiretta con V nodi. Si noti che la costante di normalizzazione $\kappa(\theta) = \sum_{Z \in \mathcal{Y}} exp\left\{\theta^T g(Z)\right\}$ non è facilmente stimabile in quanto considera tutte le possibili permutazioni della rete con V nodi. Il numero di reti con V nodi è pari a $2^{\frac{V(V-1).}{2}}$; perciò, anche se la rete avesse un numero di nodi contenuto, le possibili permutazioni sarebbero estremamente elevate. Considerando, ad esempio, una rete con $V_1 = 10$ nodi si avrebbero35184372088832 possibili configurazioni e con l'aggiunta di un solo nodo si passerebbe da un numero di permutazioni dell'ordine di 10^{13} all'ordine di 10^{16}.

In questo modello, ogni singolo elemento del vettore dei parametri θ corrisponde ad una singola statistica scalare di rete di g. Il parametro regola il modo in cui la relativa caratteristica di rete influisce sulla probabilità della configurazione di rete Y.

Ponendo l'attenzione sull'interpretazione dei parametri, i termini di un ERGM sono leggermente diversi da quelli di un modello statistico tradizionale. In un modello tradizionale, i dati sono costituiti da un insieme di osservazioni, per ciascuna delle quali sono disponibili un certo numero di variabili. Una o più di queste variabili diventa la variabile di risposta,

mentre le altre sono utilizzate come predittori. In una rete, le osservazioni constano ancora di una variabile risposta (lo stato di una coppia di nodi, tipicamente misurato dalla presenza o dall'assenza di uno o più legami nel caso di una rete diretta) e di statistiche sufficienti misurate separatamente su nodi o archi (per esempio, la durata del legame misurato). In un ERGM, però, i predittori sono funzioni dei legami stessi. I parametri di un modello ERGM possono essere interpretati in termini di propensione a produrre grafi con valori più alti ($\theta_i > 0$) o più bassi ($\theta_i < 0$) delle statistiche sufficienti, rispetto a un modello di riferimento. Il modello ERGM può essere utilizzato per decifrare l'effetto delle caratteristiche della rete nella generazione della rete osservata empiricamente. L'analisi su θ indica gli effetti di ciascuna statistica della rete, indipendentemente dagli effetti delle altre statistiche incluse in g. Questa capacità dell'ERGM di separare gli effetti consente simultanee considerazioni delle determinanti generative della struttura di una rete. Ad esempio, il numero di triangoli (gruppi di tre nodi tutti adiacenti tra loro) è spesso usato come misura della transitività in una rete (Burda et al., 2004). Le reti sparse tenderanno ad avere meno triangoli delle reti che sono molto dense. Pertanto, l'aggiunta di statistiche a g che misurino sia il numero di archi che il numero di triangoli consente di fare inferenza sulla possibilità che ci sia un numero insolitamente alto o basso di triangoli in Y, relativamente al numero di archi complessivo della rete Y.

Tra θ e Y nel modello ERGM esiste una relazione precisa. Se la probabilità di Y è $P(Y, \theta_0)$, allora il valore atteso di $g(Y)$ è uguale a $g(Y_0)$. In altre parole, la parametrizzazione dell'ERGM, con le statistiche g ed i valori dei parametri stimati, che massimizzano la verosimiglianza, da come risultato una distribuzione di reti in cui i valori attesi delle statistiche sufficienti della rete sono uguali ai valori delle statistiche calcolati su Y_0 (Park and Newman, 2004). Dunque, se i valori dei parametri sono posti uguali alle stime di massima verosimiglianza, calcolate sulla rete d'interesse, il modello probabilistico risultante genererà reti che presentano caratteristiche che sono, in media, pari alle caratteristiche della rete

d'interesse.

Per questa famiglia di modelli, si assume che gli archi della rete Y abbia distribuzione appartenente alla famiglia esponenziale; se consideriamo una rete binaria, si assumerà,generalmente, che gli archi abbiano distribuzione di Bernulli, $Y_{ij} \sim Bern(\pi_{ij})$.

Numerose sono le assunzioni sulla dipendenza degli archi che sono state presentate. Si può assumere che gli archi siano indipendenti e si verificano casualmente con una probabilità fissata α come nel modello di Erdős et al. (1960). L'ipotesi di dipendenza in questo caso è tanto semplice quanto irrealistica nelle applicazioni: tutti i possibili archi sono indipendenti l'uno dall'altro. Altre ipotesi un po' più raffinate per le reti, possono essere i modelli $p1$ (Holland and Leinhardt, 1981) e l'estensione $p2$ (Lazega and Van Duijn, 1997), ma risultano altrettanto irrealistiche. Frank and Strauss (1986) hanno introdotto un modello con dipendenza markoviana, in cui si presume che un arco tra i e j non sia indipendente da un qualsiasi altro arco che lega i o j ad un altro nodo della rete. Due archi che condividono un nodo sono condizionatamente dipendenti, dati tutti gli altri archi della rete. Se due archi sono condizionatamente dipendenti, se il valore di uno cambia, la probabilità dell'altro arco ne è influenzata, anche se tutti gli altri legami della rete rimangono gli stessi. Si assume quindi che due archi sono indipendenti, se non hanno alcun nodo in comune, condizionatamente al resto della rete.

2.2 Modello TERGM

Il modello TERGM è un estensione del modello ERGM concepita per considerare la dipendenza temporale delle reti con osservazioni longitudinali, a tempo discreto. Tale estensione si ottiene inserendo nella specificazione di un modello ERGMparametri che tengono in considerazione delle precedenti realizzazioni della rete nella determinazione delle sue caratteristiche attuali. In altre parole, alcune delle statistiche sufficienti sono funzioni dirette di una realizzazione precedente della rete.

L'equazione del modello ERGM in Equazione 2.1 per la rete Y al tempo t, da ora definita come Y_t che rappresenta un singolo livello della rete multi-livello che si vuole studiare, può essere modificata per includere la dipendenza temporale a un numero $k \in \{0, 1, ..., t-1\}$ di osservazioni precedenti della rete. Le osservazioni delle reti ritardate vengono inserite in g:

$$P(Y_t|Y_{t-k}, ..., Y_{t-1}, \theta) = \frac{exp\left\{\theta^T g(Y_t, Y_{t-1}, ..., Y_{t-k})\right\}}{\kappa(\theta, Y_{t-1}, ..., Y_{t-k})}. \tag{2.2}$$

La specificazione di k è importante perchè deve catturare tutta la dipendenza temporale di Y_t. Ovvero, le reti antecedenti a Y_{t-k} sono indipendenti da Y_t condizionata a $Y_{t-k}, ..., Y_{t-1}$. Nel seguito con Y faremo riferimento ad una rete multi-livello, $Y = \{Y_1, ..., Y_T\}$ dove T è il numero di livelli della nostra rete.

In *Equazione* 2.2 viene specificato il modello TERGM per il singolo livello t della rete multi-livello, ossia Y_t. Per ottenere la probabilità congiunta delle reti tra i tempi $k+1$ e T si calcola il prodotto delle probabilità delle singole reti condizionate alle precedenti; ciò è sensato, se k è stato scelto correttamente, in quanto sono condizionatamente indipendenti. In questo modo, un TERGM può tenere in considerazione della dipendenza temporale:

$$P(Y_{k+1}, ..., Y_T|Y_1, ..., Y_k, \theta) = \prod_{T=K+1}^{T} P(Y_t|Y_{t-k}, ..., Y_{t-1}, \theta). \tag{2.3}$$

Con la precedente specificazione della distribuzione congiunta è richiesto che siano note le distribuzioni delle prime k reti. Data una serie di reti di lunghezza T, $Y_0^1, ..., Y_t^1$, allora $\theta_0 = argmax_\theta[\prod_{t=k+1}^{T} P(Y_0^t, \theta)]$.[1] sia il vettore dei parametri stimati. Analogamente all'ERGM statico (per reti a livello singolo), se la probabilità di Y_T è $P(Y_t, \theta_0)$, allora il valore atteso di $g(Y^t, ..., Y^{t-K})$ è pari a $\frac{1}{T-k}\sum_{t=k+1}^{T} g(Y_0^t, ..., Y_0^{t-k})$. Pertanto, il modello TERGM con parametri stimati tramite verosimiglianza sulla serie di interesse mostrerà, in media, le caratteristiche medie di quella

[1] Tale specificazione suggerisce di escludere dalla stima del modello le prime k reti di cui dobbiamo comunque conoscere la distribuzione per la stima del modello. È possibile stimarli agevolmente tramite modello ERGM statici e considerare tali distribuzioni indipendenti dal parametro θ

rete multi-livello.

2.3 Metodi di stima

Passando alla trattazione dei metodi di stima per il modello TERGM, la stima diretta della verosimiglianza della funzione di probabilità non è computazionalmente trattabile, nella maggioranza dei casi, a causa della dimensione di Y, che rende difficile la determinazione della costante di normalizzazione. Infatti, come già precedentemente detto, anche per reti con un esiguo numero di nodi il numero di possibili permutazioni che si configurano per la rete sono elevate; non è computazionalmente pratico calcolare direttamente la funzione di probabilità P per le reti che non abbiano un piccolo insieme di vertici. Pertanto, per ricavare la verosimiglianza di una data configurazione di rete, si deve guardare a metodi di approssimazione della verosimiglianza per via simulata. Si mostra che, sebbene promettenti, i metodi basati sulla simulazione (MCMC-MLE) hannodifficoltà di applicazione nei casi reali perchè non si sa quante simulazioni siano necessarie. Una possibile alternativa ai metodi di simulazione è la stima della massima pseudo-verosimiglianza (MPLE), deterministica e con proprietà asintotiche. Questo secondo approccio tende a sottostimare l'incertezza delle stime, quindi, si introduce un ultimo metodo di massima pseudo-verosimiglianza con intervalli di confidenza *bootstrap*, che superi l'approccio standard dell'incertezza con MPLE.

2.3.1 *Markov chain Monte Carlo maximum likelihood estimation*: MCMC-MLE

La *Markov chain Monte Carlo maximum likelihood estimation*[2] (MCMC-MLE) utilizza un campione di reti da una catena di Markov per approssimare la funzione di probabilità, in Equazione 2.1.

[2]I metodi MCMC sono tra i metodi più utilizzati per campionare da distribuzioni di probabilità complesse. Questi metodi traggono vantaggio dall'utilizzo di apposite catene di Markov e dalle loro proprietà di stazionarietà. Due algoritmi della MCMC sono: l'algoritmo Metropolis-Hastings e l'algoritmo Gibbs Sampler.

La procedura di stima si sostanzia come segue: definiti il numero di elementi del campione della MCMC pari a m, il numero di nodi della rete pari a V e l il numero di livelli della rete multi-livello considerata, si inizializza il vettore dei parametri con $\theta^0 = 0$, o, in alternativa, si può definire uno specifico θ^0 in base a ulteriori informazioni a nostra disposizione. Si passa, poi, alla formazione del campione di reti, per via simulata, considerando la stima corrente del vettore dei parametri attraverso una MCMC. Il campione di reti si campionerà da$\tilde{Y} \backsim P(Y, \theta^{[i]})$; per eseguire questo passo, è necessario poter simulare da una rete. Una volta ottenute le nuove reti e definita la funzione di verosimiglianza approssimata, $\widehat{C(\theta)} = ln(\sum_{j=1}^{m} exp[(\theta - \theta^{[i]})^T g(\tilde{Y}_j)])$, si andranno a calcolare le nuove stime dei parametri che la massimizzano, $\theta^{[i-1]} = argmax_\theta[\sum_{t=1}^{T} \theta^T g(Y)_t - \widehat{C(\theta)}]$. Le procedure di campionamento simulato, approssimazione e massimizzazione si ripetono fino a quando non si verificano cambiamenti piccoli (inferiori ad una soglia fissata) delle stime dei parametri; si arriva allora convergenza dell'algoritmo di stima (Geyer and Thompson, 1992). La procedura di stima MCMC-MLE genera stime che convergono asintoticamente alla stima di massima verosimiglianza; cioè se numero di reti simulate, utilizzate per l'approssimazione della funzione probabilità, tende all'infinito (Strauss and Ikeda, 1990). La procedura MCMC-MLE presenta, però, un limite. Non è mai noto nelle applicazioni pratiche, se m, il numero di simulazioni utilizzate in campionamento tramite MCMC, sia sufficientemente grande.

2.3.2 Stima di massima pseudo-verosimiglianza

Un secondo metodo di approssimazione è la stima della massima pseudo-verosimiglianza (MPLE). Secondo Besag (1974), si supponga che Y sia composto da elementi tali per cui Y_{ij} è uguale a 1 se i nodi i e j sono adiacenti e 0 altrimenti. Invece di approssimare direttamente la verosimiglianza, MPLE utilizza il prodotto delle probabilità condizionate per ogni elemento della rete escludendo l'arco $\{i, j\}$ invece che la verosimiglianza congiunta degli elementi. La probabilità condizionata dell'arco

$\{i, j\}$ uguale a 1 è pari a:

$$\pi_{ij}(\theta) = P(Y_{ij} = 1|Y_{-ij}, \theta) = \frac{1}{1 + exp\{-\theta^T \delta_{ij}(g(Y))\}}, \tag{2.4}$$

dove Y_{-ij} indica tutti gli archi della rete diversi dall'arco $\{i, j\}$ e $\delta_{ij}(g(Y))$ indica il vettore di cambiamento in funzione di $g(Y)$ per Y_{ij} che cambia stato da 0 a 1, mantenendofissato il resto della rete Y. Per il prombela di massimizzazione viene utilizzato l'algoritmo di *hill-climbing*:

$$argmax_\theta \sum_{t=1}^{T} \sum_{\langle i,j \rangle} ln[(\pi_{ij}^t(\theta))^{Y_{ij}^t}(1 - \pi_{ij}^t(\theta))^{1-Y_{ij}^t}], \tag{2.5}$$

dove T è il numero di reti nel campione e $\langle i, j \rangle$ denota tutte le coppie di nodi. Pertanto, il calcolo dell'MPLE non implica una simulazione.
È stato dimostrato che l'MPLE converge in probabilità alla stima esatta MLE, è uno stimatore consistente; ciò significa che la stima MPLE converge in distribuzione alla stima MLE all'aumentare delle dimensioni della rete analizzata (Strauss and Ikeda, 1990). Anche questo approccio di stima presenta una problematicità sostanziale: le misure d'incertezza basate sull'MPLE sono sottostimate, a volte anche gravemente. Infatti, la distribuzione campionaria della massima verosimiglianza nei TERGM è una normale multivariata con media uguale al vettore dei parametri di massima log-verosimiglianza e varianza uguale all'inverso della matrice hessiana negativa della log-verosimiglianza. Come accennato in precedenza, l'MPLE fornisce un'approssimazione consistente dell'MLE. Tuttavia, l'hessiano della funzione log-pseudo-verosimiglianza calcolata con MPLE sottostima la varianza dell'MPLE, che si traduce nella sottostima dell'ampiezza degli intervalli di confidenza.
Confrontandolo con il metodo precedente (MCMC-MLE) la procedura di massima pseudo-verosimiglianza presenta dei vantaggi. Guardando al guadagno in termini di efficienza per MCMC-MLE rispetto a MPLE, è richiesto per il metodo MCMC-MLE uno sforzo di simulazione che si moltiplica all'aumentare delle dimensioni della rete (numero di nodi e di livelli della rete). Questo perchè, grazie alla sua consistenza, le pre-

stazioni dell'MPLE migliorano con l'aumentare delle dimensioni della rete. Pertanto, per mantenere un particolare livello di efficienza relativa del MCMC-MLE rispetto all'MPLE, sono necessarie ulteriori simulazioni quando la rete ha un numero di nodi maggiore; cioè, per assicurarsi che le stime di MCMC-MLE mantengano un determinato standard di efficienza al crescere di V, anche il numero di simulazioni m deve aumentare. In questo modo, l'onere computazionale richiesto per la procedura MCMC-MLE diventa insostenibile a fronte dei vantaggi che può assicurare. Dunque, il vantaggio computazionale di MPLE è più significativo con dati voluminosi, in termini di nodi o dimensioni temporali T, poiché le simulazioni richiedono più tempo e memoria quando invece MPLE guadagna in efficienza.

2.3.3 Stima di massima pseudo-verosimiglianza con intervalli di confidenza *bootstrap*

Si presenti la procedura di stima proposta da Desmarais and Cranmer (2012) per la stima di massima pseudo-verosimiglianza con intervalli di confidenza *bootstrap*. Considerando il caso di dati reali di cui non si conosce il processo generatore, c'è sempre un certo grado di incertezza quando si determinano i parametri del modello sulla base dei dati. Si supponga, ad esempio, che si voglia determinare un parametro in un ERGM che corrisponde a una statistica che misura la transitività in una rete indiretta (es. numero di triangoli). Se la rete osservata mostra un valore elevato per la statistica relativa alla transitività della rete, è improbabile che tale rete sia il risultato di un ERGM con un parametro di transitività non positivo. Ma guardando ad una singola realizzazione, la probabilità di ottenere una rete con alta transitività avendo parametro negativo non è nulla; tutte le reti possono essere osservate con qualsiasi parametrizzazione ma con diverse probabilità. Quando si valutano i valori dei parametri che hanno generato la rete osservata, è importante essere in grado di fare affermazioni precise e accurate che riassumono l'incertezza sui parametri. Per superare il problema sugli intervalli di confidenza del metodo di sti-

ma MPLE si guarda ai metodi *bootstrap* per la costruzione di intervalli di confidenza consistenti per i modelli TERGM. Si presume che le reti siano indipendenti o condizionatamente indipendenti l'una dall'altra. Questo include anche assunzione d'indipendenza condizionale di Markov di ordine k nel TERGM.

Sia $\hat{\Theta}_s$, un campione di s stime di θ costruite calcolando $\hat{\theta}$ su s campioni di $T-k$ reti estratte con reinserimento da $\{Y_{k+1}, Y_{k+2}, ..., Y_T\}$; ovvero, si vanno a campionare $T-k$ livelli dalla rete multi-livello su cui si applica il modello TERGM. È importante notare che i livelli vengono ricampionati senza modificare le configurazioni all'interno delle reti. L'algoritmo stima gli intervalli di confidenza dell'MPLE in maniera consistente, mostrando che l'MPLE è uno *M-estimator* multivariato[3], poiché è stato dimostrato che il ricampionamento bootstrap di osservazioni multivariate si traduce in stime consistenti degli intervalli di confidenza per ogni *M-estimator* multivariato consistente (Lahiri, 1992).

Sia h una qualsiasi funzione a valori scalari dei dati Y e sia θ il vettore dei parametri. Ogni stimatore $\hat{\theta}$ che risolva l'equazione

$$\sum_{t=k+1}^{T} \frac{\partial \sum_{\langle ij \rangle} ln[(\pi_{ij}^t(\theta))^{Y_{ij}^t}(1-\pi_{ij}^t(\theta))^{1-Y_{ij}^t}]}{\partial \hat{\theta}} = 0, \tag{2.6}$$

è un *M-estimator*. Pertanto, il campione *bootstrap* di MPLE fornisce una stima consistente degli intervalli di confidenza per l'MPLE.

Questo metodo non risolve solo la problematicità legata all'incertezza del metodo MPLE, ma presenta vantaggi anche rispetto la metodo MCMC-MLE. La stima attraverso MCMC-MLE è più lenta della MPLE con intervalli di confidenza *bootstrap*, divenendo quasi impraticabile per reti che non abbiano un numero piccolo di livelli. Pertanto, MPLE con intervalli di confidenza *bootstrap* è l'unica opzione in molte applicazioni empiriche. Sebbene esistano diverse problematicità legate a questo approccio di stima, le stime MCMC-MLE possono risultare più accurate per reti piccoli e con pochi livelli in quanto la consistenza delle stime di

[3] *M-estimator* multivariato è una classe di stimatori robuti che ottimizzano la funzione data; sia la verosimiglianza che la stima ai minimi quadrati sono un caso particolare di *M-estimator*.

MPLE con intervalli di confidenza *bootstrap* aumenta con il crescere del numero di osservazioni.

Capitolo 3

Caso di studio: progetto *Caviar*

3.1 I dati

Il progetto *Caviar* è stata un indagine unica nel suo genere che aveva come obiettivo l'identificazione della rete di spaccio di Montreal. Questa rete è stata osservata tra il 1994 e il 1996 da un'indagine svolta dalla Polizia di Montreal, la *Royal Canadian Mounted Police*, e altre forze dell'ordine nazionali e regionali di altri paesi (come Inghilterra, Spagna, Italia, Brasile, Portogallo e Colombia).
Il caso fu unico in quanto prevedeva un approccio investigativo che verrà poi definito *"seize and wait" strategy*, "prendi e aspetta". A differenza della maggior parte delle operazioni svolte dalle forze dell'ordine, il mandato stabilito dal Progetto *Caviar* era di sequestrare le partite di droga identificate ma di non arrestare nessuno dei partecipanti che erano stati imputati come responsabili del carico. Ciò è avvenuto nel corso di un periodo lungo 2 anni. Pertanto, sebbene 11 spedizioni siano state sequestrate durante questo periodo, gli arresti sono avvenuti solo al termine delle indagini. Ciò che offre questo caso è una rara opportunità per studiare l'evoluzione di un fenomeno di rete criminale mentre viene ostacolata dalle forze dell'ordine. La strategia investigativa intrinseca consente una valutazione del cambiamento nella struttura della rete e uno sguardo interno su come i partecipanti alla rete reagiscono e si adattano ai crescenti vincoli loro imposti.

La principale fonte di dati è costituita dalle informazioni pervenute grazie alle intercettazioni telefoniche tra i partecipanti della rete. Queste trascrizioni sono state utilizzate per creare la matrice complessiva del sistema di comunicazione dell'operazione di traffico di droga nel corso dell'indagine. Gli individui che rientravano nella rete di sorveglianza non erano tutti partecipanti all'operazione di tratta. Una prima estrazione di tutti i nomi che compaiono nei dati di sorveglianza ha portato all'identificazione di 318 individui. Da questo pool, 208 persone non sono state coinvolte nelle operazioni di tratta. La maggior parte è stata semplicemente nominata durante le numerose trascrizioni di conversazioni, ma non è mai stata rilevata. Altri che sono stati individuati non avevano un chiaro ruolo partecipativo all'interno della rete (ad es. familiari o legittimi imprenditori). La rete finale era così composta da 110 partecipanti. I dati a disposizione sono composti da 11 matrici, una per ogni sequestro effettuato dalla polizia, e contengono i dati relazionali dei soggetti coinvolti in quelle occasioni, riportando il numero di chiamate effettuate tra i soggetti e una matrice riassuntiva (rete appiattita o *flattened network*) di tutti i 110 soggetti con tutte le relazioni intercorse nel corso degli undici eventi. Per tale motivo si è deciso di affrontare l'analisi della rete multi-livello temporale come indiretta e pesata. La volontà di considerare gli archi della rete come indiretti è dovuta al fatto che la direzione della chiamata non è stata ritenuta interessante; l'interesse sta proprio nel fatto che due soggetti abbiano comunicato per organizzare l'attività di traffico e che entrambi siano in possesso di nuove informazioni grazie a questa connessione. Il numero di chiamate (complessivo di ricevute ed effettuate) tra due nodi permette di sottolineare l'intensità della relazione tra i soggetti. Per i 110 individui che partecipano alla rete non è disponibile un nominativo, pertanto, nel corso delle analisi, ci si riferirà a questi soggetti attraverso un identificativo univoco numerico, da 1 a 110. L'obiettivo di questa analisi è quello di studiare l'evoluzione della struttura di rete tramite l'analisi dei gruppi e valutare il miglior adattamento ai dati di un modello TERGM.

Grado		***Closeness centrality***		***Betweenness***	
nodo i	n_i	**nodo** i	c_i	**nodo** i	g_i
1	0.550	1	0.172	1	0.584
12	0.257	51	0.163	12	0.342
3	0.248	32	0.160	3	0.296
87	0.147	7	0.159	51	0.179
76	0.138	12	0.157	32	0.178
37	0.101	3	0.157	76	0.146
41	0.101	35	0.154	9	0.126
89	0.082	9	0.153	87	0.106
83	0.073	99	0.152	14	0.103
8	0.073	14	0.150	41	0.091

Tabella 3.1: Indici descrittivi a livello di nodo normalizzati.

3.2 Analisi descrittive

Con riferimento agli indici descrittivi presentati nel Cap. 1.4, si indaga la rete per coglierne i primi aspetti caratteristici. Guardando alla rete appiattita si ottengono informazioni di natura complessiva sulla struttura della rete che non permettono di far emergere le caratteristiche peculiari dei singoli livelli che sono indagate nella sezione successiva. Infine, si guarda alla rete multi-livello nel suo complesso per cogliere similarità tra livelli. Per tali analisi sono utilizzate le funzioni implementate nei pacchetti **igraph** (Csardi and Nepusz, 2006) e **multinet** (Magnani et al., 2021) del software R. In tutti le immagini delle reti che seguiranno, solo i nodi con il grado maggiore al novantesimo percentile sono esplicitati con il numero di riferimento univoco assegnato ad ognuno dei 110 soggetti indagati.

3.2.1 Analisi descrittiva della rete appiattita

In figura 3.1 è rappresentata la versione appiattita della rete.
Da tale rappresentazione emerge una rete connessa con 110 nodi e 205 archi con densità del 0.034. Ciò indica che di tutte le possibili connessioni solo il 3.4% è stato realizzato; dati i valori del diametro della rete pari a 5 e lunghezza del cammino medio pari a 2.65 sembra emergere che la rete funzioni efficientemente nella trasmissione delle informazioni, dati i valori bassi degli indici appena citati.

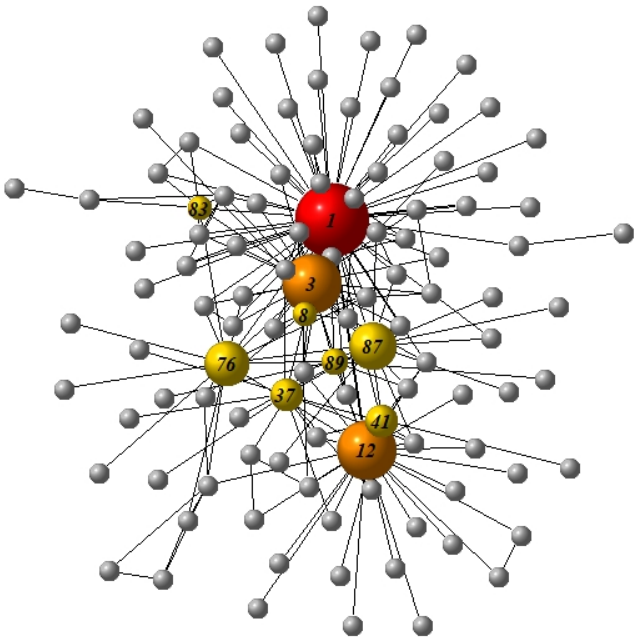

Figura 3.1: Rete appiattita con *hub* in evidenza rispetto al grado del nodo: maggiore di 50 in rosso, maggiore di 25 in arancione, maggiore di 10 in giallo.

Gli indici descrittivi a livello di nodo, riportati nella tabella 3.1, rappresentano i valori più alti di grado, *closeness centrality* e *betweenness* calcolati su tutti i nodi e fanno emergere i soggetti che fungono da intermediari facilitando la comunicazione tra nodi; i principali sono riferiti agli identificativi 1, 12 e 3. Di rilievo assoluto appare essere il nodo 1 che ha connessioni dirette con il 55% degli altri nodi presenti nella rete e un valore della *betweenness* elevato per la forte presenza del nodo 1 nei cammini minimi che connettono due nodi presenti nella rete. I valori della *closeness centrality* sono abbastanza omogenei tra i nodi che ne assumono i valori massimi, come si vede in Tabella 3.1, e non ci sono nodi che assumono valori estremamente elevati, come emerge per gli altri due indici presenti in tabella. Come si vede anche nella Figura A.1, oltre il 50% dei nodi assume valori di *closeness centrality* appartenenti all'intervallo $[0.10, 0.15]$. Si può, quindi, pensare che in tale rete, grazie alla presenza di nodi che fungono da *hub* e con alto indice di *betweenness*, la

Tempo	Densità	Diametro	Lunghezza dello *shortest path*
1	0.171	4	2.028
2	0.101	4	2.181
3	0.106	4	2.130
4	0.091	5	2.390
5	0.078	4	2.365
6	0.134	3	2.168
7	0.078	4	2.135
8	0.067	5	2.884
9	0.078	5	2.847
10	0.058	5	2.614
11	0.061	8	3.573

Tabella 3.2: Reti a livello singolo con densità, diamentro e lunghezza dello *shortest path*.

velocità di comunicazione, riassunta con l'indice di *closeness centrality* per tutti i nodi della rete, è pressoché costante a dispetto della bassa densità della rete. L'indice di transitività è pari a 0.123 e, pertanto, solo il 12.3% dei nodi adiacenti ad un vertice sono anche connessi tra loro formando triadi.

3.2.2 Analisi descrittiva delle reti a livello singolo

Passando all'analisi delle 11 reti a livello singolo, si considerano per ogni rete solo i nodi che effettivamente partecipano in quel frangente di tempo, come si può vedere in Figura 3.2. In Appendice sono riportati, in Tabella A.1, il numero di nodi, archi e chiamate totali per ogni singola rete. In Figura 3.2 si vedono, colorati di rosso, i nodi che fungono da *hub* che, oltre a essere coerenti con ciò che è emerso dall'analisi della rete appiattita, confermano l'idea di occupare una posizione di rilievo all'interno dell'organizzazione. Il nodo 1 è presente in tutte le 11 reti occupando una posizione di rilievo; anche i nodi 3 e 12 sembrano ricoprire posizioni di spicco.

I dati disponibili in Tabella 3.2, relativi agli indici descrittivi a livello di rete, mostrano un andamento particolare. Infatti, se guardiamo ai valori assunti dalla densità e dalla lunghezza dello *shortest path* sembrano avere rispettivamente un tendenza a decrescere e a crescere con il passare del tempo. Questo può essere spiegato dall'aumento dei nodi che partecipano alle singole reti. Considerando, poi, che il valore del diametro della

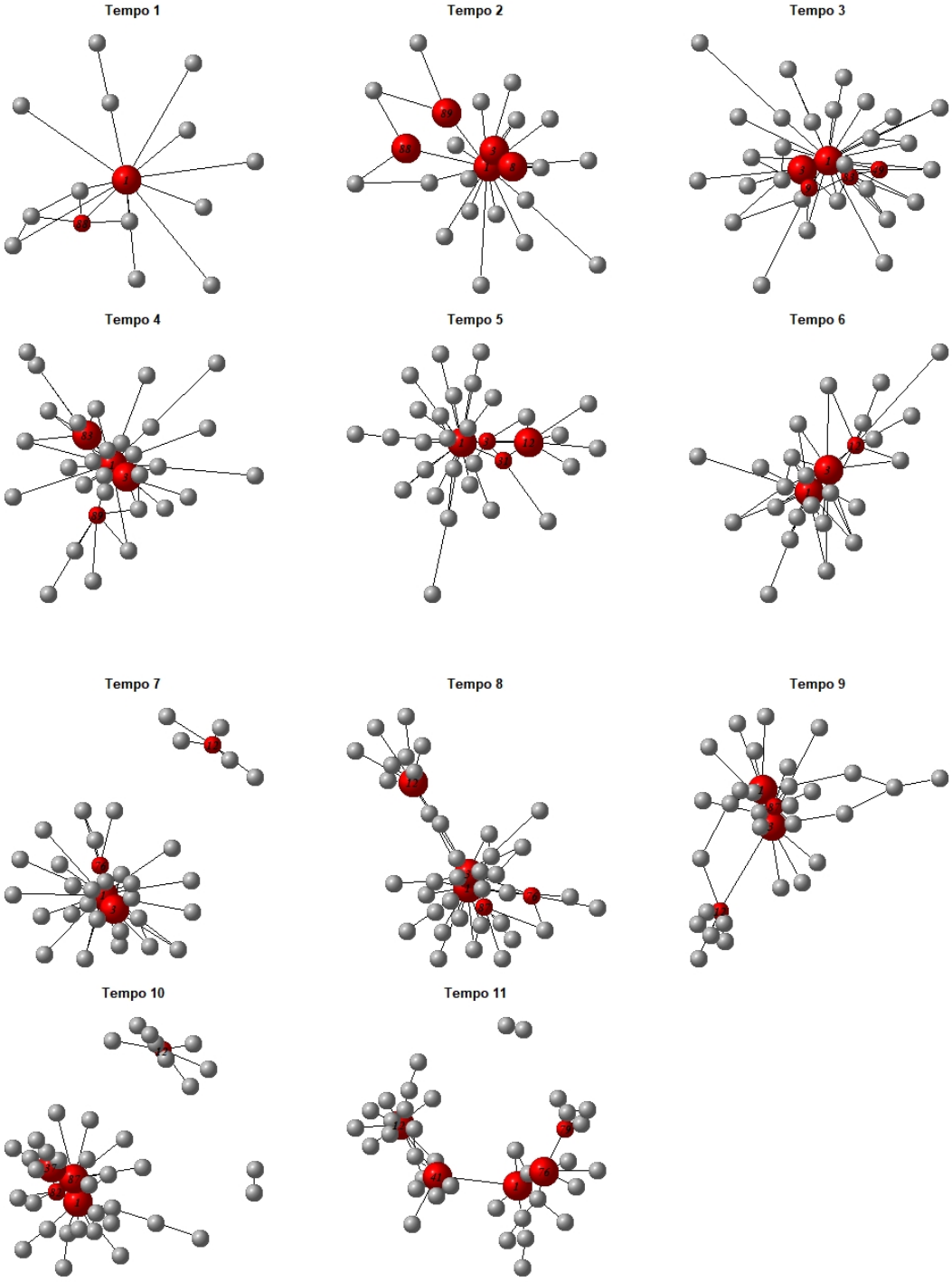

Figura 3.2: Rete a livello singolo con *hub* in evidenza: in rosso nodi con grado maggiore al 90-esimo percentile.

rete che resta stabile (escluso il tempo 11) anche se il numero di nodi è cresciuto appare una chiara modificazione della struttura comunicativa della rete con il trascorrere delle osservazioni.

Si guardi invece alla distribuzione della *closeness centrality*, riportata in Tabella 3.3, il cui supporto diminuisce con il trascorrere del tempo; tale tendenza fa emergere come la comunicazione tra i partecipanti alla rete diventi più lenta, probabilmente a causa del fatto che i ripetuti sequestri effettuati dalla polizia hanno portato alla modifica dei consolidati metodi organizzativi interni all'organizzazione.

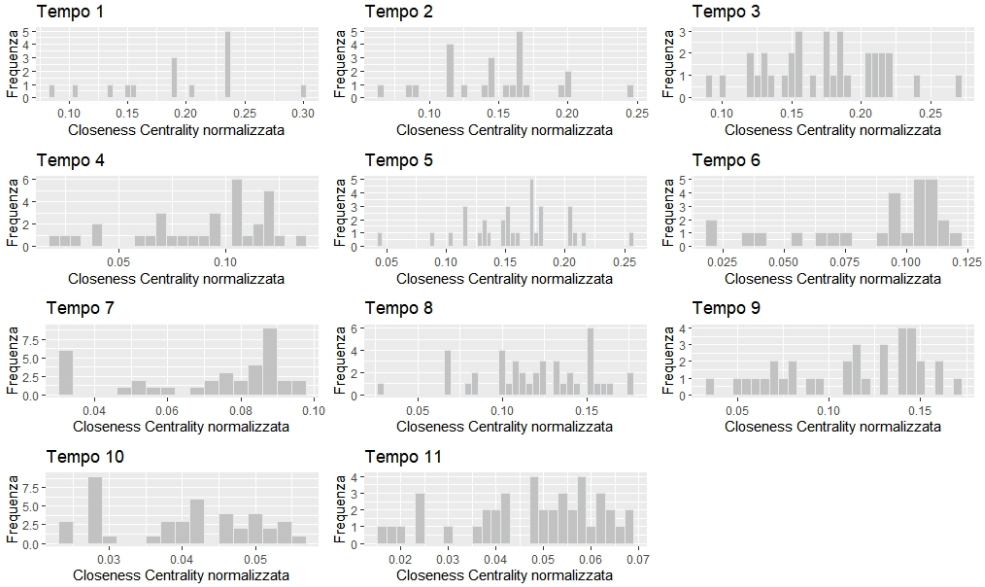

Figura 3.3: Istogrammi della *closeness centrality* per le reti a livello singolo.

3.2.3 Analisi descrittiva della rete multi-livello

Passando infine ad analizzare la rete multi-livello, ci si concentra sugli indici a livello di nodo, solo per gli attori che già dalle precedenti analisi sono apparsi avere un ruolo centrale nella struttura dell'organizzazione, per poi guardare agli indici a livello di *layer* per valutare la somiglianza tra i livelli. Non ci si concentra su tutti i nodi perchè in molti casi hanno un ruolo marginale e non sono presenti in tutte le singole reti quindi non apporterebbero informazioni di rilievo.

Per quanto riguarda gli indici a livello di nodo, come si vede in Tabella

3.3, il nodo 1 occupa una posizione di rilievo assoluto nei primi 5 periodi. Nei ultimi 6 periodi, invece, emerge che nella struttura organizzativa non è più presente un singolo vertice ma altri nodi (come 3, 12) ottengono ruoli di rilevanza.

	1	2	3	4	5	6	7	8	9	10	11
Nodo 1	12	19	27	23	22	18	24	20	10	13	7
Nodo 3	2	3	9	7	5	14	10	13	11	1	1
Nodo 12		1	2	1	8	10	5	10	8	7	12
Nodo 76		2	2	2	2	7	5	6	5	3	7
Nodo 85	3	2	4	5	3	5	4	3	3	3	3

Tabella 3.3: Grado sugli 11 livelli per i 5 nodi di maggior rilievo nella rete multi-livello; il valore mancante indica che il nodo non è presente in quel livello.

Ispezionando poi i valori di $Neighbors_{XOR}$, in Tabella 3.4, emerge come nessun livello appaia essenziale per le connessioni dei nodi principali; ovvero, anche se quel livello venisse eliminato dalla rete multi-livello i nodi che stiamo considerando manterrebbero buona parte dei loro vicini connessi sugli altri livelli. Considerando inoltre che i soggetti che partecipano all'attività criminale nei vari tempi sono diversi e alcuni di questi appaiono solo su qualche livello della rete, si possono interpretare questi valori come segno di un gruppo di nodi costantemente presente nella rete e bene collegato. Indicando la presenza di una struttura definita di soggetti che cooperano all'interno dell'organizzazione criminale.

Tempo	1	2	3	4	5	6	7	8	9	10	11
Nodo 1	1	2	5	2	3	0	3	1	1	3	1
Nodo 3	0	1	3	0	0	2	1	3	2	0	0
Nodo 12		0	0	0	1	1	1	3	2	1	5
Nodo 76		0	0	0	1	2	0	1	0	1	2
Nodo 85	1	0	0	1	0	0	1	0	0	0	0

Tabella 3.4: *Indice di* $Neighbors_{XOR}$ per i 5 nodi di maggior rilievo nella rete.

Quanto detto viene anche confermato dal $DimensionRelevance_{XOR}$, come si vede in Tabella 3.5, che assumendo per tutti e 5 i nodi valori bassi indica come nessun livello sia decisivo per la presenza nella rete e per la comunicazioni di questi nodi.

Si osserva che, tra i principali nodi considerati, le distanze multi-livello che si ottengono sono nella maggior parte dei casi di lunghezza 1 e che solo nei casi residuali è richiesto, prevalentemente, un solo passaggio in-

Tempo	1	2	3	4	5	6	7	8	9	10	11
Nodo 1	0.02	0.04	0.09	0.04	0.05	0.00	0.05	0.02	0.02	0.05	0.02
Nodo 3	0.00	0.03	0.10	0.00	0.00	0.07	0.03	0.10	0.07	0.00	0.00
Nodo 12		0.00	0.00	0.00	0.04	0.04	0.04	0.12	0.08	0.04	0.19
Nodo 76		0.00	0.00	0.00	0.06	0.12	0.00	0.06	0.00	0.06	0.12
Nodo 85	0.11	0.00	0.00	0.11	0.00	0.00	0.11	0.00	0.00	0.00	0.00

Tabella 3.5: Indice di $DimensionRelevance_{XOR}$per i 5 nodi di maggior rilievo nella rete.

termedio per essere collegati.

Guardando agli indici che confrontano di livelli si osservano i valori assunti dalla *Pair D-Correlation* riportati in Tabella A.2. Coerentemente con l'ordinamento temporale dei livelli, tale indice assume valori massimi per i tempi adiacenti a quello per cui sono calcolati e andando a decrescere nei loro valori considerando tempi più lontani. Ciò conferma l'interesse di andare a modellare la rete tramite un modello TERGM che tiene conto della dipendenza temporale dei livelli.

3.3 Analisi dei gruppi

Dopo aver effettuato le analisi descrittive, si è osservata la divisione in comunità della rete per valutare la presenza di una chiara struttura interna all'organizzazione criminale ed evidenziare eventuali modifiche di tale struttura nel corso delle 11 osservazioni.

Attraverso l'applicazione del metodo di Louvain, si sono identificate le comunità a partire dalla rete appiattita.

Come si vede in Figura 3.4, sono state identificate complessivamente 6 comunità con assortatività elevata pari a 0.644. Tale valore indica un'elevata densità interna ai gruppi e pochi archi tra di essi. A parte la comunità di due elementi in grigio, tutte le altre presentano almeno uno dei nodi che, già in fase di analisi descrittiva, erano emersi come fulcri per il diffondersi delle informazioni. Questo può indicare tali nodi come vertici del sistema organizzativo di spaccio di droga e come referenti per gli altri nodi del gruppo a cui appartengono.

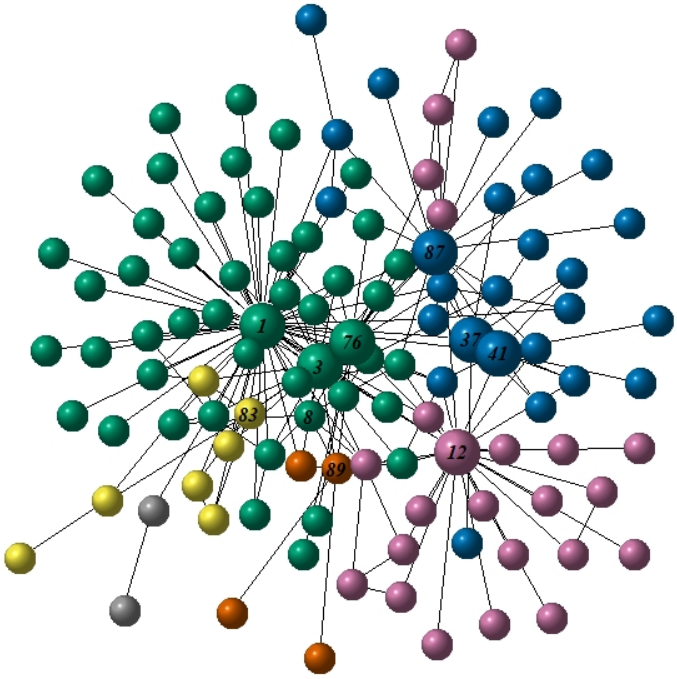

Figura 3.4: Rete appiattita con comunità.

Interessanti risultati sono emersi anche dall'analisi dei gruppi per le 11 reti a livello singolo. Si veda come in Tabella 3.2 l'indice di assortatività nei vari tempi assuma valori sempre più elevati facendo si che le comunità ottenute siano progressivamente più coese. Infatti, già valori dell'indice superiori al 0.3, nelle applicazioni reali, sono ritenuti significativi di comunità ben definite all'interno di una rete. Si veda anche la Figura 3.5, nella quale sono rappresentate le reti con le comunità ottenute. Con la progressione degli eventi, le comunità si delineano più a sè stanti e con dei nodi di riferimento che restano coerenti con quelli identificati per la rete appiattita. Inoltre, quanto affermato per l'assortatività delle reti, è visibile anche in Figura 3.5 dove le comunità sembrano effettivamente consolidarsi chiaramente. Tale fenomeno può essere imputato all'azione delle forze di polizia che hanno portato la rete a modificare la propria struttura dando un ruolo centrale di vertice organizzativo a pochi nodi che si rivolgono poi autonomamente alle comunità di nodi di cui sono il riferimento.

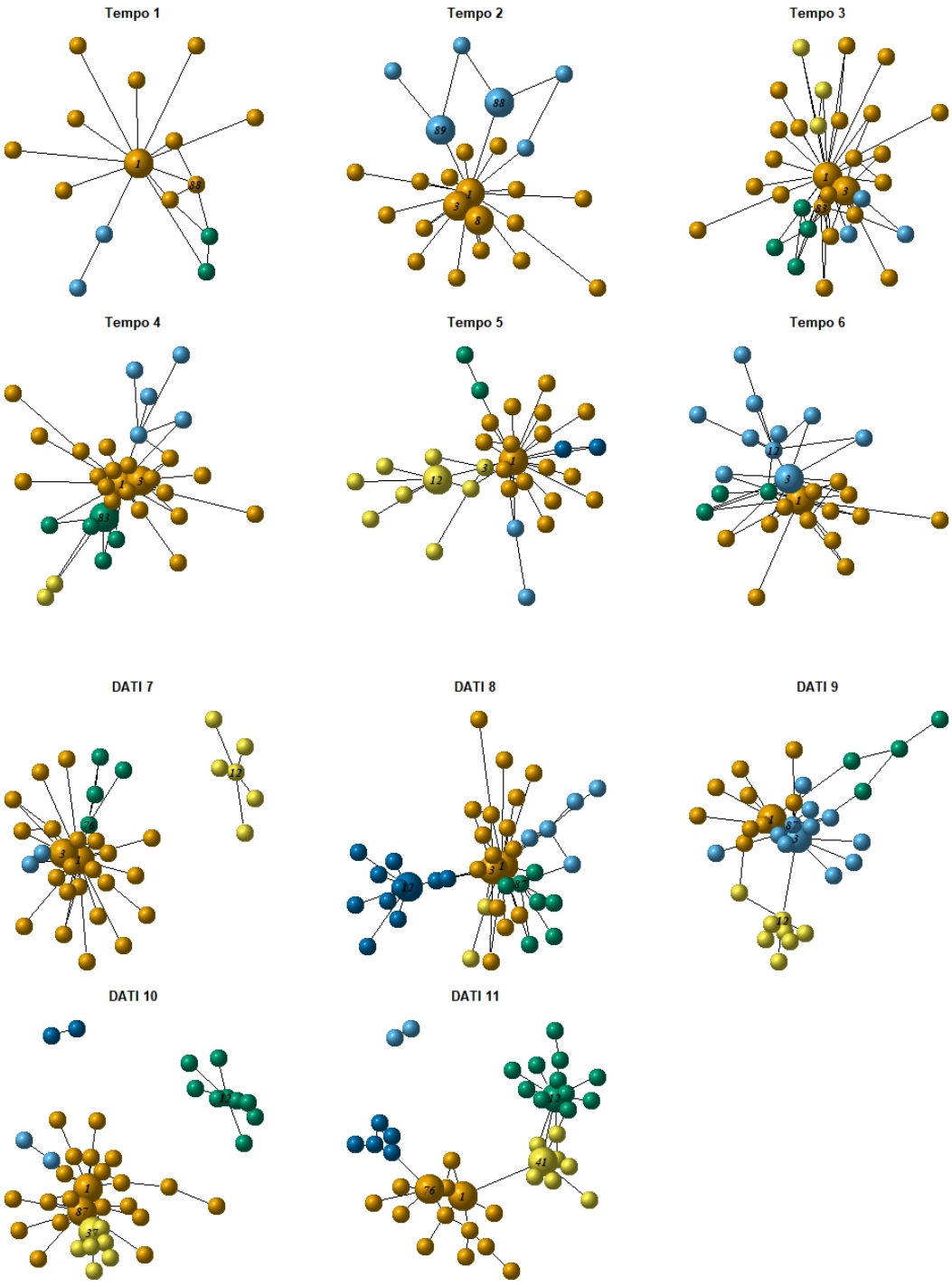

Figura 3.5: Rete a livello singolo con comunità.

Tempo	Assortatività	# di gruppi
1	0.289	3
2	0.699	2
3	0.357	4
4	0.635	4
5	0.572	5
6	0.347	3
7	0.885	4
8	0.688	5
9	0.748	4
10	0.770	5
11	0.865	5

Tabella 3.6: Indice di assortatività e numero di gruppi per le reti a livello singolo.

Considerando, infine, la divisione in comunità per la rete multi-livello nel suo complesso otteniamo 7 comunità, come si vede dalla Figura 3.6 (in Appendice è riportata la composizione delle comunità in Tabella A.3). Emerge chiaramente come nel corso dei vari tempi la struttura della rete si divida in comunità distinte, come già detto in precedenza considerando le reti a livello singolo. Va specificato che il metodo di Louvain generalizzato per reti multi-livello, utilizzato per la determinazione delle comunità, è stato utilizzando testando diversi valori del parametro ω e ottenendo risultati tra loro coerenti; si è quindi deciso di fissare ω pari a 1 massimizzando la modularità. Data questa assunzione ogni nodo, inserito in una comunità, non può cambiare comunità nel corso degli eventi. Dato che l'algoritmo di stima non considera l'ordinamento temporale degli eventi e guarda alla rete multi-livello nel suo complesso la restrizione imposta ai nodi, per quanto lontana dalla realtà, ci permette di vedere con maggiore chiarezza lo sviluppo delle comunità nei vari tempi, come si vede ad esempio per la comunità di colore rosso (comunità 3).
Al "Tempo 1" la rete presenta la comunità in giallo (comunità 1) e altre 3 comunità residuali con pochi nodi. La comunità 1, coesa attorno al nodo centrale che è 1, comprende, infatti, quasi la totalità dei nodi della rete. Iniziano poi a svilupparsi comunità concorrenti con il proprio nodo di riferimento. Se prendiamo, ad esempio, la comunità di colore rosso (comunità 3), questa appare al "Tempo 2" con il solo nodo 12 e, percorrendo i livelli della rete, si configura come una comunità assestante

e ben delineata rispetto alla comunità gialla iniziale. Il nodo 3 è invece afferente alla comunità blu petrolio (comunità 4). La comunità verde (comunità 6) ha uno sviluppo simili a quello della comunità rossa ma a partire dal "Tempo 7".

Si conferma l'intuizione precedente per cui, dati i risultati delle analisi di *clustering*, la struttura dell'organizzazione criminale sia mutata nel corso del tempo vedendo la formazioni progressiva di comunità che prendono sempre più spazio nel corso degli eventi e per le quali, nella maggior parte dei casi, è possibile identificare un nodo di riferimento sensato in quanto stiamo parlando di una rete sociale di un organizzazione criminale.

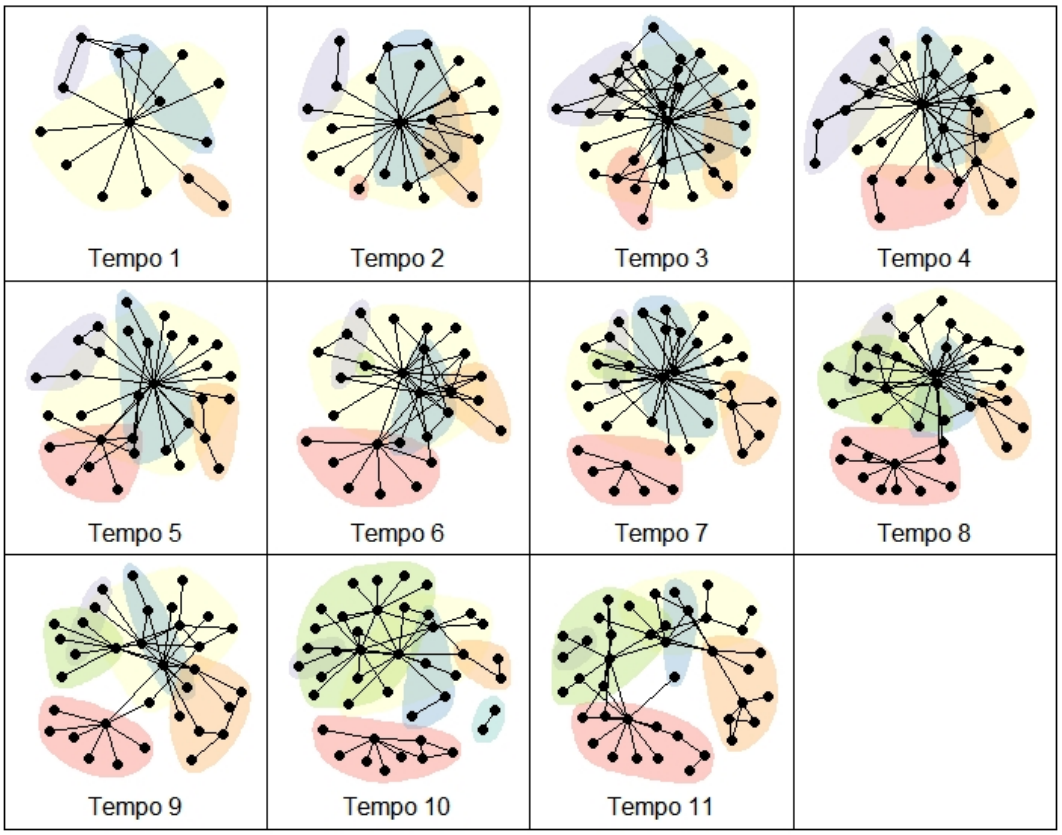

Figura 3.6: Rete multi-livello con comunità. Comunità 1 in giallo, 2 in viola, 3 in rosso, 4 in blu petrolio, 5 in arancione, 6 in verde e 7 in azzurro.

3.4 Modello

Grazie al pacchetto **btergm** (Leifeld et al., 2018) che utilizza il metodo di stima di massima pseudo-verosimiglianza con intervalli di confidenza *bootstrap*, descritto da Desmarais and Cranmer (2012), si prosegue l'analisi valutando l'adattamento dei dati attraverso il modello TERGM, concludendo poi valutandone la capacità previsiva.
Non essendo disponibili variabili esplicative per i nodi della rete, si è deciso di considerare il gruppo di appartenenza, calcolato sulle singole reti. Questa variabile associata ai nodi di ogni singolo livello della rete è definita *gruppo*. Si crea inoltre una variabile associata ai possibili archi di ogni livello della rete, definita come *mat.gruppo*. Si è costruita una matrice 110×110 dove per ogni elemento (i, j) con $i \neq j$ rappresenta l'arco tra i nodi i e j; gli elementi diagonali sono posti pari a 0. Se l'arco tra i e j collega due nodi appartenenti allo stesso gruppo nella matrice, in (i, j), si inserisce il valore 1 altrimenti 0. Per questa matrici si è fatto riferimento ai gruppi definiti dall'algoritmo di Louvain generalizzato per reti multi-livello con paramentro ω pari a 1. Il modello che si è scelto è quello che considera i seguenti termini:

- *edges*: aggiunge una statistica di rete pari al numero di archi della rete e restituisce, complessivamente, una misura di densità della rete.
- *gwesp*: aggiunge una statistica di rete uguale alla distribuzione geometrica pesata dai nodi condivisi tra nodi adiacenti, con parametro di decadimento non negativo (Caimo, 2019).
- *edgecov(mat.gruppo)*: aggiunge una statistica di rete al modello uguale alla somma dei valori della covariata per ogni arco che appartiene alla rete.
- *nodefactor("gruppo")*: questo termine aggiunge più statistiche di rete al modello, pari al numero di modalità assunte dalla variabile. Ognuna di queste statistiche fornisce il numero di volte in cui il nodo con quella modalità dell'attributo appartiene ad un arco della rete.

- *memory(type="stability", lag=1)*: questo termine va a controllare la stabilità della memoria degli archi; controlla che ogni arco presente o non presente in un determinato livello (tempo) della rete venga trasferito al momento successivo come arco presente o non presente nella rete. Un valore positivo e significativo del coefficiente stimato indica che c'è una grande stabilità dei legami tra nodi nel corso del tempo.

In Tabella 3.7 si vedono le stime dei parametri e i relativi intervalli di confidenza, ottenuti con 5000 replicazioni *bootstrap*.

	Stima coef.	2.5%	97.5%
edges	-31.78	-33.63	-31.43
gwesp.fixed.0	0.96	0.80	1.20
edgecov.grup.mat	0.28	0.23	0.33
nodefactor.gruppo.1	14.40	13.89	15.49
nodefactor.gruppo.2	14.24	13.63	15.24
nodefactor.gruppo.3	13.91	13.16	14.30
nodefactor.gruppo.4	13.92	13.65	14.84
nodefactor.gruppo.5	12.98	0.18	14.49
nodefactor.gruppo.6	13.60	13.30	14.63
nodefactor.gruppo.7	14.33	13.76	15.49
edgecov.memory	2.25	2.13	2.38

Tabella 3.7: Stime dei coefficienti e intervalli di confidenza *bootstrap*.

Tutti i coefficienti stimati sono significativi in quanto nessuno degli intervalli contiene il valore 0. Si possono, inoltre, fare le seguenti considerazioni:

- Guardando a *edges* si vede che la probabilità che due nodi qualsiasi all'interno della rete siano connessi è molto bassa, ma ciò è sensato in quanto è legato alla densità della rete.
- La statistica relativa alla covariata degli archi *mat.gruppo* (*edgecov.grup.mat*) incrementa la probabilità della formazione di un arco tra nodi appartenenti allo stesso gruppo rispetto a nodi di gruppi distinti del 57%. Questo va a segnalare come vi sia un forte legame di appartenenza al gruppo; o meglio, se un soggetto collabora ad una delle attività illecite verrà inserito in un gruppo con relativo nodo di riferimento (come emerso dalle precedenti analisi) e tenderà a connettersi solo con altri individui del suo gruppo che probabilmente

gestiscono aspetti simili all'interno dell'organizzazione. Un'immagine che potrebbe aiutare a rendere più chiara la situazione è quella di un'azienda con diversi uffici che si occupano di specifici ambiti (finanza, marketing, produzione, ...). Se si deve lavorare ad un progetto i referenti di ogni ufficio si organizzano tra di loro e selezionano soggetti del loro team. I soggetti che lavorano al progetto collaboreranno principalmente con i membri dello stesso ufficio e contatteranno solo in caso di necessità o delucidazioni gli altri uffici.

- Le statistiche di rete legate alla covariata discreta dei nodi, *gruppo*, dice di quanto l'appartenenza di un nodo ad un determinato gruppo incrementi la probabilità di creare un arco nella rete; tali coefficienti indicano un effetto positivo sulla probabilità e assumendo tutti valori simili sostiene che l'appartenenza a nessuno dei gruppi in particolare porta ad una maggiore probabilità di creare un arco nella rete; appartenere ad uno specifico gruppo, rispetto agli altri, comporta una maggiore difficoltà per i nodi a partecipare alle attività dell'organizzazione.

- *Memory*, controllando la persistenza degli archi presenti o non presenti in un tempo rispetto a quelli del tempo successivo e assumendo un coefficiente di valore positivo e significativo, indica che all'interno della rete c'è una grande stabilità nei legami di alleanza nel tempo.

Proseguendo all'analisi diagnostica del modello adottato attraverso i grafici in Figura 3.7 si vede un adattamento complessivo soddisfacente. I primi cinque grafici mostrano le distribuzioni statistiche endogene della rete osservata a confronto con quelle ottenute attraverso la simulazione di 5000 reti dal modello. L'interpretazione di questi grafici è semplice; si dice che il modello si adatta meglio tanto più le mediane dei box-plot (basati sulle reti simulate) sono vicine alla linea tracciata dall'effettivo valore che queste statistiche assumono per la rete osservata. Dai grafici emerge un evidente scostamento tra valori osservati e simulati solo per il secondo grafico mostrando invece un buon adattamento in tutte le

altre situazioni. Passando all'ultimo grafico, si vede riportata la curva ROC, di colore rosso. La curva ROC ha come obiettivo per un perfetto adattamento l'angolo in alto a sinistra. Guardando il grafico si può dire che per quanto riguarda la curva ROC il modello di adatta in maniera soddisfacente con un AUC pari a 0.859.

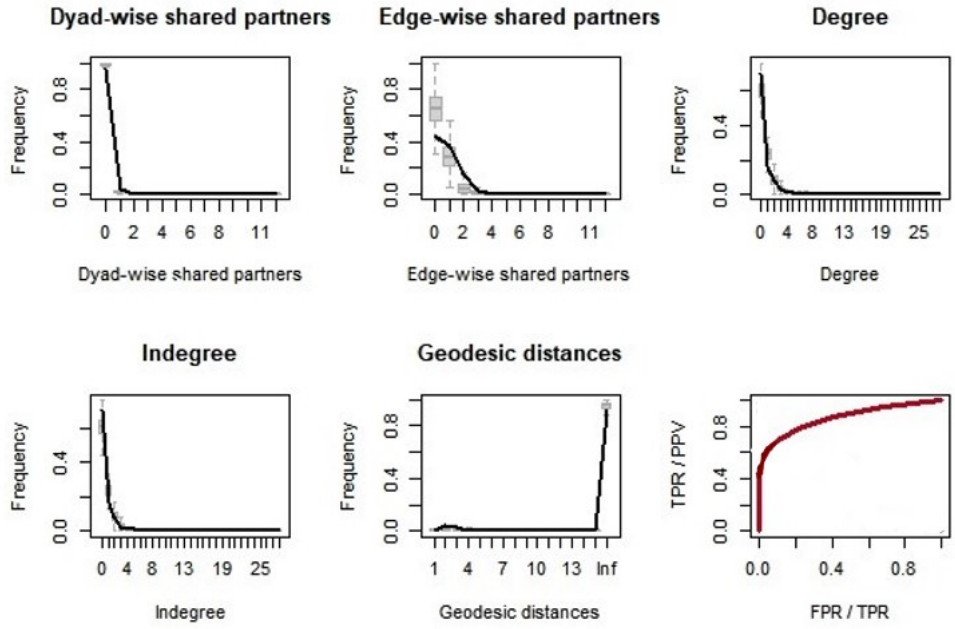

Figura 3.7: Adattamento del modello.

Valutato, complessivamente, in maniera positiva l'adattamento del modello è interessante stabilirne la capacità previsiva; si vuole verificare che, stimando nuovamente il modello sui soli primi 10 livelli temporali, la previsione per livello 11 sia soddisfacente. Per fare ciò la variabile *mat.gruppo*, inserita nel modello, viene ricalcolata attraverso l'algoritmo di Louvain generalizzato per reti multi-livello solo sui primi 10 livelli.

	Estimate	2.5%	97.5%
edges	-31.68	-33.54	-31.27
gwesp.fixed.0	0.89	0.74	1.18
edgecov.grup.mat	0.36	0.25	0.46
nodefactor.gruppo.1	13.83	13.31	15.00
nodefactor.gruppo.2	14.11	13.44	15.18
nodefactor.gruppo.3	13.68	12.06	14.29
nodefactor.gruppo.4	13.32	12.78	14.36
nodefactor.gruppo.5	13.16	0.19	14.73
nodefactor.gruppo.6	14.04	13.78	15.06
nodefactor.gruppo.7	14.33	13.65	15.51
edgecov.memory	2.21	2.09	2.31

Tabella 3.8: Stime dei coefficienti e intervalli di confidenza *bootstrap* considerando solo i primi 10 livelli della rete multi-livello temporale.

I coefficienti stimati, riportati in Tabella 3.8, sono coerenti con quanto visto prima e significativi. Guardando poi alla previsione per il livello 11, con riferimento alla Figura 3.8, i primi cinque grafici confrontano la distribuzione delle statistiche endogene di rete osservate e quelle ottenute attraverso la simulazione di 5000 reti dal modello. In effetti, le reti simulate predicono ragionevolmente bene l'ultimo livello osservato (Tempo 11) presentando lo stesso problema di adattamento nel secondo grafico già rilevato con la precedente stima. L'ultimo grafico presenta la curva ROC, di colore rosso. La curva ROC mostra una buona prestazione predittiva del modello con il valore di AUC pari a 0.773. In conclusione, il modello stimato ha un buon adattamento complessivo ai dati.

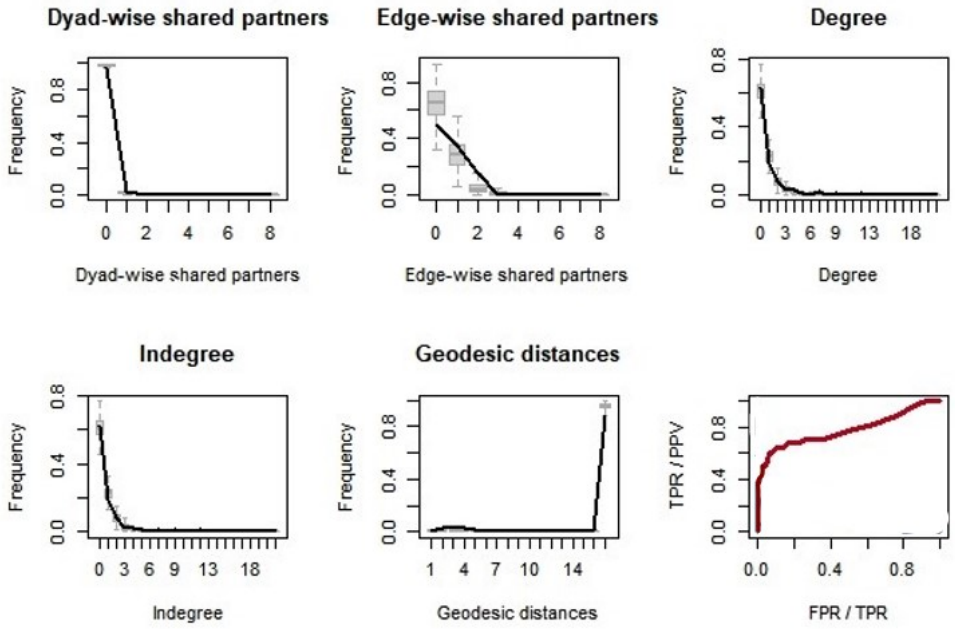

Figura 3.8: Adattamento del modello, stimato sui primi 10 livelli, ai dati relativi al livello 11.

Capitolo 4

Conclusioni

Data l'importanza che le relazioni interpersonali ricoprono nelle dinamiche della vita umana, in questo elaborato si è voluto studiare lo sviluppo della rete di un'organizzazione criminale canadese nel mirino della polizia che, attraverso una decisiva e innovativa tecnica investigativa, ha portato all'arresto dei membri dell'organizzazione coinvolti negli undici episodi di traffico di stupefacenti internazionale, sempre sventati dalla polizia canadese che provvedeva al sequestro.

Nel complesso sono stati utilizzati i dati relativi alle chiamate intercorse tra i membri dell'organizzazione nel corso di 11 eventi considerando le singole reti indirette e pesate. I pesi assegnati ad ogni nodo sono pari al numero di chiamate che due individui si sono scambiati nel corso di quell'evento.

Sfruttando le informazioni provenienti dagli indici descrittivi di rete è emerso che la rete presenta dei nodi che assumo una posizione di prominente rilievo all'interno dell'organizzazione. I principali soggetti individuati come *hub* della rete sono quelli il cui identificativo corrisponde a 1, 12 e 3. Inoltre, dagli indici descrittivi emerge che la comunicazione è rapida per tutti i soggetti grazie alla presenta di nodi che fungono a ponte e permettono una trasmissione rapida delle informazioni. Questa caratterista nel corso degli eventi sembra, però, risentire delle attività di sabotaggio della polizia.

Passando all'analisi dei gruppi si è confermata la presenza di nodi "pon-

te" all'interno dell'organizzazione e ha permesso di vederne l'evoluzione strutturale nel corso del tempo. Infatti, grazie all'applicazione del metodo di Louvain per la rete appiattita e per le singole reti e della sua versione generalizzata per la rete multi-livello è stato possibile vedere come, nel corso degli eventi, la struttura comunicativa della rete si sia modificata. Già dalle prime fasi temporali emerge come la rete sia organizzata in gruppi che hanno un individuo al vertice; inizialmente, l'individuo al vertice sembra essere solo uno, il soggetto identificato dal nodo 1. Questa struttura di comando verticale si mantiene nel corso degli eventi, ma emergono altri nodi che occupano posizioni di rilievo e fungono da ponti tra le comunità. Altra peculiarità è che il corso degli eventi porta le comunità, che sono state delineate, ad essere più coese al loro interno emostrare come quasi esclusivamente i nodi "al vertice" dei gruppi comunicano con soggetti degli altri gruppi; probabilmente ciò avviene in risposta alla crescente pressione che l'attività della polizia generava con i continui sequestri della droga in loro possesso.

Infine, l'applicazione del modello TERGM ha permesso di confermare le evidenze emerse dalle precedenti analisi.

Dal modello è emerso che: la divisione in gruppi dell'organizzazione, inserita attraverso variabili esogene di nodi e archi, porta ad un buon adattamento del modello, confermando una forte coesione interna delle comunità formatesi nel corso degli eventi.

In conclusione, per quanto visto nel corso delle analisi, si può valutare che l'attività della polizia di Montreal ha portato alla modificazione interna della struttura della rete. Non tanto per quanto riguarda la dinamica organizzativa, per cui rimane un soggetto di vertice a gestione di un gruppo di soggetti, ma per quanto riguarda l'ampliamento di tale assetto, si è visto l'affiorare di vari soggetti al comando di vari gruppi che segmentano le operazioni organizzative in diversi gruppi operativi. La divisione in gruppi ha mostrato di essere utile anche per la modellazione e la previsione dei dati. Si conclude dicendo che sarebbe interessante approfondire l'analisi di questi dati attraverso l'applicazione di altri mo-

delli per l'analisi di dati di rete, come ad esempio modelli a spazi latenti. Inoltre, se si potesse avere a disposizione maggiori variabili esplicative relative agli attori di questa rete, si potrebbero studiare le peculiarità dei soggetti che occupano posizioni di rilievo e le caratteristiche dei gruppi che vengono a formarsi.

Bibliografia

A.-L. Barabási. Linked: The new science of networks, 2003.

M. Berlingerio, M. Coscia, F. Giannotti, A. Monreale, and D. Pedreschi. Foundations of multidimensional network analysis. In *2011 international conference on advances in social networks analysis and mining*, pages 485–489. IEEE, 2011.

J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36 (2):192–225, 1974.

Z. Burda, J. Jurkiewicz, and A. Krzywicki. Network transitivity and matrix models. *Phys. Rev. E*, 69:026106, Feb 2004. doi: 10.1103/PhysRevE.69.026106. URL `https://link.aps.org/doi/10.1103/PhysRevE.69.026106`.

I. Caimo, Alberto e Gollini. Modelling weighted signed networks. In *SIS 2019: Conference of the Italian Statistical Society Milan, 18-21 June 2019*. Pearson, 2019.

G. Csardi and T. Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006. URL `https://igraph.org`.

B. A. Desmarais and S. J. Cranmer. Statistical mechanics of networks: Estimation and uncertainty. *Physica A: Statistical Mechanics and its Applications*, 391(4):1865–1876, 2012.

M. E. Dickison, M. Magnani, and L. Rossi. *Multilayer social networks*. Cambridge University Press, 2016.

E. W. Dijkstra. A note on two problems in connexion with graphs. In *Edsger Wybe Dijkstra: His Life, Work, and Legacy*, pages 287–290. 2022.

P. Erdős, A. Rényi, et al. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

O. Frank. Statistical analysis of change in networks. *Statistica Neerlandica*, 45(3):283–293, 1991.

O. Frank and D. Strauss. Markov graphs. *Journal of the american Statistical association*, 81(395):832–842, 1986.

C. J. Geyer and E. A. Thompson. Constrained monte carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(3):657–683, 1992.

P. W. Holland and S. Leinhardt. An exponential family of probability distributions for directed graphs. *Journal of the american Statistical association*, 76(373):33–50, 1981.

E. D. Kolaczyk. *Statistical analysis of network data with R / Eric D. Kolaczyk, Gabor Csardi*. Use R! Springer, New York [etc, 2014. ISBN 9781493909827.

S. Lahiri. On bootstrapping m-estimators. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 157–170, 1992.

E. Lazega and M. Van Duijn. Position in formal structure, personal characteristics and choices of advisors in a law firm: A logistic regression model for dyadic network data. *Social networks*, 19(4):375–397, 1997.

P. Leifeld, S. J. Cranmer, and B. A. Desmarais. Temporal exponential random graph models with btergm: Estimation and bootstrap confidence intervals. *Journal of Statistical Software*, 83(6):1–36, 2018. doi: 10.18637/jss.v083.i06.

M. Magnani and L. Rossi. Pareto distance for multi-layer network analysis. In *International Conference on Social Computing,*

Behavioral-Cultural Modeling, and Prediction, pages 249–256. Springer, 2013.

M. Magnani, L. Rossi, and D. Vega. Analysis of multiplex social networks with R. *Journal of Statistical Software*, 98(8):1–30, 2021. doi: 10.18637/jss.v098.i08.

P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *science*, 328(5980):876–878, 2010.

K. Orman, V. Labatut, and H. Cherifi. An empirical study of the relation between community structure and transitivity. In *Complex Networks*, pages 99–110. Springer, 2013.

J. Park and M. E. J. Newman. Statistical mechanics of networks. *Phys. Rev. E*, 70:066117, Dec 2004. doi: 10.1103/PhysRevE.70.066117. URL `https://link.aps.org/doi/10.1103/PhysRevE.70.066117`.

A. Solé-Ribalta, M. De Domenico, S. Gómez, and A. Arenas. Centrality rankings in multiplex networks. In *Proceedings of the 2014 ACM conference on Web science*, pages 149–155, 2014.

D. Strauss and M. Ikeda. Pseudolikelihood estimation for social networks. *Journal of the American statistical association*, 85(409):204–212, 1990.

J. Travers and S. Milgram. An experimental study of the small world problem. In *Social networks*, pages 179–197. Elsevier, 1977.

S. Wasserman and P. Pattison. Logit models and logistic regressions for social networks: I. an introduction to markov graphs andp. *Psychometrika*, 61(3):401–425, 1996.

S. Wasserman, K. Faust, et al. Social network analysis: Methods and applications. 1994.

Appendice A

APPENDICE

A.1 TABELLE E IMMAGINI

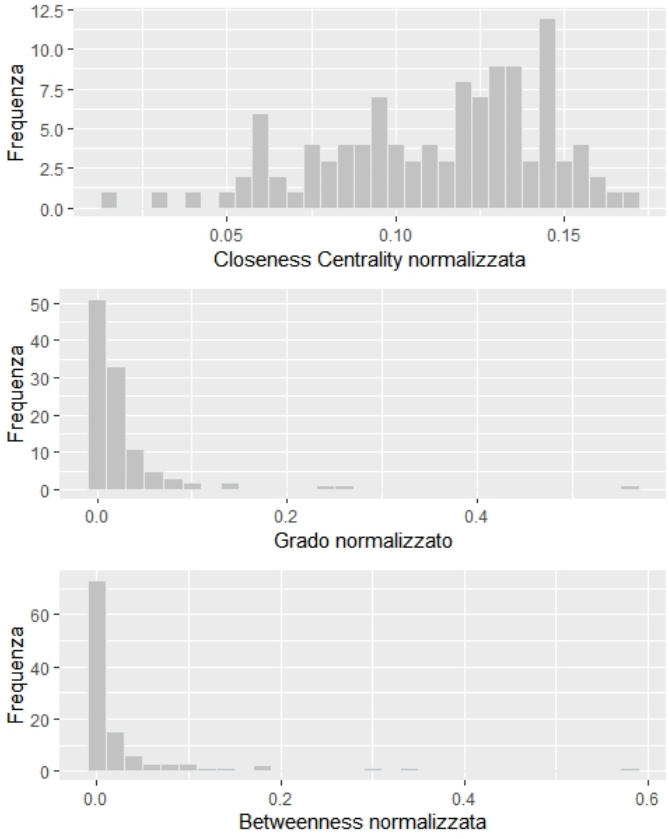

Figura A.1: Indici descrittivi a livello di nodo della rete appiattita.

Tempo	# di nodi	# di archi	# di chiamate totali
1	15	18	59
2	24	28	120
3	33	56	282
4	33	48	456
5	32	39	162
6	27	47	668
7	36	49	380
8	42	58	335
9	34	44	317
10	42	50	478
11	41	50	355

Tabella A.1: Reti a livello singolo con numero di nodi, numero di archi e il numero di chiamate totali.

Tabella A.2: Indice *Pair D-Correlation* per la rete multi-livello.

	Tempo.1	Tempo.2	Tempo.3	Tempo.4	Tempo.5	Tempo.6	Tempo.7	Tempo.8	Tempo.9	Tempo.10	Tempo.11
Tempo 1	1.00	0.35	0.19	0.22	0.21	0.12	0.16	0.09	0.09	0.08	0.01
Tempo 2	0.35	1.00	0.27	0.27	0.22	0.14	0.13	0.09	0.09	0.07	0.03
Tempo 3	0.19	0.27	1.00	0.39	0.27	0.17	0.17	0.14	0.11	0.09	0.07
Tempo 4	0.22	0.27	0.39	1.00	0.32	0.22	0.18	0.14	0.15	0.11	0.08
Tempo 5	0.21	0.22	0.27	0.32	1.00	0.39	0.28	0.20	0.19	0.14	0.11
Tempo 6	0.12	0.14	0.17	0.22	0.39	1.00	0.39	0.38	0.28	0.17	0.15
Tempo 7	0.16	0.13	0.17	0.18	0.28	0.39	1.00	0.34	0.24	0.15	0.12
Tempo 8	0.09	0.09	0.14	0.14	0.20	0.38	0.34	1.00	0.31	0.21	0.12
Tempo 9	0.09	0.09	0.11	0.15	0.19	0.28	0.24	0.31	1.00	0.25	0.25
Tempo 10	0.08	0.07	0.09	0.11	0.14	0.17	0.15	0.21	0.25	1.00	0.27
Tempo 11	0.01	0.03	0.07	0.08	0.11	0.15	0.12	0.12	0.25	0.27	1.00

Nodo	Livello	Comunità
103	Tempo 10	0
104	Tempo 10	0
77	Tempo 8	1
77	Tempo 6	1
77	Tempo 7	1
20	Tempo 8	1
20	Tempo 6	1
20	Tempo 7	1
97	Tempo 2	1
10	Tempo 2	1
10	Tempo 3	1
55	Tempo 5	1
55	Tempo 7	1
55	Tempo 2	1
55	Tempo 3	1
56	Tempo 2	1
56	Tempo 3	1
47	Tempo 4	1
47	Tempo 5	1
47	Tempo 2	1
98	Tempo 2	1
11	Tempo 9	1
11	Tempo 8	1
11	Tempo 11	1
11	Tempo 4	1
11	Tempo 5	1
11	Tempo 6	1
11	Tempo 7	1
11	Tempo 2	1
11	Tempo 3	1
48	Tempo 3	1
32	Tempo 5	1
32	Tempo 3	1
35	Tempo 8	1
35	Tempo 4	1
35	Tempo 3	1
50	Tempo 3	1
84	Tempo 10	1
84	Tempo 8	1
84	Tempo 11	1
84	Tempo 4	1
84	Tempo 5	1
84	Tempo 6	1
84	Tempo 3	1
34	Tempo 8	1
34	Tempo 5	1
34	Tempo 7	1
34	Tempo 3	1
49	Tempo 4	1
49	Tempo 3	1
64	Tempo 1	1
64	Tempo 2	1

Nodo	Livello	Comunità
90	Tempo 9	1
90	Tempo 4	1
90	Tempo 1	1
90	Tempo 2	1
90	Tempo 3	1
4	Tempo 10	1
4	Tempo 8	1
4	Tempo 4	1
4	Tempo 5	1
4	Tempo 6	1
4	Tempo 7	1
4	Tempo 1	1
4	Tempo 3	1
2	Tempo 9	1
2	Tempo 8	1
2	Tempo 4	1
2	Tempo 5	1
2	Tempo 6	1
2	Tempo 7	1
2	Tempo 1	1
2	Tempo 2	1
2	Tempo 3	1
54	Tempo 1	1
5	Tempo 4	1
5	Tempo 5	1
5	Tempo 6	1
5	Tempo 7	1
5	Tempo 1	1
5	Tempo 2	1
5	Tempo 3	1
1	Tempo 9	1
1	Tempo 10	1
1	Tempo 8	1
1	Tempo 11	1
1	Tempo 4	1
1	Tempo 5	1
1	Tempo 6	1
1	Tempo 7	1
1	Tempo 1	1
1	Tempo 2	1
1	Tempo 3	1
53	Tempo 4	1
15	Tempo 4	1
15	Tempo 5	1
15	Tempo 6	1
15	Tempo 7	1
100	Tempo 5	1
82	Tempo 9	1
82	Tempo 10	1
82	Tempo 8	1
82	Tempo 11	1
82	Tempo 5	1
82	Tempo 6	1
67	Tempo 8	1

Nodo	Livello	Comunità
36	Tempo 9	1
36	Tempo 8	1
36	Tempo 11	1
62	Tempo 7	1
69	Tempo 7	1
28	Tempo 8	1
28	Tempo 7	1
68	Tempo 7	1
46	Tempo 9	1
46	Tempo 10	1
46	Tempo 11	1
96	Tempo 9	1
96	Tempo 10	1
96	Tempo 11	1
30	Tempo 9	1
70	Tempo 10	1
86	Tempo 10	2
86	Tempo 8	2
86	Tempo 11	2
86	Tempo 4	2
86	Tempo 5	2
86	Tempo 2	2
86	Tempo 3	2
52	Tempo 4	2
52	Tempo 3	2
107	Tempo 4	2
107	Tempo 3	2
6	Tempo 9	2
6	Tempo 8	2
6	Tempo 4	2
6	Tempo 5	2
6	Tempo 6	2
6	Tempo 7	2
6	Tempo 1	2
6	Tempo 2	2
6	Tempo 3	2
83	Tempo 9	2
83	Tempo 10	2
83	Tempo 8	2
83	Tempo 11	2
83	Tempo 4	2
83	Tempo 5	2
83	Tempo 6	2
83	Tempo 7	2
83	Tempo 1	2
83	Tempo 2	2
83	Tempo 3	2
63	Tempo 4	2
106	Tempo 4	2
108	Tempo 5	2
66	Tempo 11	3
26	Tempo 11	3
12	Tempo 9	3

Nodo	Livello	Comunità
12	Tempo 10	3
12	Tempo 8	3
12	Tempo 11	3
12	Tempo 4	3
12	Tempo 5	3
12	Tempo 6	3
12	Tempo 7	3
12	Tempo 2	3
12	Tempo 3	3
13	Tempo 9	3
13	Tempo 8	3
13	Tempo 11	3
13	Tempo 4	3
13	Tempo 5	3
13	Tempo 6	3
13	Tempo 3	3
51	Tempo 4	3
51	Tempo 3	3
31	Tempo 4	3
31	Tempo 5	3
31	Tempo 6	3
14	Tempo 9	3
14	Tempo 10	3
14	Tempo 8	3
14	Tempo 11	3
14	Tempo 4	3
14	Tempo 6	3
14	Tempo 7	3
18	Tempo 9	3
18	Tempo 10	3
18	Tempo 8	3
18	Tempo 11	3
18	Tempo 5	3
18	Tempo 6	3
18	Tempo 7	3
17	Tempo 9	3
17	Tempo 10	3
17	Tempo 8	3
17	Tempo 11	3
17	Tempo 5	3
17	Tempo 6	3
17	Tempo 7	3
25	Tempo 8	3
25	Tempo 5	3
25	Tempo 6	3
23	Tempo 8	3
33	Tempo 8	3
80	Tempo 8	3
75	Tempo 7	3
22	Tempo 10	3
22	Tempo 8	3
22	Tempo 7	3
16	Tempo 9	3
16	Tempo 10	3

Nodo	Livello	Comunità
16	Tempo 8	3
16	Tempo 11	3
16	Tempo 7	3
29	Tempo 9	3
24	Tempo 10	3
24	Tempo 11	3
58	Tempo 10	3
58	Tempo 11	3
65	Tempo 10	3
65	Tempo 11	3
9	Tempo 10	4
9	Tempo 8	4
9	Tempo 4	4
9	Tempo 5	4
9	Tempo 6	4
9	Tempo 7	4
9	Tempo 2	4
9	Tempo 3	4
99	Tempo 3	4
85	Tempo 9	4
85	Tempo 10	4
85	Tempo 8	4
85	Tempo 11	4
85	Tempo 4	4
85	Tempo 5	4
85	Tempo 6	4
85	Tempo 7	4
85	Tempo 1	4
85	Tempo 2	4
85	Tempo 3	4
8	Tempo 9	4
8	Tempo 10	4
8	Tempo 8	4
8	Tempo 4	4
8	Tempo 5	4
8	Tempo 6	4
8	Tempo 7	4
8	Tempo 1	4
8	Tempo 2	4
8	Tempo 3	4
88	Tempo 9	4
88	Tempo 11	4
88	Tempo 4	4
88	Tempo 5	4
88	Tempo 7	4
88	Tempo 1	4
88	Tempo 2	4
88	Tempo 3	4
3	Tempo 9	4
3	Tempo 10	4
3	Tempo 8	4
3	Tempo 11	4
3	Tempo 4	4
3	Tempo 5	4

Nodo	Livello	Comunità
3	Tempo 6	4
3	Tempo 7	4
3	Tempo 1	4
3	Tempo 2	4
3	Tempo 3	4
74	Tempo 7	4
72	Tempo 11	5
102	Tempo 11	5
78	Tempo 9	5
78	Tempo 8	5
78	Tempo 11	5
78	Tempo 6	5
78	Tempo 7	5
76	Tempo 9	5
76	Tempo 10	5
76	Tempo 8	5
76	Tempo 11	5
76	Tempo 4	5
76	Tempo 5	5
76	Tempo 6	5
76	Tempo 7	5
76	Tempo 2	5
76	Tempo 3	5
7	Tempo 9	5
7	Tempo 4	5
7	Tempo 5	5
7	Tempo 1	5
7	Tempo 2	5
7	Tempo 3	5
89	Tempo 9	5
89	Tempo 4	5
89	Tempo 5	5
89	Tempo 1	5
89	Tempo 2	5
89	Tempo 3	5
109	Tempo 4	5
19	Tempo 10	5
19	Tempo 8	5
19	Tempo 5	5
19	Tempo 6	5
19	Tempo 7	5
59	Tempo 9	5
59	Tempo 8	5
59	Tempo 11	5
79	Tempo 9	5
79	Tempo 11	5
79	Tempo 7	5
61	Tempo 11	5
61	Tempo 7	5
101	Tempo 9	5
101	Tempo 11	5
71	Tempo 10	5
94	Tempo 11	6
92	Tempo 11	6
43	Tempo 11	6

Nodo	Livello	Comunità
87	Tempo 9	6
87	Tempo 10	6
87	Tempo 8	6
87	Tempo 11	6
87	Tempo 6	6
87	Tempo 7	6
73	Tempo 10	6
73	Tempo 8	6
38	Tempo 10	6
38	Tempo 8	6
39	Tempo 8	6
37	Tempo 9	6
37	Tempo 10	6
37	Tempo 8	6
37	Tempo 11	6
91	Tempo 8	6
81	Tempo 9	6
81	Tempo 10	6
81	Tempo 8	6
81	Tempo 11	6
81	Tempo 7	6
105	Tempo 9	6
105	Tempo 10	6
41	Tempo 9	6
41	Tempo 10	6
41	Tempo 11	6
95	Tempo 10	6
45	Tempo 10	6
21	Tempo 10	6
44	Tempo 10	6
27	Tempo 10	6
27	Tempo 11	6
42	Tempo 10	6
42	Tempo 11	6
93	Tempo 10	6
93	Tempo 11	6
40	Tempo 10	6

Tabella A.3: Comunità della rete multi-livello.