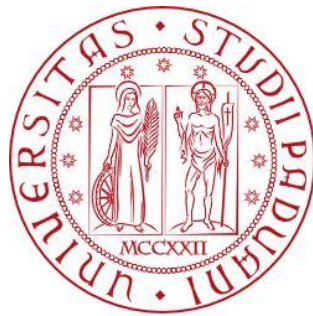


UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE
CORSO DI LAUREA MAGISTRALE IN
SCIENZE STATISTICHE



Modellazione di reti come strumento per lo studio delle relazioni internazionali

Relatore Prof. Emanuele Aliverti
Dipartimento di Scienze Statistiche

Laureando Federico Mirulla
Matricola N. 2037877

Anno Accademico 2023/2024

Indice

| | |
|---|-----------|
| Introduzione | 4 |
| 1 La geopolitica attraverso i dati | 6 |
| 1.1 Un'importanza crescente | 6 |
| 1.2 Sistemi di influenza | 8 |
| 1.3 Reti su conflitti e cooperazioni | 9 |
| 2 Dati e statistiche descrittive | 12 |
| 2.1 GDELT | 12 |
| 2.2 I dati | 13 |
| 2.2.1 La tassonomia "CAMEO" | 13 |
| 2.2.2 Le Nazioni analizzate | 17 |
| 2.2.3 Gli attributi | 19 |
| 2.3 Statistiche descrittive | 24 |
| 3 Modelli per reti sociali | 39 |
| 3.1 Le famiglie esponenziali | 39 |
| 3.2 Grafi Casuali a Famiglia Esponenziale, gli ERGM | 40 |
| 3.2.1 Un esempio di ERGM | 43 |
| 3.2.2 L'estensione multistrato | 44 |
| 3.2.3 Applicazione del modello ERGM | 47 |
| 3.3 Modelli a Effetti Additivi e Moltiplicativi | 56 |
| 3.3.1 Costruzione del modello AME | 57 |
| 3.3.2 Applicazione del modello AME | 60 |
| 3.4 Confronto dei modelli | 65 |
| Conclusioni | 68 |
| Appendice | 69 |
| Bibliografia | 76 |

Introduzione

Per decenni la geopolitica e lo studio delle interazioni tra i paesi non sono state materie considerate utili a comprendere i meccanismi che guidavano gli equilibri mondiali. Per mancanza di strumenti e di conoscenze adeguate, lo studio delle interazioni internazionali ha sempre avuto una prospettiva miope, senza la visione di un quadro complessivo che guidasse i processi decisionali delle istituzioni statali. Con l'avvento dell'età contemporanea si diede più peso a ciò che gli eventi e le interazioni tra gli stati potessero significare, con l'obiettivo di comprendere meglio il quadro generale della politica mondiale e poter prendere decisioni conseguenti. Al giorno d'oggi le possibilità di accedere a dati funzionali allo studio della geopolitica sono moltissime e, di conseguenza, la sfida è quella di sintetizzarli e ricavarne delle informazioni utili. I canali di informazione che divulgano quotidianamente una grandissima quantità di notizie costituiscono il principale veicolo per estrarre informazioni che, se immagazzinate e controllate, possono essere tradotte in dati. L'obiettivo dell'elaborato è misurare e descrivere, con l'uso di modelli statistici, i fattori che guidano le dinamiche e l'evoluzione delle interazioni tra i paesi e le loro influenze per mezzo dei dati estratti dai titoli dei quotidiani. La fonte dei dati è il progetto GDELT che monitora i principali canali di divulgazione internazionali con lo scopo di codificare e classificare le informazioni rilevanti all'interno dei titoli delle notizie riportate dai giornali. Il presente elaborato percorre, dalla genesi del dato alla modellazione, lo studio parallelo dei rapporti di conflitto e di cooperazione avvenuti all'interno di un insieme di cinquanta paesi attraverso due diversi modelli statistici: i Modelli per Grafi a Famiglia Esponenziale e i Modelli con Effetti Additivi e Moltiplicativi. I due modelli citati sono stati scelti in funzione dei loro diversi approcci: l'aspetto attrattivo dei Grafi a Famiglia Esponenziale, detti anche ERGM, consiste nella loro facilità di interpretazione e nella possibilità di valutare schemi globali, semplicemente includendo un numero di statistiche caratterizzanti la rete o i soggetti, molto inferiore al numero stesso dei soggetti. Al contrario, i Modelli a Effetti Additivi e Moltiplicativi, detti AME, permettono di misurare schemi locali, a livello micro, delle relazioni

tra nodi specifici e di catturare informazioni non osservate tramite l'utilizzo di variabili latenti.

La struttura della tesi è definita come segue.

Nel capitolo 1 viene presentato il cambiamento di concezione che la geopolitica ha subito dal '900 fino ad oggi. Si descrive l'evoluzione della percezione degli stati come entità a sé stanti fino alla valutazione di un insieme di forze sinergiche, l'una dipendente dall'altra. Vengono introdotte le strutture di rete che rappresentano la nuova concezione di relazioni internazionali e costituiscono le strutture portanti dalle quali si sviluppano le analisi.

Nel capitolo 2 si presenta la fonte dei dati, GDELT e si descrive la scelta del campione di paesi e l'assegnazione delle variabili che li caratterizzano. Vengono poi condotte le analisi esplorative sui dati e sulle strutture di rete, in particolare sul numero di legami presenti all'interno di ogni rete.

Nel capitolo 3 si introducono e applicano i modelli per le strutture di rete. I Grafi Casuali a Famiglia Esponenziale vengono utilizzati per modellare le reti sociali attraverso un approccio che sfrutta le caratteristiche endogene dei paesi e delle reti come principale spiegazione della formazione dei legami. Mentre i Modelli a Effetti Additivi e Moltiplicativi vengono implementati per studiare le reti attraverso un approccio a variabili latenti e valutare la formazione di legami attraverso fattori esogeni.

Capitolo 1

La geopolitica attraverso i dati

1.1 Un'importanza crescente

Durante il XIX secolo, sebbene la geopolitica subisse un diffuso discredito, la società assistette a una lunga tradizione di “sovranità” nazionale. Gli stati erano concepiti come oggetti monadici e in quest’ottica veniva valutata la politica mondiale, attribuendo peso alle nazioni in funzione del loro potere, visto nella capacità del singolo di influenzare l’equilibrio della geopolitica globale. Il presupposto principale di questo assetto era sovradeterminato dalle azioni delle organizzazioni che si affermavano come “sovrane”, lasciando pochissimo spazio alle interazioni a meno che queste non fossero direttamente coinvolte nel processo di cambiamento del potere delle singole nazioni. Fu proprio nel XIX secolo, grazie a politologi e geografi come R.Kjellén, K.Haushofer e H.Mackinder (Hoff e Ward, 2003), che la concezione predominante degli studiosi fu quella di considerare gli stati come entità a sé stanti che cercavano costantemente di espandersi per assicurarsi risorse e territori; teoria che trovò il suo apice nel concetto di *lebensraum* o “spazio vitale”, teorizzato per la prima volta da Friedrich Ratzel, geografo e biologo tedesco, nel suo lavoro del 1901 intitolato “Politische Geographie”.

Nei primi del '900 infatti, il *lebensraum* costituì l’idea centrale della geopolitica e, successivamente, del nazionalsocialismo, secondo cui alcuni popoli avrebbero avuto una sorta di “diritto naturale” ad espandersi sui territori limitrofi (Enciclopedia Treccani, 2022). Fu tra la fine del XIX e l’inizio del XX secolo che l’idea di considerare gli stati come entità isolate e indipendenti lasciò spazio a una nuova concezione di nazioni, intese via via sempre più come un insieme di forze sinergiche l’una interdependente dall’altra. Il 1900 fu infatti un secolo di grandi sconvolgimenti del sapere umano in diverse discipline, come la fisica, la biologia, la sociologia e l’economia: in

campo fisico, ad esempio, il concetto di interdipendenza e interazione tra le parti di un sistema emerse con l'avvento della meccanica statica e della termodinamica statica, dove si iniziò a considerare i sistemi come aggregati di molte particelle in interazione anziché come entità isolate. In biologia, il concetto di interconnessione e interdipendenza tra le parti di un organismo fu sviluppato attraverso teorie evoluzionistiche come la teoria dell'evoluzione di Darwin secondo cui le interazioni tra organismi complessi sono cruciali per la sopravvivenza e l'evoluzione. Infine, nel campo delle scienze sociali, le società iniziarono a essere considerate come sistemi complessi fatti di interazioni tra individui, istituzioni e fattori culturali ed economici. L'insieme di questi elementi, diede un impulso significativo anche alla concezione di "stato" inserito nel panorama mondiale.

Oggi sappiamo con certezza quanto alcuni fattori stabiliscono una forte connessione tra i paesi. Ad esempio, in campo economico, gli stati sono interconnessi attraverso il commercio internazionale, gli investimenti e le catene di approvvigionamento globali; o la sicurezza: la proliferazione delle armi e i conflitti regionali possono avere effetti trasversali che coinvolgono più stati sia in campo conflittuale che cooperativo. L'ambiente: i cambiamenti climatici, l'inquinamento e la perdita di biodiversità sono problemi globali che richiedono una risposta coordinata tra gli stati. La Politica: le relazioni diplomatiche e le alleanze influenzano il clima politico globale e possono essere impattanti per la pace e la sicurezza interazionale e i diritti umani. Ancora, la tecnologia: lo scambio di conoscenze, tecnologie e informazioni tra gli stati può favorire l'innovazione e lo sviluppo. La migrazione: i movimenti di persone attraverso le frontiere nazionali possono avere impatti economici, sociali e politici su entrambi i paesi di origine e di destinazione. Ignorare l'interdipendenza tra questi fenomeni sarebbe una grave omissione che affliggerebbe i tentativi di comprendere, e ancor di più di prevedere, il corso della politica nazionale e mondiale in generale (Hoff e Ward, 2003).

Oggi la geopolitica è al centro delle decisioni prese da istituzioni di ogni tipo, quali ad esempio imprenditori, partecipanti al mercato e funzionari delle banche centrali, che considerano i rischi geopolitici come determinanti chiave delle decisioni di investimento e delle dinamiche del mercato azionario. Istituti come la Banca d'Inghilterra, la Banca Centrale Europea e il Fondo Monetario Internazionale monitorano regolarmente i rischi per le prospettive economiche posti dalle tensioni geopolitiche. Secondo un sondaggio del 2017 condotto da Gallup, su oltre mille investitori, il 75 per cento degli intervistati ha espresso preoccupazioni riguardo l'impatto economico dei vari conflitti militari e diplomatici in corso nel mondo (Caldara e Iacoviello, 2022). Tuttavia, fattori sociali e politici sono stati di rado oggetto di analisi empiriche siste-

matiche da parte della comunità scientifica. Le principali limitazioni sono la mancanza di indici e lo scarso utilizzo di modelli che misurino in tempo reale le tensioni geopolitiche percepite dalla stampa, dal pubblico, dagli investitori globali e dai responsabili delle politiche.

1.2 Sistemi di influenza

Nonostante il crescente interesse verso le politiche internazionali, misurare il potere e l'influenza degli stati, che in passato erano meglio identificabili attraverso il potere militare o quello economico, è diventato più complesso. Oggi la realtà dei rapporti tra stati è sfaccettata e non c'è più lo stesso consenso a livello concettuale su cosa significhino esattamente potere e influenza. È per questo motivo che la tesi si propone di rispondere all'esigenza di analizzare i sistemi di influenza tra gli attori mondiali con metodi più articolati, che spostano la percezione dei concetti di potere e influenza verso prospettive relazionali piuttosto che materiali. Se in passato lo schema degli equilibri mondiali poteva essere pensato come piramidale, con gli stati più influenti al vertice della piramide e quelli meno potenti verso il basso, oggi la comunità scientifica fa riferimento a schemi di rete, immaginabili come vere e proprie reti in cui gli stati costituiscono i nodi e le relazioni instaurate formano le connessioni tra di essi. È in questo contesto che lo studio condotto prende luogo, proponendo degli strumenti che, dall'osservazione delle reti esistenti, misurano l'importanza di caratteristiche atte ad aumentare o diminuire la probabilità di formazione di un legame tra due stati e identificano comunità di soggetti simili nel modo di operare all'interno del panorama mondiale.

Le reti sono strutture costruite a partire da dati relazionali ovvero dati che definiscono le connessioni tra coppie di soggetti, note anche come diadi. Queste connessioni possono assumere la forma di semplici relazioni binarie e non direzionali oppure rappresentare tipi più complessi di relazioni come quelle direzionali o pesate. L'utilizzo delle strutture di rete nei lavori accademici è molto frequente, dalla genetica dove i ricercatori identificano i nodi come proteine e i collegamenti come le interazioni proteiche, fino allo studio delle reti terroristiche. I dati relazionali, a causa della loro natura di legami sociali, non possono essere studiati come un insieme di osservazioni indipendenti, ma piuttosto come una rete in cui il collegamento tra una coppia di soggetti influenza ed è influenzato da altre diadi. Grazie ai modelli di rete è possibile tener conto della dipendenza intrinseca nei dati relazionali ed estrarre informazione dai percorsi che danno forma alla struttura globale in cui i soggetti sono inseriti (Minhas et al., 2017). Questo tipo di approccio è stato reso possibile, nello studio delle relazioni tra stati, dai dati relazionali raccolti

da GDELT (Global Database of Events, Language, and Tone) che a partire dai titoli dei giornali più letti estrae informazioni sui legami che collegano i paesi di tutto il mondo, suddividendo inoltre le interazioni in “positive” e “negative” ovvero in cooperazioni e conflitti (The GDELT Project, 2013).

1.3 Reti su conflitti e cooperazioni

A titolo illustrativo, si mostra una porzione di dati che vengono analizzati. Si rimanda al capitolo 2 per una descrizione esaustiva di come sia stato scelto il campione di paesi utilizzato e una visione complessiva delle reti considerate.

Le reti sociali sono uno strumento utile e diffuso in ambito statistico, sia perché possono essere modellate, sia perché sono strutture facilmente rappresentabili con un impatto visivo significativo e interpretabile. Nel presente studio le analisi si sono sviluppate su una totalità di 24 reti, 12 per i conflitti e 12 per le cooperazioni, ognuna riferita a un mese specifico del 2023. A scopo rappresentativo, di seguito si riportano gli esempi delle due reti riferite al mese di luglio. Sono state rappresentate le reti rendendo i vertici di grandezza proporzionale alla loro grado di centralità totale, cioè una misura di conteggio degli archi in uscita e in entrata nel nodo:

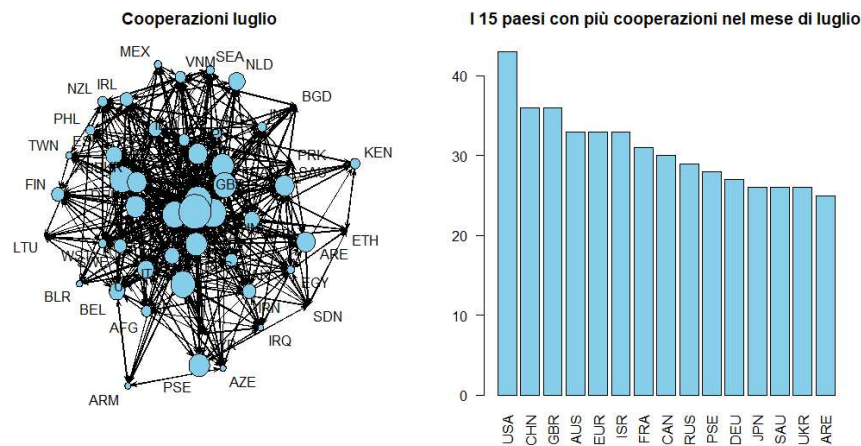


Figura 1.1: Rete delle cooperazioni di luglio 2023

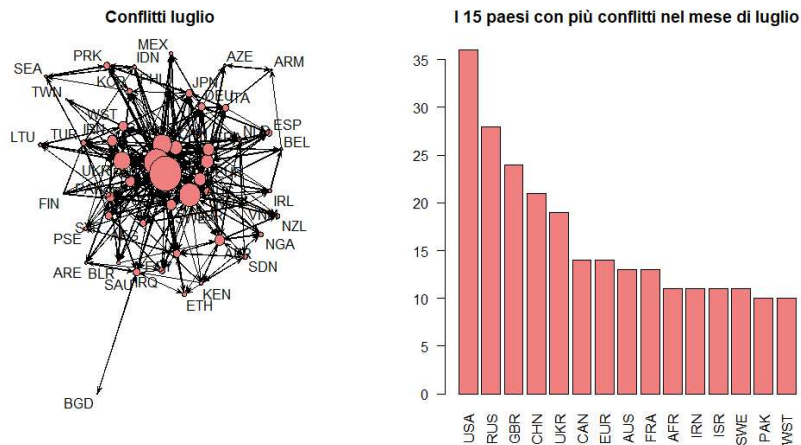


Figura 1.2: Rete dei conflitti di luglio 2023

Le figure riportate hanno una densità di legami che impedisce di identificare il nome di tutti gli stati coinvolti, specialmente quelli centrali. Per questo vengono riportati anche gli istogrammi relativi ai primi quindici paesi per quota di centralità decrescente. Nel caso delle cooperazioni di luglio 2023, in figura 1.1, è interessante notare come tutti gli attori siano coinvolti nel processo di collaborazione tra stati, il grafico infatti è approssimativamente circolare senza soggetti particolarmente limitrofi tranne per Armenia e Azerbaijan per i quali sembrano esserci principalmente archi entranti e uno bidirezionale tra i due. In generale, nella parte periferica del grafico si registrano paesi tra cui Bangladesh, Ethiopia, Kenya, la zona del sud-est asiatico, l'Irlanda, la Lituania e altri attori sicuramente con pesi nello scenario mondiale e in questo specifico caso delle cooperazioni, minore. Al centro del grafico, caratterizzati da punti decisamente più ampi, ci sono i cosiddetti "hub centrali" ossia nodi, con un numero elevato di connessioni rispetto a tutti gli altri. Gli hub centrali sono importanti poiché giocano un ruolo chiave nella struttura e nel funzionamento della rete, sono attori che occupano posizioni strategiche che permettono loro di collegare diverse parti della rete potendoli interpretare anche come punti di snodo o nodi di interconnessione (Kirkley, 2024). Hanno grande influenza e controllo sulle dinamiche di rete esercitando un ruolo significativo tra gli altri attori. Di norma gli hub sono un numero ridotto di soggetti sulla totalità presente, e in questo caso, la figura 1.1 riporta l'istogramma dei primi quindici stati per numero di legami totali. In prima posizione, con 47 legami ci sono, come ci si poteva aspettare, gli Stati Uniti, seguiti dalla Gran Bretagna con 39 legami, la Russia con 37, l'Europa con 33 e la Cina con 32 legami. Rispetto agli altri paesi, quelli ora

citati hanno sicuramente un peso significativamente più grande nella bilancia mondiale delle entità cooperanti.

I grafici in figura 1.2 rappresentano invece i rapporti conflittuali. Oltre a ritrovare alcuni dei paesi che nel grafico in figura 1.1 occupavano le stesse zone marginali vi sono inoltre, tra gli altri, l'Italia, il Sudan, gli Emirati Arabi Uniti, le Filippine, il Belgio e il Bangladesh, che in questo caso sembra avere un ruolo praticamente escluso dalle dinamiche conflittuali mondiali. La composizione degli attori centrali invece è meno varia e in gran parte si ritrovano gli stessi hub dello schema sulle cooperazioni. Il primo posto è mantenuto dagli Stati Uniti con 36 rapporti conflittuali, succeduti dalla Russia con 28, vengono poi la Gran Bretagna e la Cina con 24 e 21 legami rispettivamente. L'Ucraina si profila al quinto posto, con 19 contatti conflittuali, molto probabilmente a causa della guerra in atto dal 2022 con la Russia.

Le osservazioni riportate evidenziano come le reti sociali, anche a livello grafico, siano strumenti utili a sintetizzare e trarre informazioni dalla complessità dei rapporti tra gli stati. Infatti, la loro immediata lettura rende accessibile la conoscenza delle dinamiche in atto del sistema oggetto di studio. Le reti sono strutture essenziali a identificare i nodi più influenti nel sistema e quindi a ipotizzare gerarchie di potere ed influenza. Le azioni di un attore internazionale influenzano e sono influenzate da molti altri attori. Le reti permettono di modellare e analizzare queste interdipendenze.

Capitolo 2

Dati e statistiche descrittive

2.1 GDELT

Lo studio condotto nasce da uno dei più importanti canali da cui si possono raccogliere i dati necessari alla costruzione delle strutture di rete: le testate giornalistiche. Attraverso i notiziari di tutto il mondo è ragionevole che i dati considerati garantiscano un'ampia copertura a livello globale, una precisa collocazione temporale e una varietà di fonti. GDELT è un progetto che mira a raccogliere, monitorare e analizzare notizie e informazioni provenienti dai notiziari di tutto il mondo. Si tratta di un enorme archivio di eventi che copre una vasta gamma di argomenti, tra cui politica, conflitti, disastri, economia e ambiente. La banca dati contiene circa un quarto di miliardo di eventi geolocalizzati, con copertura globale, dal 1979 ad oggi e con aggiornamento delle informazioni giornaliero. I dati si basano su articoli giornalistici provenienti da una varietà di fonti internazionali di notizie: vengono esaminati e identificati gli eventi inclusi in testate come BBC, United Press, Washington Post e molte altre. Ulteriori fonti includono raccolte di articoli nazionali e internazionali del New York Times, dell'Associated Press e le notizie nazionali e internazionali da Google News, ad eccezione di sport, intrattenimento e notizie strettamente economiche.

Dalla grande quantità di materiale che affluisce a GDELT, vengono immagazzinate informazioni relative agli eventi globali, come proteste, conflitti, accordi diplomatici, elezioni, catastrofi naturali, discorsi politici, cambiamenti di governo e altri eventi di rilevanza internazionale. Altre sezioni riportano dati sulle lingue utilizzate dai media per tracciare gli eventi e la conseguente diffusione geografica. Parte fondamentale del lavoro di GDELT è dedicato a catalogare “il tono” delle notizie riportate, quindi un'analisi sulle tonalità positive, negative o neutre che un articolo può avere nel descrivere una no-

tizia che può essere di aiuto nel prevedere cambiamenti sociali e politici o per esaminare come i diversi media trattano le stesse notizie (The GDELT Project, 2022).

2.2 I dati

I dati relazionali prelevati da GDELT sono semplici e intuitivi: si tratta essenzialmente della composizione di tre tipologie di informazioni, due delle quali identificano i soggetti coinvolti e la terza che definisce l'esistenza di un legame o meno. Di tutta la banca dati disponibile sono stati utilizzati dodici database, immagazzinati nella sezione "GDELT 1.0 Event database". Fanno riferimento rispettivamente ai dodici mesi del 2023, ognuno corrispondente al giorno 15 di ogni mese in modo da sovrapporre il meno possibile notizie che vengono riproposte nel tempo, e contemporaneamente coprire una fascia temporale consistente che potesse apportare maggiori informazioni. Un dataset tipo contiene molteplici informazioni, divise in tre macro aree di semantica: variabili che localizzano temporalmente l'evento, variabili che identificano i due attori coinvolti e l'azione compiuta con relativa classificazione di intensità e le variabili che localizzano geograficamente i soggetti e l'evento accaduto (The GDELT Project, 2013). Per ogni attore vi sono diverse variabili afferenti per esempio al gruppo etnico di appartenenza, al nome dell'istituzione specifica che compie l'azione e alla religione di appartenenza, ma di fatto queste variabili, per quanto interessanti, presentano più valori mancanti che esistenti e perciò sono state eliminate dai dataset. Per far fronte a questa mancanza sono state inserite manualmente, in una fase successiva, variabili demografiche caratterizzanti i paesi, che andranno a costituire gli attributi dei soggetti.

Dei dodici dataset di partenza sono state mantenute quindi le due variabili riguardanti i paesi e la variabile di codifica dell'evento che segnala se gli stati coinvolti hanno avuto una relazione di conflitto o di cooperazione. Ogni coppia di paesi che compare nei dataset costituisce una diade ovvero la componente elementare per la costruzione delle strutture di rete.

2.2.1 La tassonomia "CAMEO"

La terza variabile che compone i dataset costruiti è relativa alla classificazione dell'evento che il primo paese compie sul secondo. Il metodo di classificazione si basa su una struttura di codifica degli eventi chiamata "CAMEO", acronimo di "Conflict and Mediation Event Observations". È un sistema di codifi-

ca standardizzato e ampiamente utilizzato per categorizzare e classificare gli eventi di conflitto, mediazione e cooperazione nelle relazioni internazionali. Per diverse decadi, altre due codifiche dominarono la ricerca dei dati sugli eventi: il WEIS di Charles McClelland e il COPDAB, “Conflict and Peace Data Bank” sviluppato da Edward Azar (Schrodt, 2012). Entrambi creati durante la Guerra Fredda, schematizzavano un mondo politico in cui gli stati sovrani reagivano l’uno all’altro principalmente attraverso la diplomazia ufficiale e le minacce militari. Sebbene innovativi ai tempi, questi sistemi di codifica non sono ottimali per affrontare questioni contemporanee come i conflitti etnici, la violenza a bassa intensità, l’attività criminale organizzata e l’intervento multilaterale. CAMEO era originariamente destinato a supportare un progetto per lo studio della mediazione nei conflitti interstato, invece si è sviluppato come uno schema di codifica di nuova generazione progettato sia per correggere alcuni dei problemi di WEIS e COPDAB, ma soprattutto per la codifica automatizzata degli eventi. La versione odierna di CAMEO è strutturata in venti macro categorie. All’interno dei dataset le venti categorie sono specificate in alcune variabili che aggregano gli eventi con diverse risoluzioni di specificità di modo che i titoli dei notiziari che affluiscono a GDELT possano essere ben classificati (Schrodt, 2012).

Nella costruzione dei dataset è stata mantenuta solo una variabile che permette una classificazione ad un più alto livello di aggregazione, organizzando in ultima analisi l’intera tassonomia degli eventi CAMEO in quattro categorie principali: Cooperazione Verbale, Cooperazione materiale, Conflitto Verbale e Conflitto Materiale. La struttura finale dei dataset ha previsto la suddivisione ulteriore di ognuno dei dodici dataset in due, separando i dati relativi ai conflitti da quelli sulle cooperazioni arrivando alla totalità di 24 dataset.

Nella tabella 2.1 viene riportata la macro suddivisione degli eventi che è stata utilizzata per condurre le analisi parallelamente sulle cooperazioni e sui conflitti.

| Cooperazioni | Conflitti |
|--|-------------------------------|
| Dichiarazioni pubbliche | Richieste |
| Appelli | Esprimere disapprovazione |
| Intenzioni di cooperare | Respinte |
| Consultazioni | Minacce |
| Impegno nella cooperazione diplomatica | Proteste |
| Impegno nella cooperazione materiale | Mostrare una postura militare |
| Fornire aiuto | Ridurre le relazioni |
| Concessioni (Concedere/facilitare) | Costrizioni |
| Investigazioni | Aggressioni |
| | Combattimenti |
| | Violenza di massa |

Tabella 2.1: Suddivisione delle casistiche in conflitti e cooperazioni

La suddivisione comprende insiemi di eventi per un ampia gamma di possibilità, sia per i conflitti che per le cooperazioni, per cui è utile riportare alcune casistiche di circostanze analizzate accompagnate dal relativo titolo della testata giornalistica.

Viene considerato come cooperazione:

- Fare commenti ottimistici, cioè esprimere ottimismo, sicurezza, fiducia. Ad esempio: *“Il presidente turco Turgut Özal ha dichiarato mercoledì di essere fiducioso che gli Stati Uniti rimuoveranno gli ostacoli che danneggiano le relazioni tra i due alleati NATO.”*
- Richiedere o esprimere l’intento di cooperazione in ambito materiale, economico, militare, giudiziario, di intelligence e diplomatico. Alcuni titoli di giornali possono essere: *“Venerdì, i leader aziendali indiani hanno chiesto un maggiore impulso verso il libero scambio nonostante le crescenti tensioni tra India e Pakistan.”* Oppure: *“La Turchia ha dichiarato lunedì di aver chiesto a Teheran e Damasco di fornire informazioni urgenti riguardo armi e munizioni sequestrate la scorsa settimana nel sud-est della Turchia a bordo di sei camion in viaggio dall’Iran alla Siria.”*. Oppure ancora: *“Un funzionario dell’ONU ha dichiarato lunedì che le parti turco-cipriota e greco-cipriota di quest’isola mediterranea divisa hanno concordato di condividere informazioni venerdì sulle persone scomparse di entrambe le comunità.”*
- Richieste di aiuti militari, economici o umanitari, ad esempio: *“Oggi Oxfam Canada ha lanciato un appello alla comunità internazionale*

affinché aiuti a salvare decine di migliaia di civili afgani minacciati dalla fame.”.

- Cooperazioni diplomatiche che possono essere: complimentarsi, difendere verbalmente, mobilitare il sostegno a nome di qualcuno, per esempio: *“Un alto ministro ha dichiarato in un commento pubblicato giovedì che l’Arabia Saudita ha mobilitato gruppi di pressione negli Stati Uniti per aiutare a sostenere i diritti dei palestinesi nella loro lotta contro Israele.”.*

Vengono invece classificati come conflitti eventi del tipo:

- Accuse di crimini, corruzione, abuso dei diritti umani, aggressione, crimini di guerra, spionaggio e minacce, Ad esempio: *“Il primo ministro del Kosovo è stato incriminato dal tribunale dell’ONU per i crimini di guerra per il suo presunto coinvolgimento in atrocità durante la lotta contro le forze serbe e si dimetterà.”.*
- Rifiutare cooperazioni materiali, economiche, militari, giudiziarie, di intelligence: *“La Jugoslavia ha categoricamente respinto giovedì un ultimatum australiano per consegnare una guardia coinvolta in una sparatoria di fronte al consolato jugoslavo a Sydney.”.*
- Ridurre o rompere relazioni in termini diplomatici, aiuti materiali, assistenza economica, militare o umanitaria. Per esempio: *“Il Giappone ha annunciato martedì di aver interrotto gli aiuti economici alla Jugoslavia, in linea con gli sforzi occidentali per porre fine ai combattimenti lì.”.*
- Assassini o tentati assassini: *“I guerriglieri di Hezbollah hanno ucciso il vice capo dell’alleato miliziano di Israele nel sud del Libano domenica, hanno dichiarato fonti da entrambe le parti.”.*
- Combattimenti, che spaziano da conflitti a bassa intensità a scontri con artiglieria e carri armati e uso di armi aeree: *“Aerei britannici, utilizzando missili guidati con precisione, hanno ucciso 4 iracheni in un attacco a un presunto deposito di armi a Basra.”.*
- Espulsioni di massa, uccisioni di massa e pulizie etniche: *“Le forze serbe erano coinvolte in pulizie etniche in Kosovo contro la maggioranza della popolazione albanese della provincia, secondo il governo degli Stati Uniti.”.*

- Uso di armi di distruzione di massa, armi chimiche, biologiche o nucleari.

Le casistiche elencate rappresentano solo alcuni esempi della classificazione che GDELT adopera sulle informazioni ricevute. Prima di ciò, ogni titolo di giornale viene esaminato nel contenuto attraverso analisi di linguaggio che permettono di estrapolare le informazioni dalla frase. Poniamo per esempio, il seguente titolo di giornale: *“L’Angola ha chiesto il supporto militare al Portogallo, in particolare istruttori per le sue forze armate...”*. In questo caso, attraverso ricerche testuali sulla stringa, GDELT individua automaticamente chi sono gli attori coinvolti e l’azione che un soggetto compie nei confronti dell’altro. Le informazioni utili del titolo che vengono prelevate sono: “Angola”, “ha chiesto il supporto militare” e “Portogallo”. Sono rispettivamente il primo soggetto, l’azione che compie e il secondo soggetto. Così procedendo GDELT incanala le informazioni che riceve nei suoi database scomponendo i titoli in parole o sotto stringhe che rappresentano, attraverso eventuali codifiche e ulteriori passaggi, le variabili.

2.2.2 Le Nazioni analizzate

Dai 24 dataset risultanti sono state eliminate le diadi di paesi che comparivano un numero di volte insufficiente ai fini della consistenza delle analisi. È stato stabilito un insieme di attori comune sulla base dei soggetti che hanno avuto una quota minima di interazioni per poter ottenere delle densità di rete adatte all’applicazione dei modelli. I paesi scelti, che comporranno i nodi delle reti, sono risultati essere:

| | |
|--------------------------|----------------------------|
| AFG: Afghanistan | ITA: Italia |
| AFR: Africa | JPN: Giappone |
| ARE: Emirati Arabi Uniti | KEN: Kenya |
| ARM: Armenia | KOR: Corea del Sud |
| AUS: Australia | LTU: Lituania |
| AZE: Azerbaijan | MEX: Messico |
| BEL: Belgio | NGA: Nigeria |
| BGD: Bangladesh | NLD: Paesi Bassi |
| BLR: Belarus | NZL: Nuova Zelanda |
| CAN: Canada | PAK: Pakistan |
| CHN: Cina | PHL: Filippine |
| DEU: Germania | PRK: Corea del Nord |
| EGY: Egitto | PSE: Palestina |
| ESP: Spagna | RUS: Russia |
| ETH: Etiopia | SAU: Arabia Saudita |
| EUR: Europa | SDN: Sudan |
| FIN: Finlandia | SEA: Sud Est Asiatico |
| FRA: Francia | SWE: Svezia |
| GBR: Regno Unito | SYR: Siria |
| IDN: Indonesia | TUR: Turchia |
| IND: India | TWN: Taiwan |
| IRL: Irlanda | UKR: Ucraina |
| IRN: Iran | USA: Stati Uniti d'America |
| IRQ: Iraq | VNM: Vietnam |
| ISR: Israele | WST: Emisfero Ovest |

Tabella 2.2: Lista dei 50 paesi analizzati

Di seguito, in figura 2.1, viene riportata una piantina del mondo che evidenzia i paesi in studio:

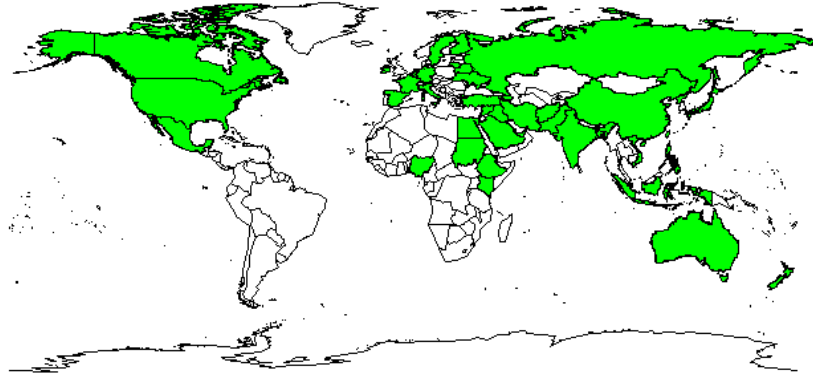


Figura 2.1: Mappa dei 50 paesi analizzati

Come riportato in tabella 2.2, la fase di selezione iniziale ha incluso 3 entità che non fanno riferimento a una singola nazione ma a regioni più grandi: l'emisfero Ovest, il Sud Est Asiatico e l'Africa. Sono state mantenute ugualmente al fine di avere degli attori che rappresentassero macro entità e le loro modalità di rapporto all'interno del panorama complessivo.

Stabilito il campione di diadi, formato dai paesi di interesse, le reti sono definite da nodi e legami che le costituiscono. Ogni paese è stato caratterizzato attraverso attributi specifici che riguardano diversi aspetti della sua composizione andando così a comporre manualmente parte delle variabili esplicative da inserire nei modelli.

2.2.3 Gli attributi

L'insieme di nodi che compone le reti rappresenta un gruppo eterogeneo di nazioni. Le caratteristiche che differenziano i paesi sono molte e con intensità differenti. Per modellare la probabilità di interazione tra i nodi all'interno delle reti sono stati scelti alcuni indici demografici e altri aspetti culturali ed identitari che costituiranno le covariate di ogni diade. Gli attributi assegnati agli attori sono:

- Tasso di fecondità totale: è un indicatore utilizzato in demografia, chiamato comunemente “numero medio di figli per donna”. È un impor-

tante segnale per comprendere la dinamica demografica di una popolazione e può influenzare diversi aspetti socio-economici e politici di una società. Il tasso di fecondità può variare notevolmente tra diverse regioni e paesi, a seconda di fattori come la cultura, l'accesso ai servizi sanitari, il livello di istruzione delle donne, le politiche pubbliche sulla famiglia, la transizione demografica ed altri elementi. Un tasso di fecondità al di sotto del livello di sostituzione, generalmente considerato intorno a 2.1 figli per donna, può portare a una diminuzione della popolazione nel lungo termine, mentre un tasso di fecondità superiore a questo livello comporta la crescita della popolazione.

- Prodotto Interno Lordo PIL è una misura economica che rappresenta il valore di tutti i beni e i servizi prodotti all'interno dei confini di un paese in un dato periodo di tempo. È un indicatore chiave per comprendere il comportamento di un paese sia dal punto di vista demografico che politico. Il benessere economico è una delle principali preoccupazioni della popolazione, per questo il PIL può essere utilizzato per valutare il livello generale di prosperità e sviluppo economico di un paese. Una popolazione con un PIL più alto tende generalmente ad avere un maggiore accesso a servizi essenziali, istruzione, assistenza sanitaria e altre risorse. In questo caso è stato considerato il tasso di crescita annuale del Prodotto Interno Lordo, cioè la variazione percentuale nel valore del PIL di un paese rispetto all'anno precedente. Sappiamo infatti che un'economia forte e in crescita tende ad essere associata a una maggiore stabilità politica. I governi che riescono a mantenere un PIL in crescita e un basso tasso di disoccupazione sono spesso visti come efficaci e possono godere di un maggiore sostegno pubblico. Dall'altra parte, una contrazione economica o un PIL stagnante possono portare a disordini sociali e instabilità politica.
- Indice di sviluppo umano, HDI è stato creato dal Programma delle Nazioni Unite per lo Sviluppo (UNDP) nel 1990. È stato introdotto per fornire una misura più completa dello sviluppo di un paese, andando oltre il semplice focus sul reddito o il PIL per includere anche indicatori di benessere sociale e qualità della vita delle persone. È una misura riassuntiva del raggiungimento medio nelle principali dimensioni dello sviluppo umano: una vita lunga e sana, essere istruiti ed avere un tenore di vita decente. La dimensione della salute è valutata dalla speranza di vita alla nascita, quella dell'istruzione è misurata dalla media degli anni di istruzione per gli adulti di 25 anni e più e dagli anni di istruzione previsti per i bambini in età scolare. La dimensione del tenore

di vita è misurata dal reddito nazionale lordo pro capite. L'HDI può esser utilizzato per mettere in discussione le scelte politiche nazionali, chiedendosi come due paesi con lo stesso livello di PIL pro capite possano ottenere risultati di sviluppo umano diversi, stimolando il dibattito sulle priorità delle politiche governative. L'HDI semplifica e cattura solo parte di ciò che comporta lo sviluppo umano poiché non informa sulle disuguaglianze, la povertà o la sicurezza umana (UNPD, 2024).

- Densità della popolazione: è un indicatore chiave per valutare e gestire vari aspetti legati alla distribuzione e alla crescita della popolazione in un'area geografica. Tramite la densità della popolazione si può stimare l'utilizzo delle risorse come acqua, suolo, cibo e infrastrutture. Anche la pianificazione urbana dipende dalla densità di popolazione in quanto fornisce informazioni preziose per la gestione del territorio e lo sviluppo delle città, oltre a influenzare la qualità della vita dei residenti. Ad esempio, aree con densità troppo elevate possono avere problemi come scarsa qualità dell'aria, carenza di spazi verdi e maggiore stress sociale. L'impatto ambientale può essere un'altra conseguenza dell'indice in quanto una densità elevata può aumentare il danno sull'ambiente, ad esempio attraverso l'inquinamento atmosferico, la deforestazione, il consumo di energia e le emissioni di gas serra. Infine, la densità della popolazione può influenzare l'accesso alle risorse e ai servizi essenziali come l'istruzione e l'assistenza sanitaria.
- Religione: la variabile "Religione" rappresenta la confessione più presente all'interno della nazione. Ha sette modalità corrispondenti alle religioni più diffuse al mondo: Buddismo, Cristianesimo, Induismo, Islam ed Ebraismo, la settima modalità rappresenta gli stati a maggioranza non religiosa. Questa variabile svolge un ruolo significativo come indicatore culturale e sociale di un paese: le credenze religiose possono influenzare la struttura familiare, i ruoli di genere e la riproduzione; ad esempio, in alcuni casi, è enfatizzato il ruolo tradizionale del padre come capofamiglia e la madre come figura principale nella cura dei figli, mentre in altri, possono essere promossi modelli familiari più egualitari. Oltre a quella familiare la religione influenza anche la struttura sociale e gerarchica all'interno di una comunità fino ad instaurare dinamiche di casta o di lignaggio, o viceversa, possono riflettere principi di uguaglianza e giustizia sociale. È un dato di fatto che le credenze religiose possono contribuire a definire i valori culturali e morali di una società, influenzando le norme sociali, l'etica e il comportamento individuale e

collettivo oltre che la partecipazione politica: i gruppi religiosi possono anche esercitare un'autorevolezza significativa sulle decisioni politiche e sulle politiche pubbliche.

- Sistema di governo: Il sistema di governo di uno stato ha certamente un impatto considerevole sul comportamento della nazione caratterizzandolo anche in termini diretti di belligeranza o cooperazione. Per esempio, monarchie assolute, giunte militari ma anche monarchie semi costituzionali si caratterizzano per la concentrazione del potere nelle mani dei vertici senza, in certi casi, alcun controllo costituzionale o limitazione da parte di un parlamento o di altre istituzioni. Questi sistemi di governo possono favorire comportamenti belligeranti in base ad interessi personali o visioni di potere. Altri sistemi come le repubbliche parlamentari o quelle presidenziali avranno sicuramente caratteri diversi da sistemi di accentramento di potere: saranno più inclini a favorire la cooperazione internazionale in quanto le decisioni sono prese attraverso il dibattito e il consenso parlamentare ed il governo o il presidente ne sono responsabili di fronte al popolo. Poiché questa variabile possiede delle categorie al cui interno sono raccolti paesi molto diversi tra loro, viene data una breve definizione di quelle che sono le forme di governo considerate (Wikipedia, 2024):

- **Governo provvisorio**: stati che hanno un sistema di governo in transizione o in tumulto. Questi regimi non hanno una base costituzionale. Esempi sono l'Afghanistan o la Libia.
- **Monarchia assoluta**: monarchie in cui l'esercizio del potere da parte del monarca non è vincolato da alcuna legge costituzionale sostanziale. Il monarca agisce sia come capo di stato che come capo di governo. È il caso per esempio dell'Arabia Saudita.
- **Giunta militare**: un comitato dei leader militari della nazione controlla il governo per la durata di uno stato di emergenza durante il quale le disposizioni costituzionali sono sospese. Questa forma caratterizza alcuni stati africani come il Mali e il Sudan.
- **Stato mono-partitico**: sistema in cui il potere politico è per legge concentrato all'interno di un unico partito politico le cui operazioni sono ampiamente fuse con la gerarchia governativa. La Cina è un esempio.
- **Monarchia semi costituzionale**: il primo ministro è l'esecutivo attivo della nazione, ma il monarca ha ancora considerevoli poteri politici che può utilizzare a propria discrezione. Stati a monarchia semi costituzionale sono il Marocco e gli Emirati Arabi Uniti.

- **Repubblica semi presidenziale:** è un sistema di governo in cui coesistono un presidente e un primo ministro, con poteri esecutivi divisi tra di loro. Il presidente è eletto direttamente dal popolo e ha poteri esecutivi significativi, mentre il primo ministro è responsabile davanti al parlamento. Questo sistema combina elementi di un sistema presidenziale e di uno parlamentare, con il presidente che detiene un ruolo di rilievo ma il governo che dipende anche dal sostegno parlamentare. Alcuni esempi sono l'Egitto, il Portogallo e lo Sri Lanka.
- **Repubblica presidenziale:** il presidente è il capo del governo, è eletto e rimane in carica indipendentemente dal parlamento. Generalmente non esiste un primo ministro, e, se presente, nella maggior parte dei casi serve puramente a discrezione del presidente. Sono repubbliche presidenziali l'Indonesia e gli Stati Uniti.
- **Monarchia costituzionale:** sistema in cui il capo di stato è un monarca costituzionale che normalmente esercita i suoi poteri solo con il consenso del governo, del popolo e/o dei loro rappresentanti e la sua autorità è stabilita e limitata dalla legge costituzionale. Sono esempi il Regno Unito e il Giappone.
- **Repubblica parlamentare:** il capo del governo è selezionato dal parlamento ed è anche responsabile nei suoi confronti. Il capo dello stato è solitamente chiamato presidente ed è separato dal capo del governo, svolgendo un ruolo ampiamente apolitico e cerimoniale. In questi sistemi, il capo del governo è solitamente chiamato primo ministro, cancelliere o premier. Sono repubbliche parlamentare ad esempio l'India e l'Iraq.

Ognuno dei cinquanta nodi delle reti è stato caratterizzato dalle sei variabili socio demografiche presentate. Nella prossima sezione le analisi esplorative illustrano come le modalità delle variabili esplicative siano distribuite tra i soggetti e come si possono estrapolare informazioni utili dalle strutture di rete.

2.3 Statistiche descrittive

Come primo passo, sono state condotte le analisi descrittive sulle variabili sopra elencate, e a seguire vengono riportate ulteriori analisi descrittive sulle strutture di rete. In questo paragrafo si riportano alcune osservazioni sul campione che ne dimostrano l'eterogeneità in termini assoluti di composizione, e la presenza di pattern all'interno di specifici gruppi di paesi.

In primo luogo si riportano, nella tabella 2.3 e in figura 2.2, gli indicatori essenziali delle distribuzioni delle variabili quantitative insieme ai relativi boxplot:

| | Min. | 1° Qu. | Mediana | Media | 3° Qu. | Max. |
|-----------|---------|--------|---------|---------|---------|----------|
| HDI | 0.462 | 0.711 | 0.795 | 0.786 | 0.925 | 0.952 |
| GDP | -18.300 | 0.025 | 1.950 | 1.504 | 4.125 | 8.700 |
| Fecondità | 0.883 | 1.515 | 1.673 | 2.115 | 2.594 | 5.063 |
| Densità | 3.442 | 50.754 | 121.709 | 205.653 | 270.720 | 1328.680 |

Tabella 2.3: Statistiche descrittive dei variabili

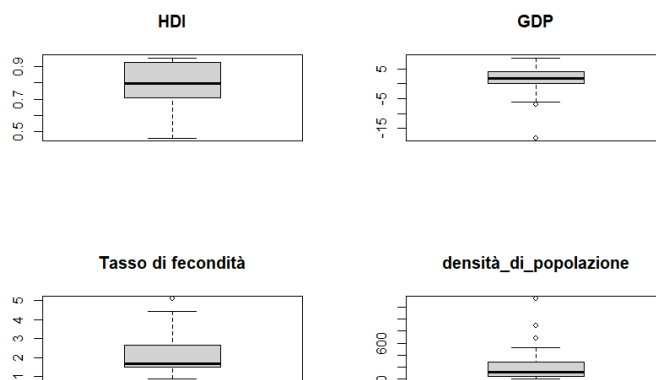


Figura 2.2: Box-plot delle variabili numeriche

Religioni:

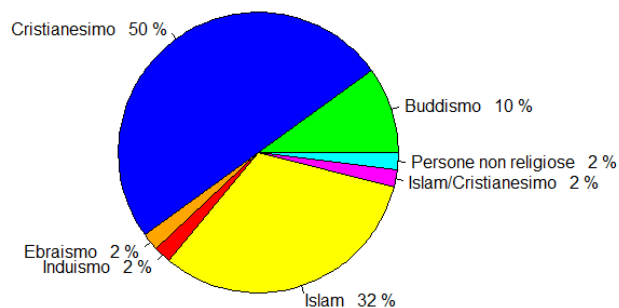


Figura 2.3: Religioni in percentuale

Sistemi di governo:

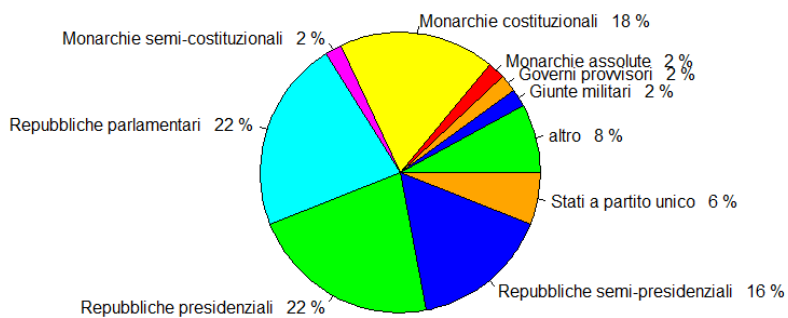


Figura 2.4: Sistemi di governo in percentuale

L'intero campione rappresenta una realtà eterogenea: tutte e quattro le variabili spaziano entro realtà differenti di cui fanno parte paesi di tutte

le fasce di sviluppo umano ed economico. Dalle prime statistiche, sull'indice di sviluppo umano e il tasso di crescita economica si riconosce che sono presenti paesi sotto sviluppati, in via di sviluppo e sviluppati. I quantili sulla densità e sul tasso di fecondità invece testimoniano una diversificazione di progresso nella transizione demografica, concetto che descrive il cambiamento di una società da un regime demografico caratterizzato da alti tassi di mortalità e natalità a uno caratterizzato da bassi tassi di mortalità e natalità.

I grafici a torta in figura 2.3 e 2.4 sono riferiti alle due variabili qualitative: l'appartenenza religiosa e il sistema di governo vigente. Il grafico sulle religioni rende evidente quanto sia grande la disparità numerica nella diffusione del cristianesimo e dell'islam rispetto alle altre confessioni, anche se rimane considerevole il 10% del buddismo principalmente localizzato nella fascia asiatica del globo. Anche il grafico sulle forme di governo racconta una realtà sfaccettata: le modalità più presenti nel campione sono le repubbliche presidenziali di cui fanno parte per esempio Siria, Turchia, Kenya, Stati Uniti d'America, e le repubbliche parlamentari come Irlanda, India, Pakistan e Italia, entrambe con il 22%. Con il 18% le monarchie costituzionali di cui fanno parte la Nuova Zelanda, la Norvegia, la Spagna e la Svezia e le repubbliche semi presidenziali come l'Egitto, la Francia, l'Azerbaijan e la Russia, con il 16%.

In figura 2.5 è riportato il grafico a barre di alcuni dei paesi del campione rappresentante l'indice di sviluppo umano. L'indice tocca il minimo di 0.462 e non va oltre lo 0.7 per paesi come Afghanistan, Ethiopia, Sudan, Pakistan, Kenya e India. Nei casi peggiori questo significa che l'aspettativa di vita è bassa, con uno scarso accesso ai servizi di base come istruzione e assistenza sanitaria oltre che a un tenore di vita povero. Diametralmente opposti si trovano i paesi occidentali, e altri come gli Emirati Arabi Uniti, la Corea del sud e il Giappone. In questi casi l'HDI non scende sotto lo 0.9 a significare che viene assicurata una vita lunga e sana, un accesso più ampio all'istruzione di qualità e un reddito pro capite significativamente più elevato rispetto ai paesi con HDI più basso.

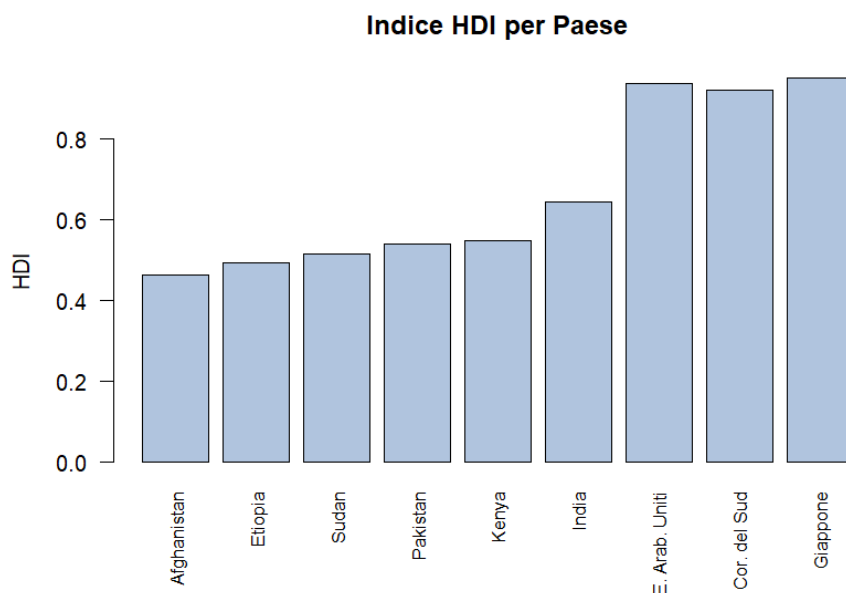


Figura 2.5: Indice di sviluppo umano

La figura 2.6 mostra il tasso di crescita annuale del PIL. Anche in questo caso tra i paesi si registra una variabilità molto ampia che riflette la diversità di sviluppo economico che può investire un paese in un anno. Nel grafico a barre vi sono pressoché tutti gli scenari possibili di sviluppo economico: paesi come Siria, Palestina e Afghanistan hanno subito una forte recessione economica principalmente dovuta a pesanti conflitti e una forte instabilità politica registrando tassi di decrescita di -7 circa per tutti e tre. I paesi considerati sviluppati, come quelli europei, il Canada e l’Inghilterra, mostrando un andamento economico quasi stabile o con tassi di crescita lievemente positivi. Infine ci sono paesi come India, Bangladesh e il sud-est asiatico e altre economie che hanno visto un’esplosione del loro fatturato negli ultimi decenni, paesi anche detti “in via di sviluppo”.

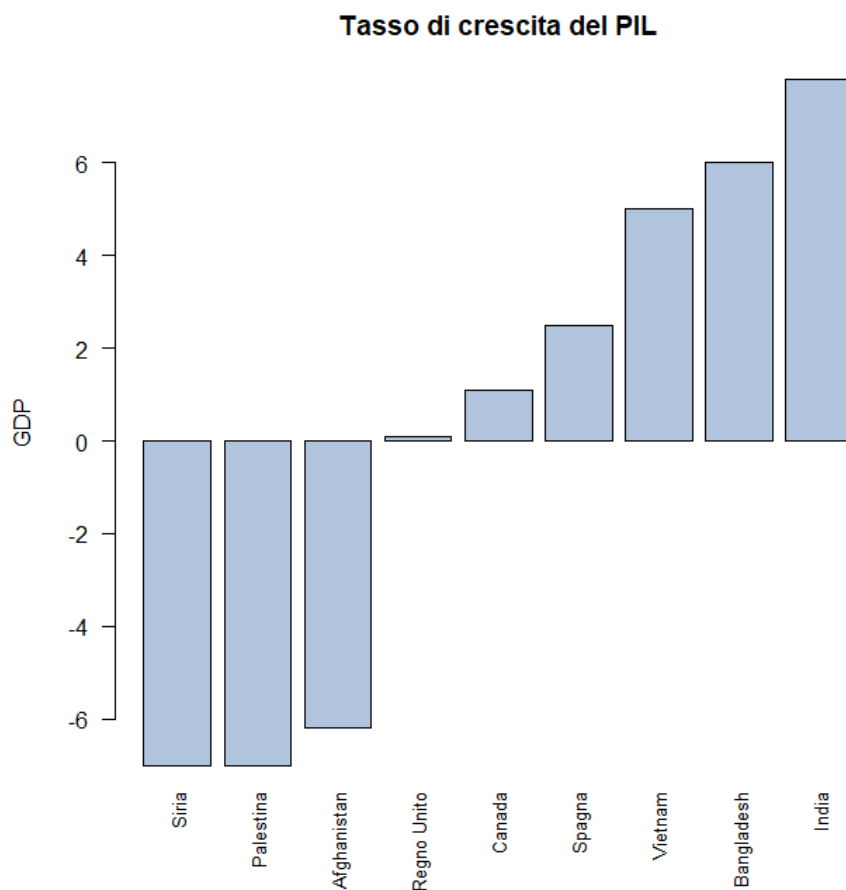


Figura 2.6: Tasso di crescita annuale del PIL

Il grafico in Figura 2.7 mostra l'andamento dei due indici precedenti congruentemente di alcuni paesi estratti dal campione. Tra i due indici non sussiste una relazione lineare, perciò ad un GDP elevato non corrisponde necessariamente un HDI elevato, e viceversa. Il gruppo di osservazioni che si accumula nella parte alta del grafico rappresenta i paesi con un'economia che non ha subito incrementi significativi con gli indici di sviluppo umano più alti, si tratta tendenzialmente dei paesi sviluppati con economie stabili. Quelli invece con economie in forte espansione sono spesso caratterizzati da uno sviluppo umano mediocre su cui concorrono diversi fattori negativi tra cui la dilagante povertà nei ceti sociali bassi e una forbice delle disuguaglianze sempre più ampia. Sono emblematici gli esempi dell'India, le Filippine o l'Armenia. Solo i soggetti con entrambi gli indici bassi o negativi riportano una

corrispondenza tra i due: si tratta di paesi fortemente compromessi da conflitti e disordini sociali che non permettono né lo sviluppo economico né tanto meno quello umano, è il caso di Afghanistan, Siria, Palestina e Sudan.

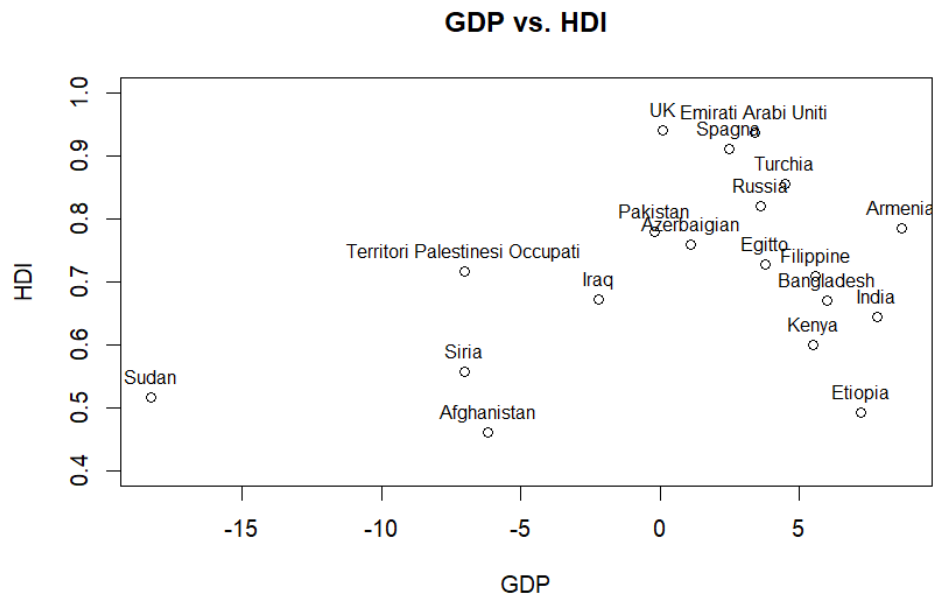


Figura 2.7: GDP vs. HDI

La figura 2.8 confronta la densità della popolazione con il tasso di fecondità. A primo impatto si potrebbe ipotizzare una relazione esponenziale negativa tra le due variabili per cui a quote alte di densità della popolazione corrisponde un tasso di fecondità basso, se non al di sotto del livello di ricambio generazionale (2.1). È il caso del Bangladesh, considerabile quasi come un outlier, ma anche Taiwan, Corea del Sud e India: paesi che raccontano un passato recente di incremento delle nascite seguito dal progresso economico e sociale che ha permesso il miglioramento dell'istruzione, un maggiore accesso ai servizi sanitari e cambiamento culturale. Il raggruppamento vicino all'origine del grafico è composto da tutti i paesi che per propria storia hanno già affrontato la fase di forte incremento di nascite, tendenzialmente durante il secolo scorso, e che ora stanno subendo il conseguenziale invecchiamento della popolazione che come effetto porta alla diminuzione della densità di popolazione. Sono emblematici gli esempi di Italia e Giappone, ma

anche Francia e Germania, conosciuti per la popolazione a maggioranza ultra sessantenne e il basso tasso di fecondità (1.3 circa per i primi due e 1.6 circa per i secondi due), che non permette il ricambio generazionale. Nella sezione superiore del grafico vi sono paesi africani e mediorientali con una densità della popolazione moderata e in crescita che stanno affrontando ora il picco di nascite, in attesa che anche per loro lo sviluppo umano ed economico comporti una riproduzione più moderata. Fanno parte di questo sottogruppo paesi africani come la Nigeria o l’Etiopia con rispettivamente cinque e quattro figli per donna in media e paesi mediorientali tra cui Afghanistan e Iraq con 4.4 e 3.4 figli medi per donna.

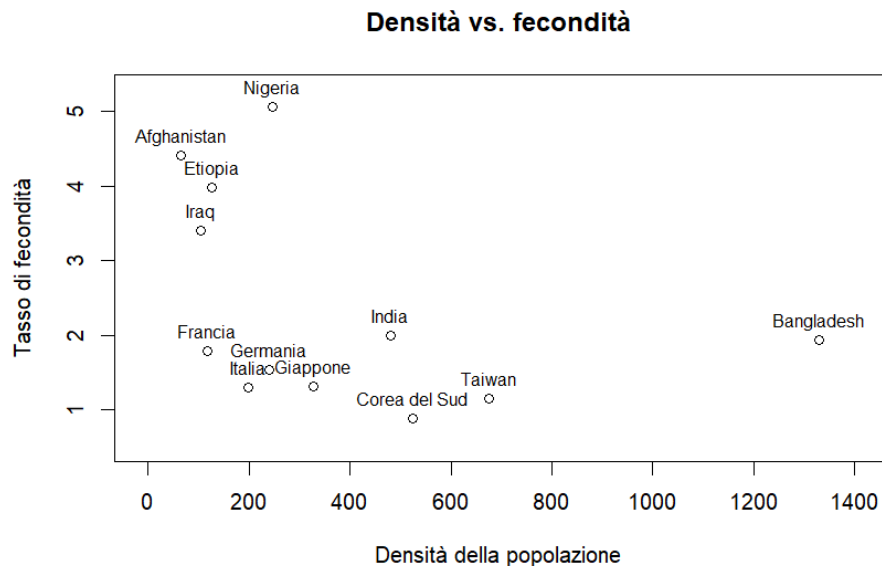


Figura 2.8: Densità vs. fecondità.

Le ultime due variabili categoriali che caratterizzano i paesi sono la confessione religiosa e la forma di governo.

Il grafico in Figura 2.9 mostra la distribuzione delle confessioni religiose congiuntamente al tasso di fecondità per ogni paese: partendo dal basso è interessante vedere come i paesi colorati in verde, cioè quelli a maggioranza buddista, abbiamo tassi di fecondità molto bassi. Si tratta di Cina, Giappone, Taiwan e Korea del Sud. Poco sopra vi è una fascia di paesi cristiani, in blu, che riflette principalmente la tendenza occidentale a fare pochi figli. Salendo ancora nel grafico si incontrano paesi a

maggioranza islamica, con tassi di fecondità ancora alti, principalmente appartenenti al continente africano e le regioni mediorientali.

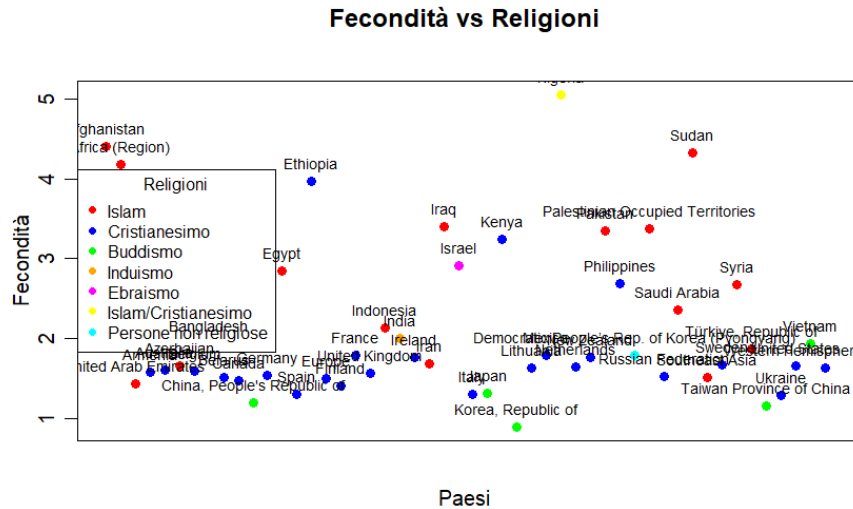


Figura 2.9: Fecondità vs. religione.

Infine, la figura 2.10 mostra come le forme di governo vigenti possano influenzare il tasso di crescita annuale del PIL, il GDP. Considerando che non si tratta di una relazione di causa effetto, è possibile però individuare dei pattern che legano le due variabili: il colore blu, che corrisponde alle repubbliche parlamentari ha due fasce di concentrazione, quella bassa con quote di GDP rasenti allo zero o negative a cui appartengono repubbliche di antica formazione (con l'eccezione dell'Iraq) come Germania (1949), Finlandia (1917) e Irlanda (1922), e la fascia superiore che delinea indici di sviluppo economico molto alti corrispondenti a repubbliche parlamentari più giovani, per esempio l'Armenia (1991), l'Etiopia (1991) e il Bangladesh (1971). Nella fascia mediana il grafico è popolato per lo più dalle monarchie costituzionali, in rosso, con indici di sviluppo economico bassi anche se per lo più positivi, paesi di questo tipo son il Belgio, il Canada, il Regno Unito e la Svezia, monarchie costituzionali formatesi durante il 1800 se non prima. Infine, il gruppo colorato in giallo, nella parte superiore del grafico, rappresenta le repubbliche presidenziali che nel 2023 hanno avuto un tasso di crescita del PIL intorno a 5. Di questi fanno parte, tra gli altri, Indonesia, Iran, Kenya, Filippine, Messico e Nigeria, tutti paesi che hanno

stabilito la propria forma di governo durante la seconda metà del 1900 in seguito alla conquista dell'indipendenza da paesi colonizzatori come Inghilterra, Paesi Bassi e Stati Uniti d'America.

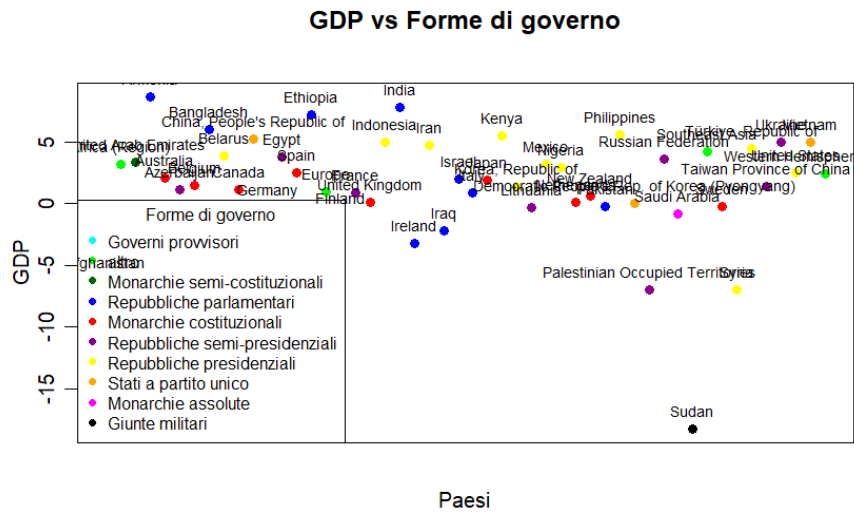


Figura 2.10: GDP vs. forme di governo.

Oltre alla analisi descrittive condotte sulle variabili e il campione di nazioni, sono state studiate le statistiche descrittive delle reti. Attraverso il conteggio degli archi in diverse modalità si possono ottenere informazioni importanti sulle strutture di rete. Di seguito vengono riportati i grafici delle 24 reti in studio e le relative analisi descrittive.

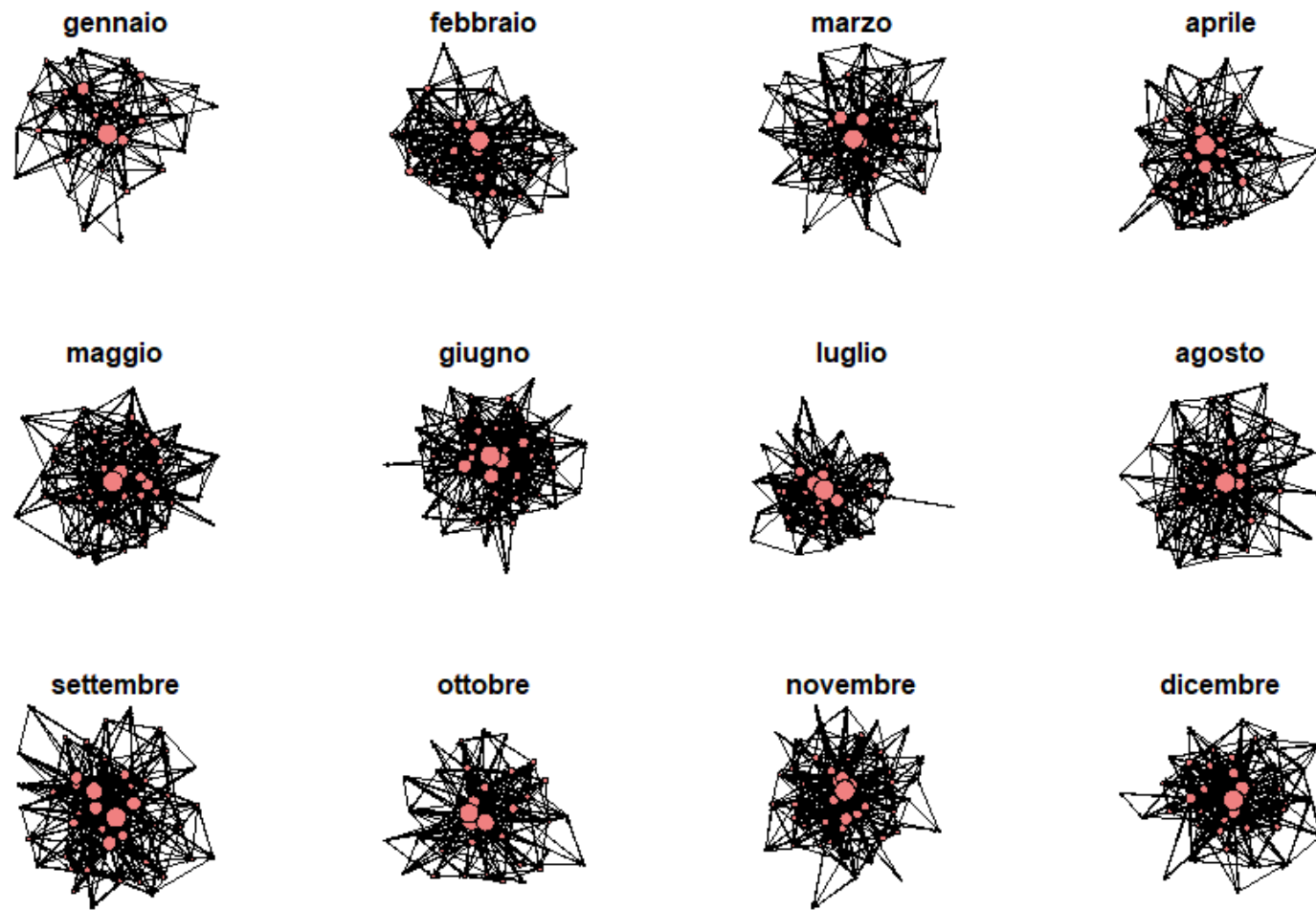


Figura 2.11: Grafici delle 12 reti di conflitto

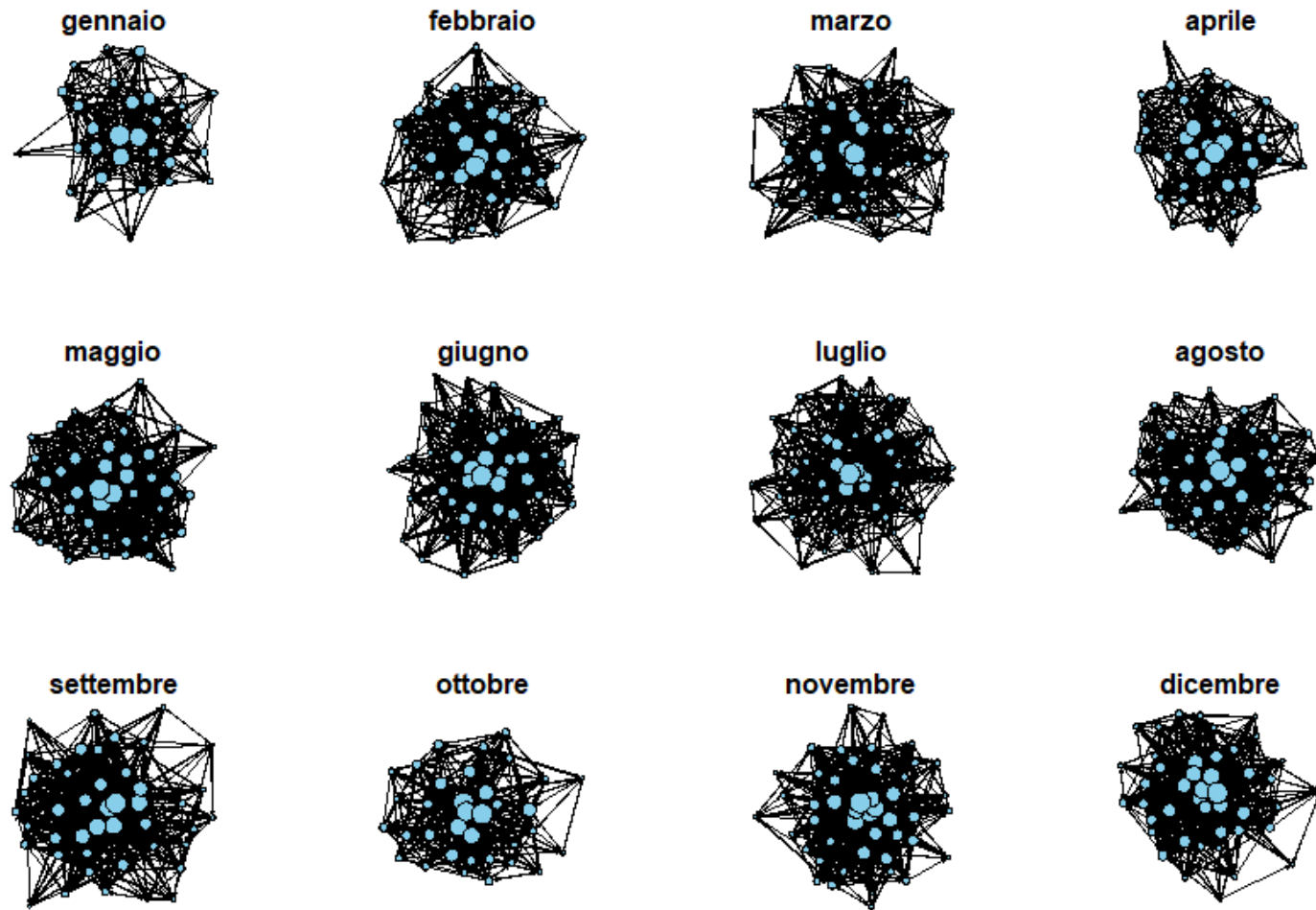


Figura 2.12: Grafici delle 12 reti di cooperazione

La prima osservazione utile riguarda il numero di archi presente in ogni rete mensile. In figura 2.11 sono riportati i grafici a barre che rappresentano la somma dei legami ricevuti o inviati da ogni paese per ogni mese del 2023. Ci sono principalmente due osservazioni di cui tenere nota. La prima riguarda la diversa scala a cui fanno riferimento i conteggi: nel primo caso, quello dei conflitti, l'istogramma ha una soglia massima che si osserva intorno ai cinquecento contatti mensili con un minimo di poco più di duecento legami nel mese di gennaio. Nel secondo caso, quello delle cooperazioni in blu, la scala di conteggio è nell'ordine dei mille, quindi doppiamente più ampia di quella precedente, mantenendo sempre il minimo numero dei rapporti nel mese di gennaio. La seconda osservazione riguarda la stagionalità dei rapporti. Come verrà confermato anche in seguito, i contatti positivi e negativi tra gli stati seguono un andamento che non sembra essere del tutto casuale. Come anticipato, le reti con meno legami sono quelle di gennaio con al seguito aprile e ottobre, mentre si verificano picchi dei rapporti nei mesi estivi e poi di nuovo nei mesi invernali. Infine, osservando i due andamenti simili, si può ipotizzare che l'aumento e la diminuzione dei legami nei due grafici abbiano una certa correlazione in quanto spesso i rapporti di assistenza sono consequenziali a quelli di conflitto.

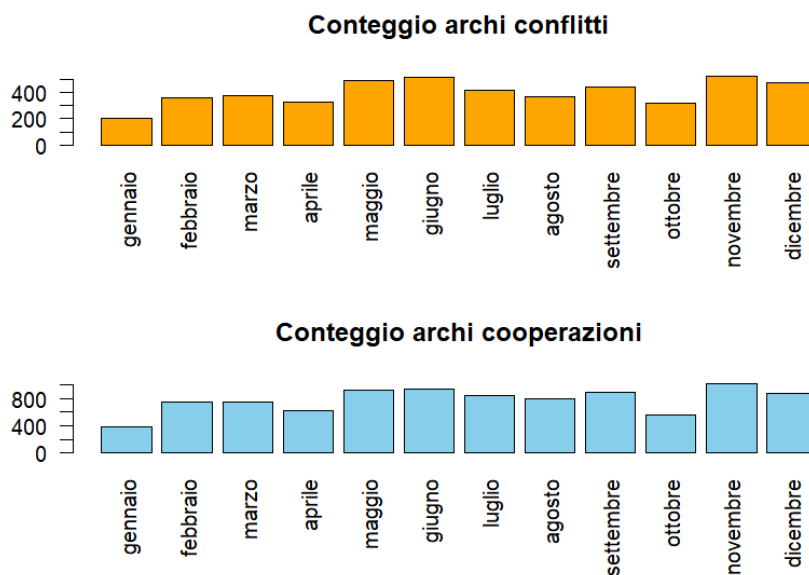


Figura 2.13: Conteggio degli archi.

I grafici in figura 2.12 e 2.13 rappresentano invece la somma dei con-

teggi mensili delle connessioni ricevute (a sinistra) e di quelle inviate (a destra) per gruppo religioso. I conteggi sono stati rapportati al numero di stati appartenenti a una determinata confessione in modo da avere una misura standardizzata. Sicuramente di spicco è il conteggio riferito ai paesi con confessione ebraica di cui fa parte solo uno stato, Israele. È al primo posto con un distacco di centinaia di legami rispetto ad altre confessioni, sia nel conteggio delle relazioni di contrasto sia in quelle di collaborazione. È interessante sottolineare che la confessione di Israele, l'ebraismo, è un tratto identitario dello stato e del popolo che vi abita che possa quindi influenzare anche il carattere dello stato nella sua varietà e diversità di rapporti con gli altri attori mondiali. Dall'altra parte, i conteggi minimi, per entrambe le coppie di grafici, si hanno in corrispondenza della rappresentanza non religiosa, di cui fa parte solo la Corea del Nord che sappiamo essere una dittatura totalitaria, chiusa in se stessa e nei confronti di altre nazioni.

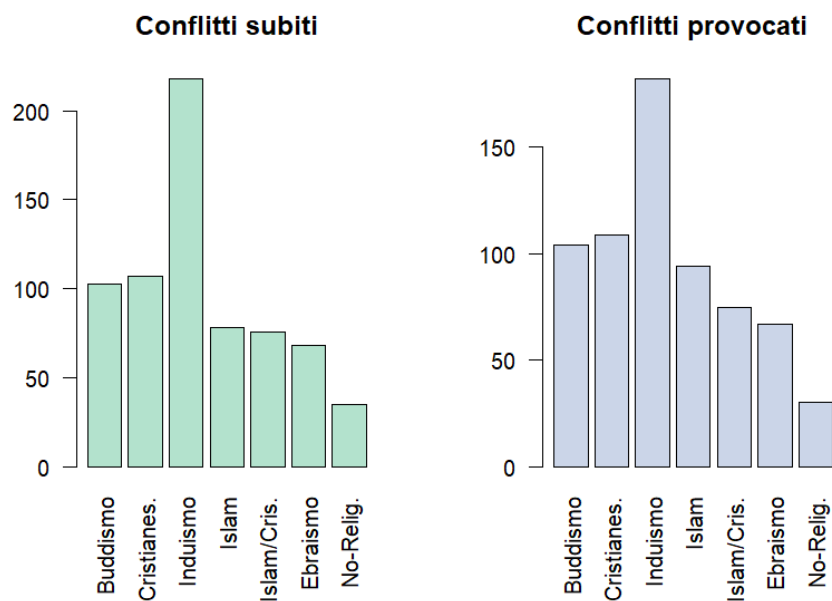


Figura 2.14: Conflitti su religioni.

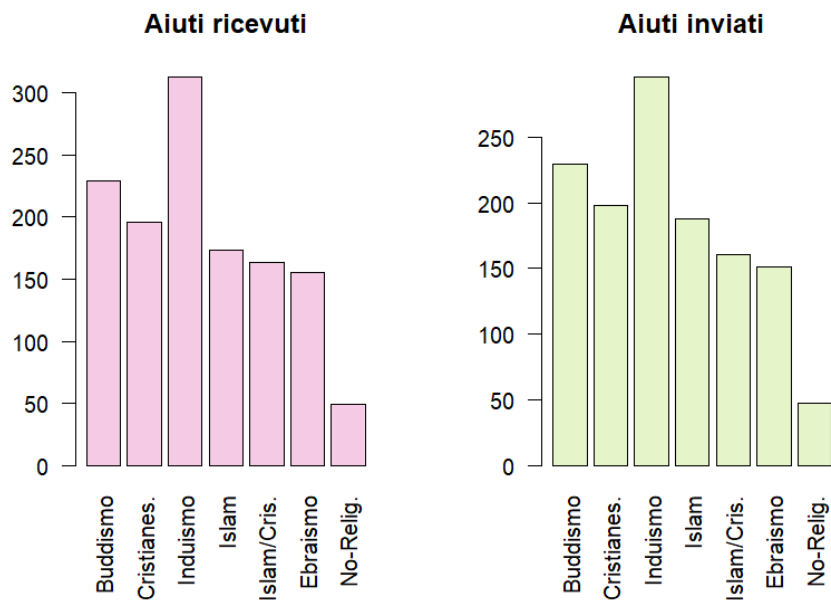


Figura 2.15: Aiuti su religioni

In ultima analisi si è scelto di riportare alcune statistiche di rete combinate. I grafici 2.14 e 2.15 sono la trasposizione grafica delle matrici di conteggio che per ogni mese registrano il numero di conflitti e cooperazioni verificatesi tra nazioni con determinate differenze di indice di sviluppo umano, in figura 2.14, e di tasso di crescita del Prodotto Interno lordo annuale, in figura 2.15. Nella prima coppia di grafici si possono osservare alcuni aspetti: in primo luogo si rileva nuovamente la stagionalità dei rapporti dettata dalle macchie più scure in corrispondenza dei mesi estivi e invernali, sia per i conflitti che per le cooperazioni. In secondo luogo si nota come i contatti avvengano tra paesi che non hanno grandi differenze di indice di sviluppo umano. Questo può significare che da una parte paesi sviluppati non entrano in conflittualità con paesi poco sviluppati ma dall'altro lato sono limitati anche i rapporti di sussistenza tra nazioni umanamente sviluppate e nazioni umanamente sottosviluppate. La figura 2.15 invece, rappresenta le differenze di prodotto interno lordo tra gli le coppie di soggetti coinvolti nei legami e presenta uno spettro di colori più ampio e quindi una maggior variabilità dei legami. I rapporti di cooperazione e conflitto si verificano anche tra paesi economicamente asimmetrici, seppur in quantità via via decrescente con l'aumentare della differenza di GDP. Bisogna tener conto che con l'aumentare della differenza dell'indice diminuisce il numero di

paesi con un GDP molto basso e quindi anche i possibili collegamenti di questo tipo. Rimangono più densi, in termini di rapporti, i periodi estivi e invernali.

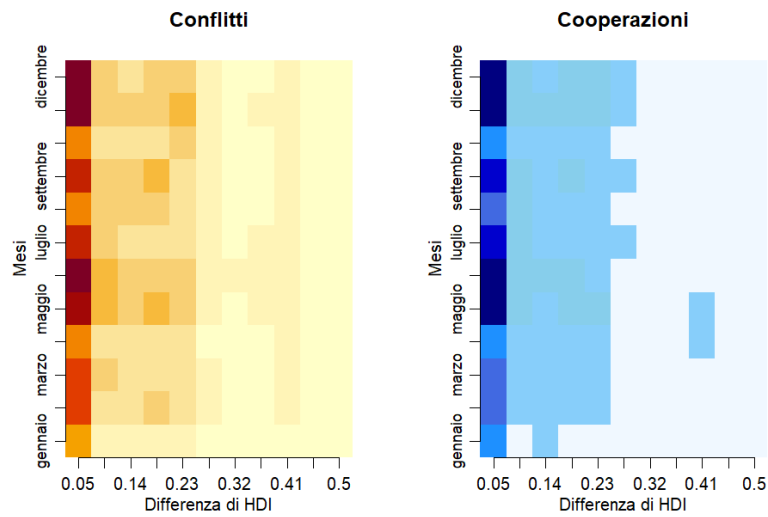


Figura 2.16: Mesi su differenza di HDI

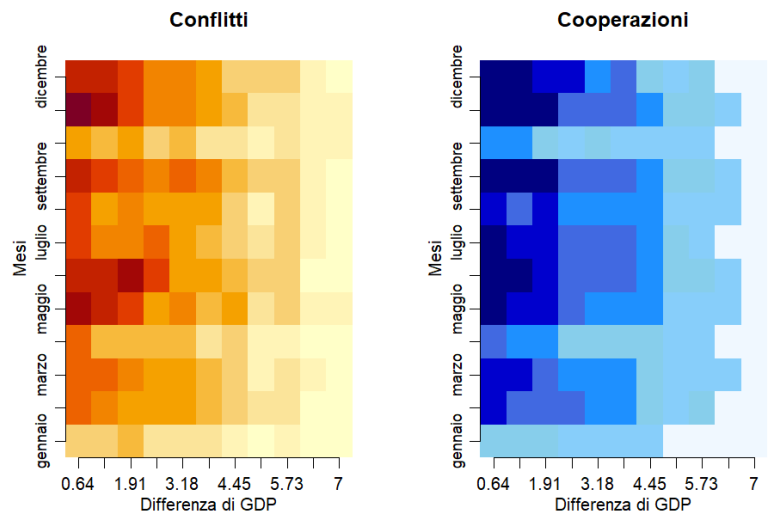


Figura 2.17: Mesi su differenza GDP

Capitolo 3

Modelli per reti sociali

In questo capitolo si presentano due importanti modelli per le strutture di rete: i Modelli per Grafi Casuali a Famiglia Esponenziale e i Modelli a Effetti Additivi e Moltiplicativi.

3.1 Le famiglie esponenziali

Una famiglia esponenziale è una famiglia F di distribuzioni per una variabile casuale univariata o multivariata con parametro $\theta \in \Theta \subseteq \mathbb{R}^p$, $p \geq 1$, e densità

$$p(y, \theta) = h(y) \cdot c(\theta) \cdot \exp \{ \omega(\theta)^\top \cdot t(y) \} \quad (3.1)$$

dove $h(\cdot) \geq 0$ e $\omega(\theta) = (\omega_1(\theta), \dots, \omega_k(\theta))$ è una funzione dei parametri con dominio $\Omega = \omega(\Theta) \subseteq \mathbb{R}^k$ e $t(y) = (t_1(y), \dots, t_k(y))^\top$ rappresenta il vettore di statistiche sufficienti.

Il supporto di Y è la chiusura di $\{y \in \mathbb{R}^d : h(y) > 0\}$, quindi Y è lo stesso per tutti i valori di θ . La funzione $c(\theta)$ è una costante di normalizzazione che nel caso continuo,

$$c(\theta) = \int_Y h(y) \exp\{\omega(\theta)^\top t(y)\} dy.$$

Affinché F sia non banale e θ identificabile, Θ deve contenere almeno due elementi e $\omega(\cdot)$ deve essere iniettiva, quindi $p \leq k$ se $\omega(\cdot)$ è continua e differenziabile. La rappresentazione (3.1) è chiamata minimale se coinvolge il numero minimo possibile di funzioni $\omega_j(\theta)$ e statistiche associate $t_j(y)$, cioè se:

1. Θ contiene almeno $k+1$ elementi, dove k è il numero di componenti della funzione $\omega(\cdot)$;
2. le $k+1$ funzioni reali $1, \omega_1(\theta), \dots, \omega_k(\theta)$ sono linearmente indipendenti;
3. allo stesso modo, le $k+1$ funzioni $1, t_1(y), \dots, t_k(y)$ devono essere linearmente indipendenti.

In una famiglia esponenziale con rappresentazione minimale, k è l'ordine della famiglia e la funzione $t(y) = (t_1(y), t_2(y), \dots, t_k(y))$ è chiamata statistica canonica. Spesso $k = p$ e $\omega(\theta)$ è una funzione biunivoca, quindi è una riparametrizzazione del modello, con $\omega(\theta)$ differenziabile nell'interno di Θ , insieme alla sua inversa $\theta = \theta(\omega)$. In questo caso $\omega = \omega(\theta)$ è chiamato parametro canonico. Nel caso in cui la dimensione dello spazio parametrico Θ è ridotto rispetto a k , quindi se $p < k$, la famiglia esponenziale è detta curva.

Considerando quindi una famiglia esponenziale di ordine p nella parametrizzazione canonica si può arrivare a una densità della forma:

$$p_Y(y; \omega) = c(\theta(\omega))h(y) \exp(\omega(\theta)^\top t(y)) = h(y) \exp(\omega^\top t(y) - K(\omega)) \quad (3.2)$$

Chiamata forma canonica della famiglia esponenziale (Pace et al., 2001). Questa rappresentazione è utile in primo luogo ad ottenere la funzione generatrice dei momenti tramite la funzione $K(\omega)$, e in secondo luogo, di maggiore importanza per l'elaborato, ad evidenziare il collegamento con i modelli di grafi casuali a famiglia esponenziale, gli ERGM, dove la variabile esplicativa y rappresenta la rete osservata.

3.2 Grafi Casuali a Famiglia Esponenziale, gli ERGM

Gli ERGM sono una classe di modelli per reti, o grafi, che si basano sulla specificazione della probabilità di osservare un certo grafo Y in funzione di alcune caratteristiche strutturali del grafo stesso. Formalmente, un grafo casuale $Y \in \mathcal{Y}$ consiste in un insieme di n nodi e una collezione di variabili di connessione $\{Y_{ij} : i = 1, \dots, n; j = 1, \dots, n\}$, indicizzate da coppie di nodi ij , dove $Y_{ij} = 1$ se i nodi (i, j) sono collegati da un arco e $Y_{ij} = 0$ altrimenti. Una coppia di nodi ij è chiamata

diade e una diade è un arco se $Y_{ij} = 1$ (Wikipedia, 2024).

In questo contesto, il grafo osservato y è solitamente rappresentata come una matrice di adiacenza dove ogni elemento y_{ij} indica la presenza ($y_{ij} = 1$) o l'assenza ($y_{ij} = 0$) di un legame tra il nodo i e il nodo j . Considerando n nodi ed escludendo i “loop” cioè i legami di un nodo con se stesso (casi che nell'elaborato sono stati esclusi), il grafo avrà la seguente struttura:

$$y = \begin{pmatrix} 0 & y_{12} & y_{13} & \dots & y_{1n} \\ y_{21} & 0 & y_{23} & \dots & y_{2n} \\ y_{31} & y_{32} & 0 & \dots & y_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & y_{n3} & \dots & 0 \end{pmatrix}$$

Attraverso i grafi si possono mettere in luce caratteristiche importanti delle relazioni sociali. Una delle più importanti è la direzionabilità: vi sono infatti reti dirette e reti indirette. Nelle prime, il generico elemento y_{ij} rappresenta la relazione che il nodo i stabilisce con il nodo j e non viceversa, ed implica una raffigurazione tramite matrice di adiacenza non simmetrica. Nel caso in studio, considerando un insieme di n paesi del mondo come nodi, un possibile grafo per relazioni direzionate potrebbe raffigurare l'insieme di “tutti i paesi che hanno avuto atteggiamenti conflittuali verso altri paesi”, perciò se il paese i ha inflitto danno in qualche modo al paese j , la variabile diadica che corrisponde al rapporto d'interesse sarà $Y_{ij} = 1$. Solo se il paese j , a sua volta ha ricambiato l'azione nei confronti di i ci sarà una biunivocità di rapporto con il valore della variabile $Y_{ji} = 1$, ma non è detto che questo avvenga. Per le reti dirette, considerando sempre n attori, ci potranno essere una totalità di $2^{n(n-1)}$ configurazioni del grafo. Nel secondo caso, quello delle relazioni non direzionate, il generico elemento y_{ij} delinea la biunivocità del rapporto tra i due attori senza considerare un mittente e un ricevente, la matrice di adiacenza in questo caso sarà simmetrica e le possibili configurazioni saranno pari a $2^{n(n-1)/2}$.

Una volta costruita la struttura che rappresenta l'insieme delle relazioni di interesse, l'assunto di base di questi modelli è che il grafo osservato y possa essere spiegato attraverso un dato vettore di statistiche sufficienti $s(y)$, che sono funzioni della rete osservata e degli attributi dei nodi (Kolaczyk, 2009). In questo modo, le caratteristiche della rete, come possono essere il numero di archi, il numero di triangoli o gli scambi mutuali di rapporto, insieme alle caratteristiche dei nodi, sono

utilizzati come statistiche di rete e formeranno le covariate del modello. Queste ultime si dividono essenzialmente in due tipologie: i termini “diadi dipendenti” e quelli “diadi indipendenti”.

I termini diadi dipendenti considerano la rete come composta da attori reciprocamente dipendenti, sono più onerosi da stimare ma possono modellare aspetti essenziali della dinamica di rete: per esempio l’effetto di reciprocità tra attori, ovvero la tendenza di due nodi a formare un legame bidirezionale. Oppure, la propensione per due nodi ad avere un legame in comune attraverso un terzo nodo. I termini diadi indipendenti considerano le caratteristiche dei soggetti nella rete in modo indipendente l’una dall’altra, permettendo di esaminare come le specificità dei nodi influenzino la formazione di collegamenti tra di loro, senza considerare direttamente le dipendenze tra le coppie di nodi: per esempio, si può catturare l’effetto della “popolarità” di un nodo nella rete attraverso il conteggio delle connessioni in ingresso. Oppure si possono considerare effetti di omofilia cioè la tendenza dei nodi con caratteristiche simili a connettersi tra loro tramite per esempio il conteggio degli archi formati tra nodi che condividono gli stessi valori delle variabili.

Il modello ERGM stimato rappresenta la distribuzione di probabilità discreta delle possibili configurazioni di rete con n attori, in base alle statistiche inserite nel modello che si vuole spieghino le dinamiche di formazione del grafo, fornendo quindi delle indicazioni su quali grafi è più probabile che si verifichino dato un insieme di n soggetti.

La probabilità di un grafo y , dove $y = [y_{12}, y_{13}, \dots, y_{n(n-1)}]^T$, in un ERGM è data da (Statnet, 2023):

$$P(Y = y) = \frac{\exp\{\theta^\top s(y)\}}{\kappa(\theta)} \quad (3.3)$$

dove:

- θ è il vettore di parametri del modello.
- $s(y)$ è il vettore di statistiche del grafo (ad esempio, numero di legami, numero di triangoli, etc.).
- $\kappa(\theta)$ è la funzione di normalizzazione che assicura che la somma delle probabilità su tutti i grafi possibili sia 1.

È immediato stabilire il parallelismo con la forma canonica delle famiglie esponenziali in equazione (3.2);

Riscrivendo la probabilità di un grafo y otteniamo:

$$P(Y = y) = \exp(\theta^\top s(y) - \log(\kappa(\theta))) \quad (3.4)$$

Con:

- $\theta^\top s(y) = \omega^\top t(y)$
- $\log \kappa(\theta) = K(\omega)$ funzione di normalizzazione

Si propone di seguito un breve esempio per comprensione del modello.

3.2.1 Un esempio di ERGM

Supponiamo di avere un insieme di solamente 3 attori $V = \{1, 2, 3\}$. Sia poi \mathcal{Y} l'insieme di tutti i grafi non direzionati che possono formarsi a partire da V , per cui $Y_{ij} = Y_{ji}$, e senza loop, quindi $Y_{ii} = 0$. Sappiamo che il numero dei possibili grafi che si possono creare a partire da 3 soggetti con relazione indiretta è $2^{n(n-1)/2}$ che in questo caso corrisponde a $2^{3(3-1)/2} = 2^3 = 8$.

Si vuole costruire un modello che spieghi la formazione del grafo attraverso un vettore di due semplici statistiche di rete: il numero di archi e il numero di triangoli formati. Il vettore di statistiche sarà quindi $s(y) = [s_1(y), s_2(y)]^T$ dove $s_1(y) = \text{edges}(y)$ è definito come il numero di archi nel grafo y , e $s_2(y) = \text{triangles}(y)$ è definito come il numero di triangolazioni in y . Definiamo, a fini illustrativi, il corrispondente vettore di parametri come $\theta = (\theta_1, \theta_2)^T = (-\ln(2), \ln(3))^T$.

Abbiamo ora tutti gli elementi per un ERGM:

$$P(Y = y \mid \theta) = \frac{\exp(-\ln(2) \cdot \text{edges}(y) + \ln(3) \cdot \text{triangles}(y))}{\kappa(\theta)}$$

Avendo un numero ridotto di nodi, le possibili combinazioni di grafi sono quattro: un grafo con zero archi, tre grafi con un arco, tre grafi con due archi e un grafo con tre archi.

Considerando per adesso solo il numeratore del modello, si verificano i seguenti scenari:

Se y è il grafo con zero archi, allora $\text{edges}(y) = 0$ e $\text{triangles}(y) = 0$, quindi

$$\exp(-\ln(2) \cdot 0 + \ln(3) \cdot 0) = \exp(0) = 1.$$

Se y è un grafo con esattamente un arco, allora $\text{edges}(y) = 1$ e $\text{triangles}(y) = 0$, quindi

$$\exp(-\ln(2) \cdot 1 + \ln(3) \cdot 0) = \frac{\exp(0)}{\exp(\ln(2))} = \frac{1}{2}.$$

Se y è un grafo con esattamente due archi, allora $\text{edges}(y) = 2$ e $\text{triangles}(y) = 0$, quindi

$$\exp(-\ln(2) \cdot 2 + \ln(3) \cdot 0) = \frac{\exp(0)}{\exp(\ln(2))^2} = \frac{1}{4}.$$

Se y è il grafo con esattamente tre archi, allora $\text{edges}(y) = 3$ e $\text{triangles}(y) = 1$, quindi

$$\exp(-\ln(2) \cdot 3 + \ln(3) \cdot 1) = \frac{\exp(\ln(3))}{\exp(\ln(2))^3} = \frac{3}{8}.$$

La costante di normalizzazione $c(\theta)$ è calcolata sommando il numeratore su tutti gli otto grafi differenti $y \in \mathcal{Y}$. Questo produce:

$$\kappa(\theta) = 1 + 3 \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} + \frac{3}{8} = \frac{29}{8}.$$

Una volta calcolati questi elementi, la distribuzione di probabilità di tutti i possibili grafi è ora di facile accesso. Per ottenere la probabilità di uno specifico grafo $y \in \mathcal{Y}$ basta calcolare

$$P(Y = y \mid \theta) = \frac{\exp\{\theta^\top s(y)\}}{\kappa(\theta)} = \frac{8}{29} \exp(-\ln(2) \times \text{edges}(y) + \ln(3) \times \text{triangles}(y))$$

Risulta quindi che un grafo con zero archi ha probabilità $8/29$.

Ogni grafo con un arco ha probabilità $4/29$.

I grafi con due archi hanno probabilità $2/29$ e infine, il grafo con tre archi ha probabilità $3/29$. Ricordando che il modello è una distribuzione di probabilità, i coefficienti aiutano a capire quali sono le configurazioni dei grafi più probabili: un coefficiente negativo per il termine riferito al numero di archi, mantenendo costanti gli altri termini, indica che i grafi con più archi sono meno probabili di quelli con meno archi. Viceversa, il coefficiente positivo riferito al termine sul numero di triangoli significa che, mantenendo costanti gli altri termini, le reti con più triangoli hanno probabilità più alta di verificarsi rispetto a quelle con meno triangoli.

3.2.2 L'estensione multistrato

Un'importante estensione degli ERGM è quella dei modelli multistrato. Anche conosciuti come reti multiplex, multirelazionali o multivariate,

in una rete multistrato ogni attore può avere relazioni simultane multiple in ogni strato della rete (Krivitsky et al., 2020). Questo tipo di modelli può essere utilizzato per affrontare domande di ricerca sulla struttura sociale e sui processi che possono svilupparsi sia all'interno che tra gli strati. Le applicazioni più comuni delle reti multistrato si interessano di stabilire la probabilità di formazione di un legame tra i soggetti di uno specifico strato come può essere per esempio quello dei legami di amicizia all'interno di un gruppo di persone, data l'esistenza di legami in un altro strato del modello, che rappresenta invece i rapporti di fiducia. In altre parole, un modello di rete multistrato consente di considerare domande del tipo: “con che probabilità si forma un legame di amicizia data la presenza di un legame di fiducia tra due attori?”, “Le amicizie tendono a formarsi tra attori che considerano le stesse persone affidabili?”

In questo elaborato però, a differenza dell'esempio precedente, i modelli di rete multistrato non sono stati utilizzati per coniugare aspetti sociali diversi come possono essere la fiducia e l'amicizia, ma piuttosto, le due reti rappresentano i rapporti di conflitto o di cooperazione degli stati coinvolti, ogni mese del 2023. In questo modo i modelli potranno rispondere a domande come: “che probabilità c'è che due paesi abbiano una relazione di conflitto, data la loro condotta durante i diversi mesi del 2023?”. Oppure: “Esiste una relazione tra il tipo di cooperazioni che si verificano in un determinato mese e i fattori geopolitici o economici che caratterizzano gli stati coinvolti?”.

L'estensione del modello alle reti multistrato è di facile intuizione e di seguito se ne fornisce una breve presentazione. Si veda (Krivitsky e Koehly, 2020) per una descrizione più dettagliata.

Sia Y una rete multilivello casuale tra un insieme comune di attori $N = \{1, \dots, n\}$, su L relazioni binarie dove $Y_{i,j,l}$ indica la presenza (= 1) o l'assenza (= 0) della l -esima relazione tra gli attori i e j ($i \neq j$ per ogni l). Sia y una realizzazione di Y . Questa rete multilivello può essere considerata come una composizione di L reti binarie, y_1, \dots, y_L o come una rete multivariata, dove ogni diade $y_{i,j}$ è un vettore L binario $[y_{i,j,l}]_{l=1}^L$. Per specificare un ERGM su una tale rete, definiamo l'insieme delle relazioni potenziali all'interno di ciascun livello: Y_l , $l = 1, \dots, L$ che per un livello diretto, $Y_l \subseteq \{(i, j) : (i, j) \in N \times N\}$. Per convenienza, definiamo anche $\mathbb{Y} \equiv \mathbb{Y}_1 \cup \dots \cup \mathbb{Y}_L$ come l'insieme di diadi per le quali una relazione può esistere in minimo uno strato. Lo spazio campionario per l'ERGM multistrato è l'insieme dei possibili sottoin-

siemi delle relazioni potenziali di interesse, quindi $Y \subseteq 2^{\mathbb{Y}_1 \times \dots \times \mathbb{Y}_L}$. Un ERGM è specificato dalla sua funzione vettoriale $s(Y)$ di statistiche sufficienti che operazionalizzano le caratteristiche di interesse della rete multilivello, incluse quelle dei suoi livelli costituenti individualmente o tramite varie combinazioni di esse, e, in alcune formulazioni (Hunter e Handcock 2006; Krivitsky 2017), una mappatura $\eta(\theta)$ del vettore dei parametri del modello θ ai parametri naturali.

Considerando i livelli di un modello multistrato come condizionatamente indipendenti, può essere specificato un modello ERGM per ogni strato della rete individualmente:

$$\Pr(Y_l = y_l) = \frac{\exp\{\eta_l(\theta_l) \cdot s_l(y_l)\}}{\kappa_l(\theta_l)} \quad (3.5)$$

E con una verosimiglianza complessiva che quindi risulta essere:

$$\prod_{l=1}^L \Pr(Y_l = y_l) = \exp \left\{ \sum_{l=1}^L \eta_l(\theta_l) \cdot s_l(y_l) \right\} / \prod_{l=1}^L \kappa_l(\theta_l). \quad (3.6)$$

Nel caso di indipendenza condizionale tra i livelli della rete, la formula (3.6) rappresenta un ERGM multistrato, le cui statistiche sufficienti $s(y)$ sono date da $[s_1(y_1), \dots, s_L(y_L)]$, e il vettore dei parametri $\eta(\theta)$ è dato da $[\eta_1(\theta), \dots, \eta_L(\theta)]$. Anche se le statistiche $s_k(\cdot)$ per più livelli sono le stesse, non viene fatta alcuna assunzione se le forze sociali che esse incorporano abbiano la stessa intensità e direzione ma rimangono eterogenee. Tuttavia, è possibile considerare modelli in cui si assume che l'intensità e la direzione siano uguali - al netto di altre forze sociali, anche se i livelli stessi sono ancora stocasticamente indipendenti, contribuendo a una migliore parsimonia e maggiore potenza statistica. È il caso in cui l'effetto $s_k(\cdot)$ è considerato omogeneo tra i livelli con coefficiente $\eta_k(\theta_k)$ e può essere espresso attraverso un modello congiunto (Krivitsky et al., 2020):

$$\begin{aligned} \Pr(Y_1 = y_1, \dots, Y_L = y_L) &= \prod_{l=1}^L \frac{\exp(\eta_k(\theta_k) \cdot s_k(y_l))}{\kappa_l(\theta_l, \theta_k)} \\ &= \exp \left\{ \eta_k(\theta_k) \sum_{l=1}^L s_k(y_l) \right\} / \kappa(\theta_1, \dots, \theta_L, \theta_k) \end{aligned} \quad (3.7)$$

L'utilizzo e l'interpretazione del modello in equazione (3.7) viene mostrata nel paragrafo successivo.

3.2.3 Applicazione del modello ERGM

Le statistiche riferite ai termini diadi indipendenti rappresentano caratteristiche comuni a tutti gli strati, che verranno quindi modellate come covariate uniche trasversali ai dodici livelli attraverso la formulazione della seconda parte del modello in equazione (3.7). Procedendo con una tecnica del tipo “stepwise forward” manuale, sono stati aggiunti al modello i seguenti termini sequenzialmente:

- Edges: termine diade dipendente, rappresenta il conteggio degli archi nella rete. Il coefficiente degli “edges” corrisponde al logaritmo del rapporto tra la probabilità di un collegamento esistente e la probabilità di un collegamento inesistente tra due nodi qualsiasi nella rete, a parità di tutte le altre condizioni.
- Attributi di nodo:
 - Nodefactor(“Religioni”)
 - Nodefactor(“Forme di governo”)

I secondi due termini sono diadi indipendenti e per ogni modalità delle variabili “Religione” e “Sistema di governo” rappresentano il conteggio dei nodi con la specifica categoria e che appaiono collegati da un arco. I coefficienti relativi a questi termini forniscono indicazioni sulle modalità delle variabili categoriali più influenti alla spiegazione del grafo.

- Covariate nodali:
 - Nodecov(“GDP”)
 - Nodecov(“HDI”)
 - Nodecov(“Tasso di fecondità”)
 - Nodecov(“Densità di popolazione”)

Sono stati inseriti in seguito altri quattro termini diadi indipendenti relativi alle variabili numeriche dei nodi. Sono il corrispettivo quantitativo degli “Attributi di nodo”, perciò rappresentano i conteggi dell’attributo considerato sulle diadi connesse da un arco. Se consideriamo per esempio l’indice di sviluppo umano, il termine di “Covariate nodali” calcolerà la somma dell’indice di sviluppo umano sui nodi collegati da un arco, nell’ottica che questo conteggio possa essere utile a valutare come la somma degli HDI dei nodi collegati influenzi la formazione dei legami nella rete.

- Mutual: l'ultimo termine aggiunto al modello è diade dipendente che comporta una tempo di calcolo decisamente più lungo rispetto a un modello con soli termini diadi-indipendenti. "Mutual" rappresenta il numero di coppie di attori i e j per cui vi è un'interazione del primo nei confronti del secondo ($i \rightarrow j$) e viceversa ($j \rightarrow i$). Nell'ambito dei conflitti e delle cooperazioni questo termine può essere molto utile a spiegare la probabilità di formazione di un determinato grafo poiché spesso, atteggiamenti ostili causano risposte ostili, e viceversa, gesti cooperativi possono incentivare posizioni cooperative.

Sia per la rete dei conflitti che per quella delle cooperazioni sono stati eseguiti gli stessi modelli in modo da poter stabilire un confronto diretto.

| | Religioni | | | | | Forme di governo | | | |
|------------------------|---------------------|--------------------|---------------------|--------------------|--------------------------------|--------------------|--------------------|--------------------|---------|
| | Model 1 | Model 2 | Model 3 | Model 4 | | Model 1 | Model 2 | Model 3 | Model 4 |
| edges | -1.63*** (0.02) | -2.69*** (0.14) | -11.22*** (0.50) | -8.70*** (0.42) | | | | | |
| Cristianesimo | | 0.60*** (0.05) | 0.34*** (0.07) | 0.21*** (0.05) | Giunte militari | -0.72*** (0.13) | -0.67*** (0.17) | -0.45** (0.14) | |
| Ebraismo | | 1.91*** (0.09) | 1.18*** (0.11) | 0.74*** (0.09) | Governi provvisori | 0.17 (0.10) | 0.69*** (0.12) | 0.44*** (0.10) | |
| Induismo | | 0.75*** (0.10) | 1.69*** (0.12) | 1.08*** (0.09) | Monarchie assolute | 0.20* (0.10) | -0.64*** (0.10) | -0.41*** (0.08) | |
| Islam | | 0.17** (0.06) | 0.38*** (0.07) | 0.24*** (0.06) | Monarchie costituzionali | 0.12* (0.05) | -0.25*** (0.05) | -0.16*** (0.04) | |
| Islam/Cristianesimo | | -0.17 (0.11) | 0.04 (0.15) | 0.04 (0.12) | Monarchie semi-costituzionali | -0.35** (0.11) | -1.11*** (0.12) | -0.71*** (0.10) | |
| Persone non religiose | | -1.77*** (0.14) | -0.70*** (0.15) | -0.47*** (0.13) | Repubbliche parlamentari | -0.36*** (0.05) | -0.34*** (0.05) | -0.22*** (0.05) | |
| | | | | | Repubbliche presidenziali | 0.27*** (0.05) | 0.42*** (0.05) | 0.27*** (0.04) | |
| | | | | | Repubbliche semi-presidenziali | 0.33*** (0.05) | 0.39*** (0.05) | 0.25*** (0.04) | |
| | | | | | Stati a partito unico | 1.05*** (0.08) | 1.43*** (0.09) | 0.90*** (0.07) | |
| | Covariate numeriche | | | | | | | | |
| | Model 1 | Model 2 | Model 3 | Model 4 | | | | | |
| GDP | | | -0.02** (0.01) | -0.01** (0.00) | | | | | |
| HDI | | | 4.85*** (0.23) | 3.08*** (0.20) | | | | | |
| Tasso di fecondità | | | 0.28*** (0.03) | 0.18*** (0.03) | | | | | |
| Densità di popolazione | | | -0.00*** (0.00) | -0.00*** (0.00) | | | | | |
| Mutual | | | | 3.25*** (0.06) | | | | | |
| AIC | 26190.25 | 24983.33 | 24195.96 | 20617.73 | | | | | |
| BIC | 26198.54 | 25115.95 | 24361.73 | 20791.79 | | | | | |
| Log Likelihood | -13094.12 | -12475.66 | -12077.98 | -10287.86 | | | | | |

***p < 0.001; **p < 0.01; *p < 0.05

Tabella 3.1: Modelli per i conflitti. Stima dei coefficienti

| | Religioni | | | | | Forme di governo | | | |
|------------------------|--------------------|--------------------|--------------------|--------------------|--------------------------------|--------------------|--------------------|--------------------|---------|
| | Model 1 | Model 2 | Model 3 | Model 4 | | Model 1 | Model 2 | Model 3 | Model 4 |
| edges | -0.76*** (0.01) | -0.95*** (0.10) | -9.96*** (0.41) | -7.87*** (0.31) | | | | | |
| Cristianesimo | | 0.21*** (0.04) | -0.11* (0.05) | -0.06 (0.04) | Giunte militari | -1.19*** (0.10) | -0.57*** (0.14) | -0.33** (0.10) | |
| Ebraismo | | 1.39*** (0.08) | 0.47*** (0.09) | 0.26*** (0.07) | Governi provvisori | -0.24** (0.08) | 0.46*** (0.09) | 0.26*** (0.07) | |
| Induismo | | 0.45*** (0.08) | 1.10*** (0.09) | 0.61*** (0.06) | Monarchie assolute | 0.62*** (0.07) | -0.19* (0.08) | -0.10 (0.06) | |
| Islam | | -0.06 (0.04) | 0.15** (0.05) | 0.08* (0.04) | Monarchie costituzionali | 0.13** (0.04) | -0.19*** (0.04) | -0.11*** (0.03) | |
| Islam/Cristianesimo | | -0.20* (0.08) | -0.08 (0.11) | -0.04 (0.08) | Monarchie semi-costituzionali | 0.30*** (0.07) | -0.54*** (0.08) | -0.30*** (0.06) | |
| Persone non religiose | | -2.54*** (0.12) | -1.29*** (0.13) | -0.73*** (0.10) | Repubbliche parlamentari | -0.42*** (0.04) | -0.29*** (0.04) | -0.16*** (0.03) | |
| | | | | | Repubbliche presidenziali | -0.01 (0.04) | 0.11** (0.04) | 0.06* (0.03) | |
| | | | | | Repubbliche semi-presidenziali | 0.11** (0.04) | 0.22*** (0.04) | 0.12*** (0.03) | |
| | | | | | Stati a partito unico | 0.92*** (0.07) | 1.19*** (0.07) | 0.66*** (0.05) | |
| Covariate numeriche | | | | | | | | | |
| | | | Model 3 | Model 4 | | | | | |
| GDP | | | 0.02*** (0.00) | 0.01** (0.00) | | | | | |
| HDI | | | 5.02*** (0.19) | 2.80*** (0.15) | | | | | |
| Tasso di fecondità | | | 0.32*** (0.03) | 0.18*** (0.02) | | | | | |
| densità di popolazione | | | -0.00*** (0.00) | -0.00*** (0.00) | | | | | |
| mutual | | | | 4.61*** (0.06) | | | | | |
| AIC | 36770.91 | 34895.86 | 33747.54 | 24605.90 | | | | | |
| BIC | 36779.20 | 35028.48 | 33913.32 | 24779.97 | | | | | |
| Log Likelihood | -18384.45 | -17431.93 | -16853.77 | -12281.95 | | | | | |

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Tabella 3.2: Modelli per le cooperazioni. Stima dei coefficienti

Per analizzare i modelli più facilmente e poter avere degli strumenti per l'interpretazione dei parametri è fondamentale esporre un'altra formulazione della probabilità che un modello ERGM fornisce. L'espressione per l'intero grafo mostrata in (3.7) può essere riformulata in termini di log-odds condizionali di un singolo legame tra due attori (Statnet, 2023):

$$\text{logit}P(Y_{ij} = 1 \mid y_{ij}^c) = \theta\delta(y_{ij}) \quad (3.8)$$

Dove:

- Y_{ij} è la variabile casuale per lo stato della coppia di attori i, j (con realizzazione y_{ij}).
- y_{ij}^c denota il complemento di y_{ij} , cioè tutte le diadi nella rete eccetto y_{ij} .
- $\delta(y_{ij})$ è un vettore delle “statistiche di cambiamento” per ciascun termine del modello. La statistica di cambiamento registra come cambia il termine $s(y)$ se il legame y_{ij} viene attivato o disattivato. Quindi:

$$\delta(y_{ij}) = s(y_{ij}^+) - s(y_{ij}^-)$$

Dove: y_{ij}^+ è definito come y_{ij}^c insieme a y_{ij} impostato a 1, e y_{ij}^- è definito come y_{ij}^c insieme a y_{ij} impostato a 0.

Quindi $\delta(y_{ij})$ è uguale al valore di $s(y)$ quando $y_{ij} = 1$ meno il valore di $s(y)$ quando $y_{ij} = 0$, ma tutte le altre diadi sono come in y .

L'equazione numero (3.8) mostra che il coefficiente θ può essere interpretato come il contributo del relativo termine al log-odds di un legame individuale, condizionato al fatto che tutte le altre diadi rimangano invariate. Il coefficiente per ciascun termine del modello è moltiplicato per il numero di configurazioni che quel legame creerà, o rimuoverà per quel termine specifico (ovvero la statistica di cambiamento, $\delta(y_{ij})$). Per esempio, disponendo di un modello con coefficiente relativo al numero di triangoli formati tra gli attori e volendo calcolare il log-odds della probabilità di formazione del legame ij dovremmo verificare quante triangolazioni aggiuntive apporterebbe il legame ij al grafo e moltiplicarle per il relativo coefficiente.

Le tabelle in figura 3.1 e 3.2 riportano per colonna i coefficienti dei modelli stimati con i relativi standard error e l'indicazione della significatività delle stime. Il primo modello contiene solo il termine di conteggio

degli archi *edges* senza ulteriori covariate. La stima del coefficiente risulta negativa sia per i conflitti che per le cooperazioni, indicando una probabilità bassa di osservare grafi densi rispetto a probabilità più alte di osservare quelli con un numero più ristretto di archi. Questo effetto è più forte nella rete dei conflitti con un parametro stimato di -1.63 contro -0.76 nel modello sulle cooperazioni, rispecchiando il maggior numero di connessioni osservate. Seguendo l'interpretazione in (3.8), con la statistica di cambiamento per gli archi che equivale a 1, avremo:

Modello sui conflitti:

$$\begin{aligned}\log\text{-odds}P(Y_{ij} = 1 \mid y_{ij}^c) &= \theta \times 1 = -1.63 \\ P(Y_{ij} = 1 \mid y_{ij}^c) &= 0.164\end{aligned}$$

Modello sulle cooperazioni:

$$\begin{aligned}\log\text{-odds}P(Y_{ij} = 1 \mid y_{ij}^c) &= \theta \times 1 = -0.76 \\ P(Y_{ij} = 1 \mid y_{ij}^c) &= 0.318\end{aligned}$$

I modelli stimano una probabilità leggermente più alta di osservare la formazione di un arco nelle reti, al netto delle altre diadi, sulle cooperazioni rispetto a quelle sui conflitti.

I secondi modelli implementati contengono il termine di conteggio dei soggetti appartenenti alle diverse modalità della variabile *religioni* e *sistema di governo*. Attraverso questi fattori sono quantificate le stime dell'influenza che la religione e il sistema di governo di una nazione possono esercitare nello stabilire legami con altri stati. È interessante osservare come i coefficienti relativi alla modalità “ebraismo” siano alti in entrambi i modelli, a significare l'influenza di questa caratteristica sui rapporti internazionali. Si osserva anche, per esempio, che il coefficiente legato alla modalità “cristianesimo” è 0.60 per i conflitti e 0.21 per le cooperazioni. Per i sistemi di governo invece, modalità come “stato mono partitico”, “repubblica presidenziale”, “governo provvisorio” e “repubblica semi presidenziale”, aumentano la probabilità di osservare un legame tra due attori nelle reti dei conflitti, al netto di tutte le altre diadi. Estendendo l'esempio precedente, calcoliamo la probabilità di legame tra un paese ebraico con repubblica parlamentare *i* e uno cristiano con una repubblica semi presidenziale *j*; avremo:

Modello sui conflitti

$$\begin{aligned} \log_odds P(Y_{ij} = 1 \mid y_{ij}^c) &= \theta \times 1 + \theta_{\text{Ebraismo}(i)} + \theta_{\text{Cristianesimo}(j)} + \\ &\quad + \theta_{\text{Repubblica_parlamentare}(i)} + \theta_{\text{Repubblica_semi_presidenziale}(j)} \\ &= -2.69 + 1.91 - 0.36 + 0.60 + 0.33 = -0.21 \\ P(Y_{ij} = 1 \mid y_{ij}^c) &= 0.447 \end{aligned}$$

Modello sulle cooperazioni:

$$\begin{aligned} \log_odds P(Y_{ij} = 1 \mid y_{ij}^c) &= \theta \times 1 + \theta_{\text{Ebraismo}(i)} + \theta_{\text{Cristianesimo}(j)} \\ &\quad + \theta_{\text{Repubblica_parlamentare}(i)} + \theta_{\text{Repubblica_semi_presidenziale}(j)} \\ &= -0.95 + 1.39 - 0.42 + 0.21 + 0.11 = 0.34 \\ P(Y_{ij} = 1 \mid y_{ij}^c) &= 0.584 \end{aligned}$$

Il terzo modello in figura in tabella 3.1 e 3.2 contiene tra i suoi termini esplicativi anche quelli relativi alle variabili numeriche: *GDP*, *HDI*, *tasso di fecondità* e *densità della popolazione*. Dalle stime dei coefficienti elevate si osserva che l'effetto dell'indice di sviluppo umano sia importante all'esplicazione della probabilità di interazione tra i paesi, mentre non è lo stesso per il tasso di fecondità e la densità di popolazione che possiedono coefficienti prossimi a zero. Di seguito si riporta la stima della probabilità dei due paesi precedenti, con un tasso di crescita del PIL di 2 e 3.6 rispettivamente, un indice di sviluppo umano di 0.915 e 0.821, una densità della popolazione di 423 e 8.82 persone per chilometro quadrato e un numero medio di figli per donna di 2.9 e 1.52.

Per il modello sui conflitti, vale che:

$$\begin{aligned}
\log_odds P(Y_{ij} = 1 \mid y_{ij}^c) &= \theta \times 1 + \theta_{\text{Ebraismo}(i)} + \theta_{\text{Cristianesimo}(j)} \\
&+ \theta_{\text{Repubblica_parlamentare}(i)} + \theta_{\text{Repubblica_semi_presidenziale}(j)} \\
&+ \theta_{\text{GDP}} \times \text{GDP}(i) + \theta_{\text{GDP}} \times \text{GDP}(j) \\
&+ \theta_{\text{HDI}} \times \text{HDI}(i) + \theta_{\text{HDI}} \times \text{HDI}(j) \\
&+ \theta_{\text{Densità_popolazione}} \times \text{Densità_popolazione}(i) \\
&+ \theta_{\text{Densità_popolazione}} \times \text{Densità_popolazione}(j) \\
&+ \theta_{\text{Tasso_fecondità}} \times \text{Tasso_fecondità}(i) \\
&+ \theta_{\text{Tasso_fecondità}} \times \text{Tasso_fecondità}(j) \\
&= -11.22 + 0.34 + 1.18 + -0.34 + 0.39 - 0.02 \times 2 \\
&\quad - 0.02 \times 3.6 + 4.85 \times 0.915 + 4.85 \times 0.821 \\
&\quad + 0.00 \times 423 + 0.00 \times 8.82 + 0.28 \times 2.917294 \\
&\quad + 0.28 \times 1.521507 \\
&= -0.1 \\
P(Y_{ij} = 1 \mid y_{ij}^c) &= 0.475
\end{aligned}$$

Per il modello sulle cooperazioni abbiamo:

$$\begin{aligned}
\log_odds P(Y_{ij} = 1 \mid y_{ij}^c) &= \theta \times 1 + \theta_{\text{Ebraismo}(i)} + \theta_{\text{Cristianesimo}(j)} \\
&+ \theta_{\text{Repubblica_parlamentare}(i)} + \theta_{\text{Repubblica_semi_presidenziale}(j)} \\
&+ \theta_{\text{GDP}} \times \text{GDP}(i) + \theta_{\text{GDP}} \times \text{GDP}(j) \\
&+ \theta_{\text{HDI}} \times \text{HDI}(i) + \theta_{\text{HDI}} \times \text{HDI}(j) \\
&+ \theta_{\text{Densità_popolazione}} \times \text{Densità_popolazione}(i) \\
&+ \theta_{\text{Densità_popolazione}} \times \text{Densità_popolazione}(j) \\
&+ \theta_{\text{Tasso_fecondità}} \times \text{Tasso_fecondità}(i) \\
&+ \theta_{\text{Tasso_fecondità}} \times \text{Tasso_fecondità}(j) \\
&= -9.96 + 0.47 - 0.11 - 0.29 + 0.22 + 0.02 \times 2 \\
&\quad + 0.02 \times 3.6 + 5.02 \times 0.915 + 5.02 \times 0.821 \\
&\quad + 0.00 \times 423 + 0.00 \times 8.82 + 0.32 \times 2.917294 \\
&\quad + 0.32 \times 1.521507 \\
&= 0.577 \\
P(Y_{ij} = 1 \mid y_{ij}^c) &= 0.640
\end{aligned}$$

Infine, l'ultimo modello contiene il termine per il conteggio dei rapporti reciproci, *mutual*. Sia per le cooperazione che per i conflitti il coeffi-

ciente risulta significativo e con stime alte, la probabilità di osservare un arco biunivoco tra due attori con le precedenti caratteristiche è significativamente più alta:

Modello sui conflitti

$$\begin{aligned}
 \log_odds P(Y_{ij} = 1 \mid y_{ij}^c) &= \theta \times 1 + \theta_{\text{Ebraismo}(i)} + \theta_{\text{Cristianesimo}(j)} \\
 &+ \theta_{\text{Repubblica_parlamentare}(i)} + \theta_{\text{Repubblica_semi_presidenziale}(j)} \\
 &+ \theta_{\text{GDP}} \times \text{GDP}(i) + \theta_{\text{GDP}} \times \text{GDP}(j) \\
 &+ \theta_{\text{HDI}} \times \text{HDI}(i) + \theta_{\text{HDI}} \times \text{HDI}(j) \\
 &+ \theta_{\text{Densità_popolazione}} \times \text{Densità_popolazione}(i) \\
 &+ \theta_{\text{Densità_popolazione}} \times \text{Densità_popolazione}(j) \\
 &+ \theta_{\text{Tasso_fecondità}} \times \text{Tasso_fecondità}(i) \\
 &+ \theta_{\text{Tasso_fecondità}} \times \text{Tasso_fecondità}(j) \\
 &+ \text{Mutual} \\
 &= -8.7 + 0.21 + 0.74 - 0.22 + 0.25 - 0.01 \times 2 \\
 &\quad - 0.01 \times 3.6 + 3.08 \times 0.915 + 3.08 \times 0.821 \\
 &\quad + 0.00 \times 423 + 0.00 \times 8.82 + 0.18 \times 2.917294 \\
 &\quad + 0.18 \times 1.521507 + 3.25 \\
 &= 1.619 \\
 P(Y_{ij} = 1 \mid y_{ij}^c) &= 0.835
 \end{aligned}$$

Modello sulle cooperazioni

$$\begin{aligned}
\log_odds P(Y_{ij} = 1 \mid y_{ij}^c) &= \theta \times 1 + \theta_{\text{Ebraismo}(i)} + \theta_{\text{Cristianesimo}(j)} \\
&+ \theta_{\text{Repubblica_parlamentare}(i)} + \theta_{\text{Repubblica_semi_presidenziale}(j)} \\
&+ \theta_{\text{GDP}} \times \text{GDP}(i) + \theta_{\text{GDP}} \times \text{GDP}(j) \\
&+ \theta_{\text{HDI}} \times \text{HDI}(i) + \theta_{\text{HDI}} \times \text{HDI}(j) \\
&+ \theta_{\text{Densità_popolazione}} \times \text{Densità_popolazione}(i) \\
&+ \theta_{\text{Densità_popolazione}} \times \text{Densità_popolazione}(j) \\
&+ \theta_{\text{Tasso_fecondità}} \times \text{Tasso_fecondità}(i) \\
&+ \theta_{\text{Tasso_fecondità}} \times \text{Tasso_fecondità}(j) \\
&+ \text{Mutual} \\
&= -7.87 + 0.26 - 0.06 - 0.16 + 0.12 + 0.01 \times 2 \\
&\quad + 0.01 \times 3.6 + 2.8 \times 0.915 + 2.8 \times 0.821 \\
&\quad - 0.00 \times 423 + -0.00 \times 8.82 + 0.18 \times 2.917294 \\
&\quad + 0.18 \times 1.521507 + 4.61 \\
&= 2.616 \\
P(Y_{ij} = 1 \mid y_{ij}^c) &= 0.932
\end{aligned}$$

Gli ERGM implementati hanno mostrato che alcune caratteristiche dei paesi come l'indice di sviluppo umano o la forma di governo hanno assunto importanza all'esplicazione della probabilità di formazione di legami tra gli stati. Altri fattori, come la densità della popolazione e il tasso di fecondità, hanno assunto invece un ruolo meno importante. Il modello che include il termine di adiacenza dipendente, ha fornito una probabilità di creazione dei legami, al netto di tutti gli altri, sicuramente rilevante. Questo fattore rivela che all'interno delle reti sono presenti strutture di dipendenza cruciali da analizzare, e utili a spiegare le dinamiche internazionali. I modelli ERGM forniscono un importante mezzo per dimostrare l'importanza di fattori sociali alla creazione di legami.

Per valutare l'adattamento dei modelli stimati ci si è basati sui residui di Pearson tra valori adattati e valori osservati delle diverse statistiche incluse nel modello. Le diagnostiche presentano residui relativamente bassi ma con alcune divergenze tra valori adattati e osservati in determinati mesi. Complessivamente, anche dall'indice di Akaike e il

valore della log-verosimiglianza nelle tabelle in figura 3.1 e 3.2, il quarto modello è il migliore sia nel caso delle cooperazioni che nel caso dei conflitti. Nel passo successivo della fase di diagnostiche sono stati adattati dei modelli lineari sui residui: per ogni statistica del modello sono stati utilizzati i relativi residui di Pearson come variabile dipendente di una regressione lineare con regressore una variabile temporale, costruita tramite una variabile numerica con valori da 1 a 12. In questo modo si è potuto controllare se la variabile temporale influisse sui valori adattati del modello alle dodici reti, per ogni statistica inserita nel modello. È interessante osservare che nel caso della regressione dei residui della statistica di rete “edges” sul tempo, il relativo coefficiente risulta significativo, a supportare l’ipotesi della presenza di un effetto di stagionalità nella mole dei legami tra i paesi durante l’anno. In generale, dalle regressioni dei residui sul tempo si intuisce che un modello che incorpori un fattore di tempo possa leggermente migliorare l’adattamento. Ma complessivamente, sarebbe meglio integrare statistiche aggiuntive diadiche dipendenti per adattare meglio il modello ai dati. Questo porta però a modelli più complicati da stimare e quindi molto più pesanti computazionalmente. Una soluzione alternativa ai limiti che presentano gli ERGM è stata utilizzare un altro tipo di modello che rappresenta un diverso approccio. Si tratta dei Modelli a Effetti Additivi e Moltiplicativi che vengono presentati nella seguente sezione.

3.3 Modelli a Effetti Additivi e Moltiplicativi

L’analisi di una matrice di adiacenza, in questo caso anche detta sociomatrice, procede spesso nel contesto di uno o più modelli statistici, con i quali un analista dei dati può valutare teorie concorrenti sulla formazione delle reti. Oltre agli ERGM, in letteratura ci sono tanti approcci a metodologie statistiche che si sviluppano attorno a decomposizioni matriciali della sociomatrice, analisi fattoriale e modelli a effetti misti lineari e lineari generalizzati.

Nelle seguenti sezioni viene affrontato l’uso dei modelli a effetti additivi e moltiplicativi che consentono di quantificare similarità dei nodi in termini di caratteristiche latenti non osservate (Hoff, 2021).

3.3.1 Costruzione del modello AME

L'utilizzo dei modelli a effetti additivi e moltiplicativi nasce dall'analisi di alcuni elementi che in altri modelli non vengono considerati. Le sociomatrici, in primo luogo, mostrano per prime alcune particolarità statistiche: per esempio, valori della variabile diadica in una riga della sociomatrice possono essere correlati con quelli di un'altra riga, nel senso che valori alti e valori bassi non sono equamente distribuiti tra le righe, risultando in una sostanziale eterogeneità delle medie di riga della sociomatrice. Tale tipo di eterogeneità può essere spiegata tenendo a mente che le relazioni all'interno di una particolare riga condividono lo stesso "mittente" o indice di riga. Se il mittente i_1 è più "socievole" del mittente i_2 , ci aspetteremmo che le realizzazioni nella riga i_1 siano mediamente maggiori rispetto a quelli nella riga i_2 . In questo modo, l'eterogeneità dei nodi in termini di socievolezza contribuisce a una varianza tra le medie delle righe della sociomatrice. Allo stesso modo ma in senso contrario, l'eterogeneità dei nodi in termini di "popolarità", cioè la tendenza a ricevere più legami, contribuisce alla varianza tra le colonne delle medie delle colonne. Un approccio classico per valutare l'eterogeneità tra le righe e tra le colonne in una matrice di dati è la decomposizione ANOVA. Un modello statistico basato sulla decomposizione ANOVA postula che la variabilità delle Y_{ij} attorno a una media complessiva μ possa essere rappresentata da effetti additivi di riga e di colonna:

$$Y_{ij} = \mu + a_i + b_j + \epsilon_{i,j} \quad (3.9)$$

In questo modello, l'eterogeneità tra a_i e b_j dà origine all'eterogeneità osservata nelle medie delle righe e delle colonne della sociomatrice, rispettivamente. Sebbene sia semplice da implementare, un'analisi classica ANOVA ignora una caratteristica fondamentale dei dati diadici diretti: ogni nodo appare nel dataset sia come mittente che come ricevente di relazioni, o equivalentemente, le etichette di riga e colonna della matrice di dati si riferiscono allo stesso insieme di nodi. Nel contesto del modello ANOVA, ciò significa che ogni nodo i ha due effetti additivi: un effetto di riga a_i e un effetto di colonna b_i . Poiché ogni coppia di effetti (a_i, b_i) condivide un nodo, ci si potrebbe aspettare una correlazione tra i vettori (a_1, \dots, a_n) e (b_1, \dots, b_n) . Inoltre, ogni diade $\{i, j\}$ ha due risultati, Y_{ij} e Y_{ji} . Pertanto, dovrebbe essere considerata la possibilità che Y_{ij} e Y_{ji} siano correlati. Il modello ANOVA descritto riesce a quantificare la variazione tra le righe, la variazione tra le colonne e quella residua ma di contro, non quantifica la correlazione

mittente-ricevente che sembra evidente da molti esempi in letteratura. Un differente modello che tiene conto di quest'ultimo aspetto fu introdotto nell'ambito della psicologia sociale da Warner, Kenny e Stoto (1979). Rappresenta un modello più completo, si chiama “modello per le relazioni sociali” (SRM, social relation model), è strutturato tramite effetti casuali come quello in equazione (3.9) ma con l'assunzione addizionale che gli a_i , b_j e ϵ_{ij} siano variabili casuali con media zero, per le quali:

$$\text{Var} \left(\begin{bmatrix} a_i \\ b_i \end{bmatrix} \right) = \Sigma = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}$$

$$\text{Var} \left(\begin{bmatrix} \epsilon_{ij} \\ \epsilon_{ji} \end{bmatrix} \right) = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Un'estensione di tale modello che si avvicina al caso utile per lo studio dell'elaborato aggiunge alla struttura di covarianza del modello SRM una parte di regressione lineare che permette di quantificare l'associazione tra una particolare variabile diadica e altre variabili nodali caratterizzanti degli attori, formulato come:

$$Y_{ij} = \beta^\top x_{ij} + a_i + b_j + \epsilon_{ij} \quad (3.10)$$

dove la variabile x_{ij} è un vettore p -dimensionale di regressori e β^\top è un vettore di coefficienti di regressione da stimare. Il vettore x_{ij} può contenere variabili specifiche per i nodi o coppie di nodi. Ad esempio, potremmo avere $x_{ij} = (x_{ri}, x_{cj}, x_{dij})$ dove x_{ri} è un vettore di caratteristiche del nodo i come mittente o oggetto di riga, x_{cj} è un vettore di caratteristiche del nodo j come ricevente o oggetto di colonna e x_{dij} è un vettore di caratteristiche della coppia ordinata (ij) .

Seppur costruiti con presupposti ragionevoli, i modelli precedenti mostrano casi di carenza di adattamento in diverse applicazioni in cui hanno avuto luogo. In particolare è frequente che le reti reali mostrino pattern di dipendenza più complicati come possono essere triangolazioni o formazioni di cluster. Anche nel caso in studio, il miglioramento dell'adattamento degli ERGM attraverso termini diadi dipendenti ha suggerito che l'approccio allo studio delle reti deve comprendere all'interno del modello componenti che catturino schemi di dipendenza tra le diadi più articolati.

Per modellare schemi di rete di ordine superiore e cercare di catturare forme di dipendenza più intrinseche nei dati, bisogna andare oltre gli effetti specifici dei nodi che si combinano in modo additivo. Una soluzione è includere ulteriori effetti nodali nel modello che si combinano in modo moltiplicativo. Ad esempio, per variabili diadiche binarie e di conteggio, Hoff (2005) ha proposto di modellare Y_{ij} come una funzione di $\beta^\top x_{ij} + u_i u_j + a_i + b_j$, dove ogni u_i è un vettore a bassa dimensionalità di parametri specifici del nodo, o caratteristiche latenti. Un modello ad effetti moltiplicativi per dati diadici diretti è ottenuto sostituendo $u_i u_j$ con $u_i v_j$, in modo che u_i rappresenti caratteristiche latenti del nodo i come mittente di relazioni, e v_i rappresenti le caratteristiche del nodo i come ricevente. Combinando un termine moltiplicativo del genere con il modello in equazione (3.10), si ottiene il seguente modello ad effetti additivi e moltiplicativi per dati diadici, chiamato AME (additive and multiplicative effects):

$$Y_{ij} = \beta^\top x_{ij} + u_i^\top v_j + a_i + b_j + \epsilon_{i,j} \quad (3.11)$$

$$\{(\epsilon_{ij}, \epsilon_{ji}) : i < j\} \sim \text{i.i.d. } \mathcal{N}_2 \left(0, \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

Nello specifico, ci si riferisce al modello in equazione (3.11) come Modello AME Gaussiano, poichè con le assunzioni precedenti, i dati osservati sono considerati condizionatamente gaussiani. Gli effetti moltiplicativi introdotti tramite il modello AME possono essere visti come un mezzo per quantificare schemi di dipendenza globale di ordine superiore in una matrice di adiacenza, oppure possono anche essere interpretati come rappresentativi di variabili di regressione omesse o come la rivelazione di una struttura di gruppo tra i nodi.

L'ultimo passo per rendere il modello più adatto possibile ai dati in studio è quello di affrontare la discordanza tra la natura dei dati diadici e l'assunzione di normalità degli errori. A questo assunto viene in aiuto un'ulteriore estensione del modello AME gaussiano che permette di quantificare l'associazione tra la variabile diadica e altre variabili sulla base di una rappresentazione latente del modello di regressione probit. Si dispone quindi di un modello AME per le variabili latenti s_{ij} nella sua forma finale:

$$\begin{aligned} s_{ij} &= \beta^\top x_{ij} + u_i^\top v_j + a_i + b_j + \epsilon_{i,j} \\ y_{i,j} &= \mathbb{1}[s_{ij} > 0] \end{aligned} \quad (3.12)$$

3.3.2 Applicazione del modello AME

Ottenuta la forma finale del Modello a Effetti Additivi e Moltiplicativi, può ora essere applicato ragionevolmente ai dati in studio attraverso l'estensione multistrato che permette l'adattamento del modello a dati relazionali replicati nel tempo. L'estensione multistrato mantiene la stessa struttura di covarianza per gli effetti additivi del modello AME originale, così come mantiene la correlazione tra l'errore associato alla relazione tra i nodi i e j e l'errore della relazione opposta. Una nuova assunzione del modello multistrato è che gli errori provenienti da diadi diverse o da punti temporali diversi siano considerati indipendenti. Questo significa che, sebbene ci sia dipendenza all'interno della stessa diade in un dato momento, si assume che non ci siano correlazioni tra le relazioni di coppie diverse di nodi o tra le stesse coppie in momenti diversi. Il modello AME multistrato considera quindi tutte le osservazioni, che siano relazioni tra nodi in momenti diversi o relazioni tra nodi diversi, come campioni indipendenti provenienti da una stessa distribuzione AME.

Il modello è stato implementato in modo da avere due coefficienti regressivi per ognuna delle stesse variabili nodali usate negli ERGM, relativi al loro effetto di riga e di colonna. Oltre ai regressori, il modello contiene un effetto additivo di riga a_i e uno di colonna b_j per ogni soggetto, e due dimensioni per ognuno dei due effetti moltiplicativi u_i e v_i . In appendice, le tabelle 3.3 e 3.4 riportano le stime dei coefficienti di regressione stimati per i modelli sui conflitti e sulle cooperazioni. I coefficienti dei regressori rispecchiano alcune caratteristiche osservate con gli ERGM: il tasso di crescita del PIL, sia per i conflitti che per le cooperazioni ha una bassa influenza sulla creazione di un legame tra paesi, mentre il coefficiente relativo all'indice di sviluppo umano gioca un ruolo importante nello stabilire legami tra soggetti, in questo caso maggiore nel modello sulle cooperazioni, mentre negli ERGM maggiore per i conflitti. La densità della popolazione, come negli ERGM, è influente alla creazione dei legami, sia nel conflitti che nelle cooperazioni. I coefficienti relativi alle modalità delle variabili sulla forma di governo e sulla religione, così come i coefficienti riferiti al tasso di fecondità, sono più bassi rispetto agli stessi nei modelli ERGM, a significare che nei modelli AME queste variabili perdono di importanza nell'esplicazione dei legami nelle reti. In figura 4.1 e 4.2 vengono riportate le stime degli effetti additivi di riga e di colonna del modello sui conflitti per ogni paese, mentre in figura 4.3 e 4.4 quelli del modello sulle coope-

razioni. Gli effetti additivi di riga, relativi ai conflitti, denotano paesi come Stati Uniti, Russia, Pakistan, Regno Unito e Cina come soggetti altamente “sociali”, cioè attivi nell’innescare relazioni di conflitto. Hanno invece un comportamento contrario paesi come Lituania, Bielorussia, Finlandia e Svezia che sembrano respingere un atteggiamento ostile nei confronti di altri paesi. Stati Uniti, Russia e Regno Unito spiccano anche con i loro effetti additivi di colonna, cioè gli effetti che li denotano come soggetti “popolari” nel ricevere relazioni di conflitto, mentre mantengono un ruolo neutrale Finlandia, Lituania, Bielorussia e Svezia. Per quanto riguarda il modello sulle cooperazioni, gli effetti additivi di riga e di colonna riportano come maggiormente attivi nello stabilire legami cooperativi e nel riceverli paesi come Cina, Germania, Canada, Pakistan, Giappone e Ucraina, oltre ai paesi altamente attivi anche nei conflitti. Anche per le cooperazioni, Bielorussia, Finlandia, Lituania e Svezia sembrano mantenere un profilo defilato nello stabilire rapporti con altri paesi.

Del modello AME l’aspetto più attraente è sicuramente la stima degli effetti moltiplicativi che catturano informazioni non osservate degli attori sia come mittenti sia come riceventi. Scegliere due dimensioni per ognuna delle variabili latenti ha permesso di generare un grafico bidimensionale in cui le informazioni non osservate per ogni soggetto sono le coordinate di questo spazio. Così facendo si può ricercare coerenza o mancata coerenza di posizionamento per uno stesso soggetto nei panni di mittente (u_i) e nei panni di destinatario (v_i); si possono identificare pattern di stati che hanno caratteristiche latenti simili quindi connessi da uno specifico ruolo nel panorama mondiale oppure ricercare contrapposizioni tra attori localizzati in zone opposte del grafico. In figura 3.1 si riporta l’immagine relativa alle coordinate latenti stimate dal modello sui conflitti mentre in figura 3.2 quelle relative al modello sulle cooperazioni.

Per quanto riguarda i conflitti si riconoscono alcuni gruppi di stati che si conformano nell’area di posizionamento. Si possono stabilire tre grandi blocchi di influenza e omologazione: sulla destra ci sono principalmente paesi occidentali rappresentati per lo più da Gran Bretagna, Stati Uniti, Francia e Germania che presentano quindi informazioni non osservate molto simili. Da ponte verso gli altri due blocchi di stati che si possono identificare come mondo asiatico da una parte e medio oriente dall’altra ci sono rispettivamente Cina e Russia che presentano molto chiaramente caratteristiche latenti in continuità con il blocco occidentale. La Cina, girando in senso antiorario, anticipa paesi come

Giappone, Taiwan, Filippine, Korea, Vietnam e India che localizzano ovviamente l'area asiatica. Nell'altro versante invece, girando in senso orario, La Russia fa da ponte con caratteristiche medio orientali e nord africane per lasciare spazio, nella parte inferiore del grafico a paesi come Israele, Iran, Egitto, Siria, Iraq, Turchia e Arabia Saudita ossia quella che viene chiamata "area mena" (Middle East North Africa). Per molti paesi esiste una corrispondenza nelle coordinate sul piano di mittenza e su quello di destinazione che denota coerenza di magnitudine nel ricevere e nel rispondere a un conflitto o a un'offesa. Questo non avviene per alcuni paesi come Italia, Kenya e Ucraina che sembrano essere caratterizzati per lo più come paesi destinatari di conflitti. Infine, è utile sottolineare le contrapposizioni apparse tra blocchi di stati; oltre a quella già citata tra mondo asiatico e mondo medio-orientale c'è sicuramente un contrasto di caratteristiche non osservate tra paesi occidentali nella parte destra del grafico e paesi africani che nei mesi successivi a gennaio appaiono nella parte sinistra del grafico.

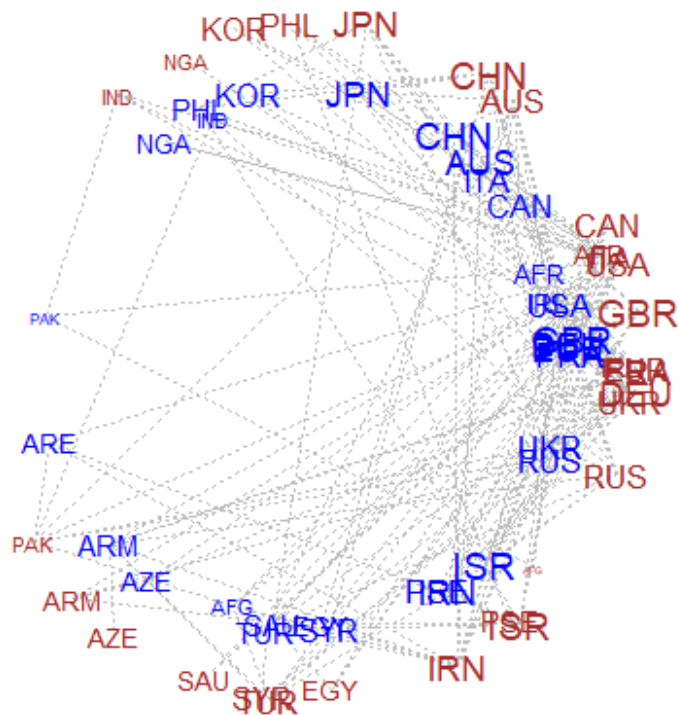


Figura 3.1: Modello AME sui conflitti di gennaio

Il corrispettivo grafico delle cooperazioni conferma l'esistenza di blocchi di stati con carattere comune. Sono particolarmente evidenti i gruppi sul lato destro e sinistro del grafico: a destra si ritrovano paesi asiatici come Giappone, Filippine, Korea e nei mesi successivi anche Vietnam e Indonesia, insieme al Messico che anche precedentemente era accomunato al gruppo asiatico; nel blocco a sinistra, oltre alla Russia ci sono Turchia, Iran, Siria, Iraq e Afghanistan che rappresentano il blocco medio orientale. Tra queste due parti e il terzo raggruppamento, posto nella parte inferiore del grafico, vi sono alcuni paesi di intermezzo come precedentemente erano Russia e Cina. Ora Canada, Australia e Nuova Zelanda fungono da interconnessione con il blocco asiatico e, Arabia Saudita, Emirati Arabi Uniti e Israele con il gruppo medio orientale. Questo terzo gruppo di paesi posizionato nella parte inferiore del gra-

fico, è più eterogeneo poiché oltre a paesi occidentali come Germania, Francia, Italia e Gran Bretagna, si trovano anche India e Nigeria. Per le cooperazioni sembra esserci maggiore corrispondenza tra aiuti inviati e supporto ricevuto il che potrebbe indicare che più facilmente la cooperazione è un'interazione biunivoca rispetto alle relazioni conflittuali. Anche in questo caso appaiono evidenti i gruppi di appartenenza che si sono venuti a creare secondo le variabili latenti stimate. Pare anche evidente, da entrambi i grafici, che le variabili latenti cioè le informazioni non osservate che portano due soggetti a collocarsi vicini nei grafici 3.1 e 3.2 siano caratteristiche proprie di parti diverse del mondo che contraddistinguono il “modus operandi” di gruppi di nazioni al loro interno omogenei. In quest'ottica potrebbe essere di grande potere esplicativo l'informazione legata alla geolocalizzazione di ogni soggetto oltre che a fattori culturali e storici che legano gli stati di una stessa regione.

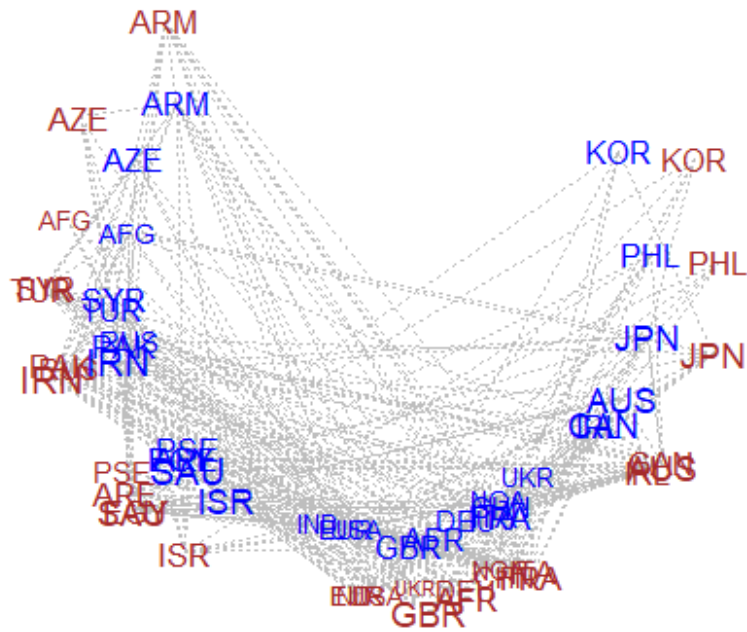


Figura 3.2: Modello AME sulle cooperazioni di gennaio

3.4 Confronto dei modelli

I modelli considerati costituiscono due diversi approcci allo studio delle reti sociali. Gli ERGM utilizzati nella tesi hanno permesso di descrivere pattern globali all'interno dei grafi attraverso un numero ristretto di statistiche sufficienti: sono state utilizzate sei covariate relative ad aspetti strutturali dei paesi e due variabili diadi dipendenti inerenti i legami formatisi all'interno delle reti. Con questo approccio si è potuto stabilire quali elementi influiscono sulla probabilità di formazione delle diadi e si è potuto ottenere un quadro di come fattori sociali e culturali influenzino le interazioni tra i paesi del mondo con un'interpretazione diretta dei parametri, rendendo gli ERGM dei modelli trasparenti

nel capire quali caratteristiche della rete sono modellate. Nonostante l'inclusione di regressori appropriati all'interno dei modelli, gli ERGM hanno mostrato dei limiti computazionali nel modellare forme di dipendenza tra le diadi che fossero più complesse rispetto a quelle inserite, come potrebbero essere forme di raggruppamento o triangolazioni di rapporti.

La stima degli ERGM attraverso la massimizzazione della verosimiglianza completa sarebbe troppo complicata da calcolare poiché richiede di considerare tutte le possibili configurazioni di rete. Viene quindi utilizzata la pseudo verosimiglianza come approssimazione della verosimiglianza, ottenuta come prodotto delle probabilità di ogni legame condizionato agli altri legami della rete. Quando sono inclusi termini diadi dipendenti nel modello, l'approccio di stima tramite la pseudo verosimiglianza perde di senso poiché, per costruzione, non considererebbe la dipendenza reciproca dei legami. In questo caso il metodo di stima cambia, e la verosimiglianza viene approssimata attraverso l'utilizzo di un algoritmo MCMC con il metodo Metropolis Hastings come campionario. Attraverso l'algoritmo MCMC vengono simulate possibili configurazioni di reti delle quali si calcolano le statistiche inserite nel modello. A ogni iterazione dell'algoritmo le statistiche calcolate per ogni simulazione vengono mediate con le precedenti ed attraverso una funzione obiettivo che minimizza la differenza quadratica tra le statistiche osservate e quelle medie stimate vengono aggiornati i parametri del modello fino a convergenza.

Il modello ERGM più adatto è stato scelto in funzione dell'indice di Akaike che ha guidato le analisi a partire da modelli insufficienti con AIC molto alto a modelli più completi con una quota di AIC significativamente minore.

Al contrario, i modelli AME utilizzati hanno permesso di caratterizzare il livello di attività dei paesi attraverso gli effetti additivi e analizzare schemi di dipendenza della sociomatrice più complessi, attraverso una matrice a bassa dimensione UV di variabili latenti. Grazie a queste ultime si sono potuti osservare schemi di somiglianza tra i paesi e ipotizzare ulteriori fattori che influenzano gli equilibri mondiali come possono essere la collocazione geografica e l'etnia di appartenenza. Nonostante i vantaggi, gli AME forniscono dei parametri di più difficile interpretazione e se ne favorisce l'utilizzo in contesti dove la matrice di adiacenza è molto grande e la riduzione dimensionale è cruciale. I modelli AME, costruiti per modellare dipendenze tra i dati, introducono una complessità significativa nella verosimiglianza del modello che

diventa troppo complicata da calcolare. Vengono perciò stimati attraverso l'algoritmo Gibbs sampling che procede simulando iterativamente i valori delle quantità sconosciute dalle loro distribuzioni condizionate, generando così una catena di Markov con una distribuzione stazionaria uguale alla distribuzione a posteriori.

In appendice si osservano i grafici riferiti alla traccia delle catene di Markov per la stima dei parametri del modello AME. Le tracce dei parametri sono considerabili casuali e fluttuano attorno a un valore stabile. I boxplot seguenti indicano che alcune caratteristiche delle reti, come la variabilità tra le medie di riga e le medie di colonna, sono catturate adeguatamente dai modelli. Vi è invece più variabilità nelle dipendenze cicliche e tra diadi che riflettono relazioni strutturali più complesse tra i nodi della rete che i modelli riescono a catturare meno adeguatamente.

Infine, i due modelli, seppur di diversa natura e costruzione, sono stati confrontati direttamente costruendo una misura della bontà di adattamento. Delle reti adattate dai modelli AME, e quelle simulate dai modelli ERGM, è stata calcolata la variabilità di riga e confrontata con la stessa calcolata sulle reti osservate. La somma del valore assoluto delle differenze di variabilità di riga tra le reti adattate e quelle osservate fornisce un indice che stabilisce quanto le reti stimate rappresentino quelle reali. Dai modelli AME si è ottenuto un indice della bontà di adattamento di 247.172 per il modello sui conflitti e 592.540 per il modello sulle cooperazioni, mentre gli ERGM hanno rivelato un adattamento peggiore con una differenza di variabilità di 716.735 per i conflitti e 918.725 per le cooperazioni. Dalla costruzione di questo indice si può affermare che i modelli AME forniscono uno strumento più accurato nel rappresentare le reti in studio. Complessivamente, nella tesi sono risultati essere più utili, ai fini esplicativi e rappresentativi delle reti osservate, i modelli AME poiché oltre alla migliore bontà di adattamento, permettono di intuire caratteristiche trasversali alla rete multistrato attraverso gli effetti additivi e le variabili latenti, portando l'accuratezza dello studio a un livello superiore.

Conclusioni

La tesi fornisce un metodo di ricerca e sviluppo nell'ambito dello studio delle relazioni internazionali. Dalle informazioni estratte dai titoli dei quotidiani si è potuto collezionare un insieme di dati relazionali dai quali si sono tratte importanti informazioni per caratterizzare la dipendenza tra gli stati coinvolti nello studio. Con la modellazione, gli ERGM hanno mostrato quali fossero alcuni tratti essenziali della descrizione di un paese e della sua rete di rapporti ma hanno anche mostrato l'incapacità di catturare dinamiche di reciproca dipendenza più complesse. Approfondire ciò che gli ERGM non sono riusciti a spiegare ha significato utilizzare un modello con un diverso approccio che raccogliesse la variabilità non osservata fra gli attori attraverso delle variabili latenti, per mezzo dell'utilizzo dei Modelli con Effetti Additivi e Moltiplicativi. Tramite questo secondo strumento si sono potute raccogliere informazioni riguardanti i raggruppamenti di paesi che condividono caratteristiche comuni non osservate. L'aspetto più interessante dell'approccio a effetti additivi e moltiplicativi è stato proprio poter immaginare quali fossero le caratteristiche non osservate: i raggruppamenti che abbiamo visto anche nelle figure 3.1 e 3.2 suggeriscono che le variabili latenti hanno sicuramente a che fare con la zona geografica, la storia e la cultura che una determinata zona del globo ha assorbito nel corso del tempo e che tutt'ora conserva come proprio retaggio. Grazie a queste osservazioni, lo studio condotto potrebbe essere proseguito e ampliato per esempio integrando i fattori appena citati o aggiungendo strutture di correlazione tra paesi nella stessa area geografica.

Appendice

Tabella 3.3: Stime dei coefficienti di regressione del modello sui conflitti

| Coefficienti di regressione | pmean | psd | z-stat | p-val |
|------------------------------------|--------------|------------|---------------|--------------|
| intercept | -0.950 | 1.221 | -0.778 | 0.437 |
| GDP.row | 0.001 | 0.013 | 0.101 | 0.919 |
| HDI.row | 0.625 | 0.654 | 0.956 | 0.339 |
| Densità della popolazione.row | 0.000 | 0.000 | -0.160 | 0.873 |
| Giunte militari.row | 0.033 | 0.377 | 0.087 | 0.931 |
| Governi provvisori.row | 0.110 | 0.307 | 0.358 | 0.720 |
| Monarchie assolute.row | -0.086 | 0.275 | -0.314 | 0.754 |
| Monarchie costituzionali.row | -0.031 | 0.136 | -0.229 | 0.819 |
| Monarchie semi-costituzionali.row | -0.108 | 0.267 | -0.404 | 0.686 |
| Repubbliche parlamentari.row | -0.017 | 0.144 | -0.116 | 0.908 |
| Repubbliche presidenziali.row | 0.063 | 0.136 | 0.461 | 0.645 |
| Repubbliche semi-presidenziali.row | 0.047 | 0.140 | 0.332 | 0.740 |
| Stati a partito unico.row | 0.168 | 0.256 | 0.658 | 0.510 |
| Fecondità.row | 0.043 | 0.091 | 0.479 | 0.632 |
| Cristianesimo.row | 0.040 | 0.172 | 0.234 | 0.815 |
| Ebraismo.row | 0.123 | 0.311 | 0.396 | 0.692 |
| Induismo.row | 0.161 | 0.319 | 0.505 | 0.614 |
| Islam.row | 0.048 | 0.177 | 0.270 | 0.787 |
| Islam/Cristianesimo.row | -0.040 | 0.351 | -0.115 | 0.909 |
| Persone non religiose.row | -0.014 | 0.351 | -0.041 | 0.967 |
| GDP.col | -0.001 | 0.013 | -0.069 | 0.945 |
| HDI.col | 0.448 | 0.655 | 0.684 | 0.494 |
| Densità della popolazione.col | 0.000 | 0.000 | -0.165 | 0.869 |
| Giunte militari.col | 0.004 | 0.389 | 0.010 | 0.992 |
| Governi provvisori.col | 0.073 | 0.301 | 0.242 | 0.809 |
| Monarchie assolute.col | -0.043 | 0.273 | -0.158 | 0.874 |
| Monarchie costituzionali.col | -0.022 | 0.155 | -0.144 | 0.885 |
| Monarchie semi-costituzionali.col | -0.095 | 0.268 | -0.354 | 0.724 |
| Repubbliche parlamentari.col | -0.023 | 0.145 | -0.161 | 0.872 |
| Repubbliche presidenziali.col | 0.050 | 0.130 | 0.386 | 0.700 |
| Repubbliche semi-presidenziali.col | 0.053 | 0.141 | 0.377 | 0.707 |
| Stati a partito unico.col | 0.186 | 0.230 | 0.810 | 0.418 |
| Fecondità.col | 0.024 | 0.096 | 0.249 | 0.803 |
| Cristianesimo.col | 0.056 | 0.150 | 0.372 | 0.710 |
| Ebraismo.col | 0.212 | 0.340 | 0.625 | 0.532 |
| Induismo.col | 0.165 | 0.269 | 0.614 | 0.539 |
| Islam.col | 0.056 | 0.153 | 0.363 | 0.717 |
| Islam/Cristianesimo.col | -0.003 | 0.337 | -0.010 | 0.992 |
| Persone non religiose.col | -0.081 | 0.354 | -0.229 | 0.819 |

Tabella 3.4: Stime dei coefficienti di regressione del modello sulle cooperazioni

| Coefficienti di regressione | pmean | psd | z-stat | p-val |
|------------------------------------|--------------|------------|---------------|--------------|
| intercept | -1.319 | 1.445 | -0.913 | 0.361 |
| GDP.row | 0.005 | 0.015 | 0.307 | 0.759 |
| HDI.row | 0.939 | 0.741 | 1.268 | 0.205 |
| Densità della popolazione.row | 0.000 | 0.000 | -0.434 | 0.664 |
| Giunte militari.row | 0.004 | 0.424 | 0.010 | 0.992 |
| Governi provvisori.row | 0.110 | 0.343 | 0.322 | 0.748 |
| Monarchie assolute.row | -0.035 | 0.305 | -0.115 | 0.908 |
| Monarchie costituzionali.row | -0.032 | 0.152 | -0.212 | 0.832 |
| Monarchie semi-costituzionali.row | -0.101 | 0.301 | -0.335 | 0.738 |
| Repubbliche parlamentari.row | -0.040 | 0.161 | -0.251 | 0.802 |
| Repubbliche presidenziali.row | 0.025 | 0.151 | 0.164 | 0.870 |
| Repubbliche semi-presidenziali.row | 0.040 | 0.156 | 0.255 | 0.799 |
| Stati a partito unico.row | 0.234 | 0.285 | 0.822 | 0.411 |
| Fecondità.row | 0.061 | 0.103 | 0.591 | 0.555 |
| Cristianesimo.row | -0.018 | 0.190 | -0.096 | 0.923 |
| Ebraismo.row | 0.081 | 0.347 | 0.234 | 0.815 |
| Induismo.row | 0.163 | 0.360 | 0.453 | 0.650 |
| Islam.row | 0.019 | 0.197 | 0.095 | 0.924 |
| Islam/Cristianesimo.row | -0.060 | 0.395 | -0.153 | 0.879 |
| Persone non religiose.row | -0.146 | 0.393 | -0.372 | 0.710 |
| GDP.col | 0.003 | 0.015 | 0.226 | 0.821 |
| HDI.col | 0.846 | 0.745 | 1.136 | 0.256 |
| Densità della popolazione.col | 0.000 | 0.000 | -0.432 | 0.666 |
| Giunte militari.col | -0.015 | 0.439 | -0.034 | 0.973 |
| Governi provvisori.col | 0.096 | 0.336 | 0.286 | 0.775 |
| Monarchie assolute.col | -0.029 | 0.308 | -0.093 | 0.926 |
| Monarchie costituzionali.col | -0.040 | 0.171 | -0.231 | 0.817 |
| Monarchie semi-costituzionali.col | -0.100 | 0.303 | -0.329 | 0.742 |
| Repubbliche parlamentari.col | -0.048 | 0.164 | -0.295 | 0.768 |
| Repubbliche presidenziali.col | 0.021 | 0.145 | 0.146 | 0.884 |
| Repubbliche semi-presidenziali.col | 0.042 | 0.157 | 0.266 | 0.790 |
| Stati a partito unico.col | 0.241 | 0.258 | 0.936 | 0.349 |
| Fecondità.col | 0.046 | 0.108 | 0.422 | 0.673 |
| Cristianesimo.col | -0.010 | 0.170 | -0.060 | 0.952 |
| Ebraismo.col | 0.122 | 0.375 | 0.326 | 0.744 |
| Induismo.col | 0.173 | 0.317 | 0.546 | 0.585 |
| Islam.col | 0.031 | 0.175 | 0.177 | 0.859 |
| Islam/Cristianesimo.col | -0.034 | 0.382 | -0.089 | 0.929 |
| Persone non religiose.col | 71 -0.186 | 0.402 | -0.464 | 0.643 |

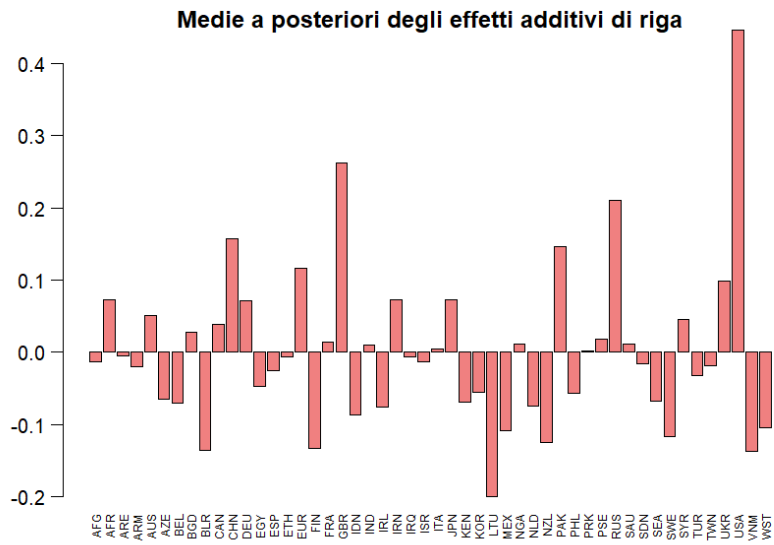


Figura 3.3: Effetti additivi di riga del modello sui conflitti

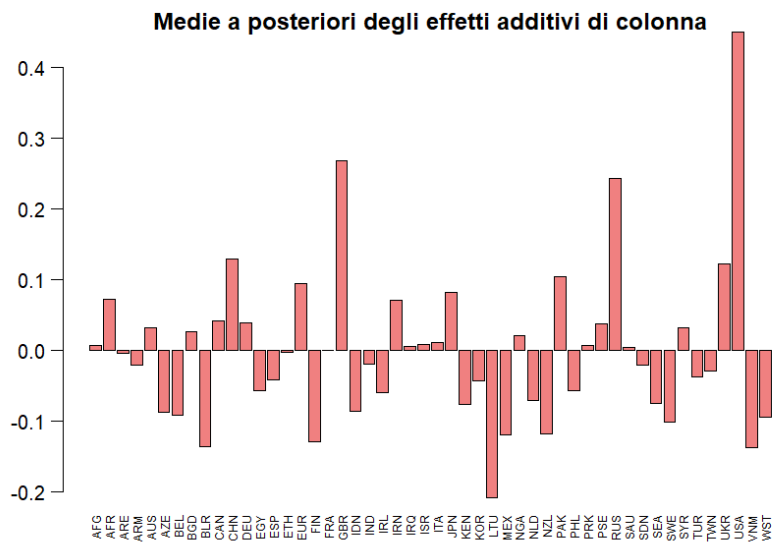


Figura 3.4: Effetti additivi di colonna del modello sui conflitti

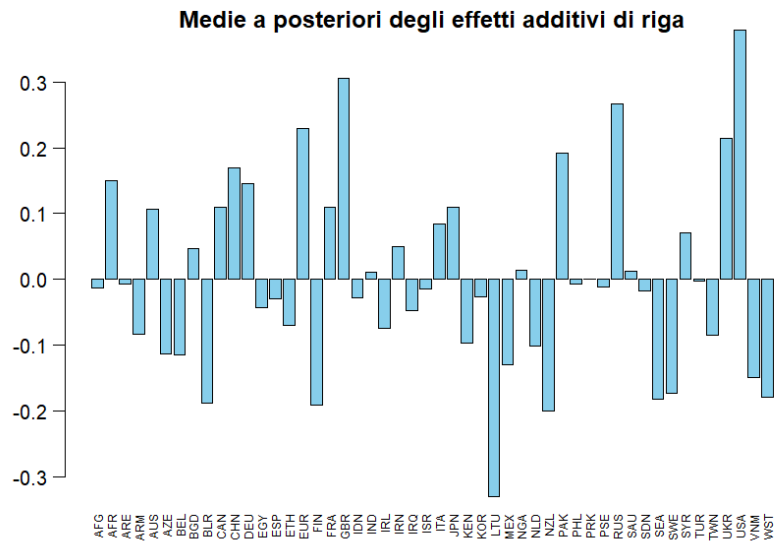


Figura 3.5: Effetti additivi di riga del modello sulle cooperazioni

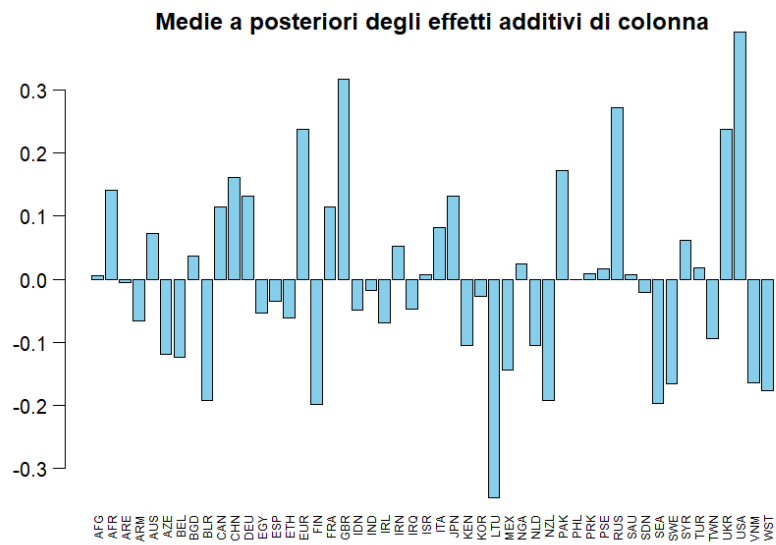


Figura 3.6: Effetti additivi di colonna del modello sulle cooperazioni

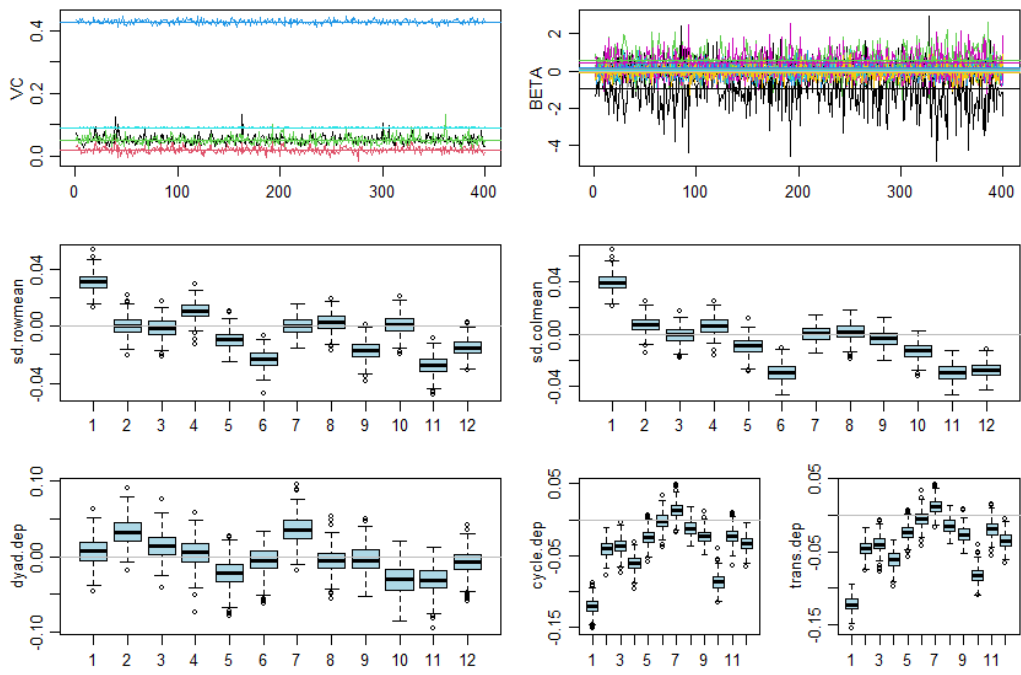


Figura 3.7: Diagnostiche del modello AME sui conflitti

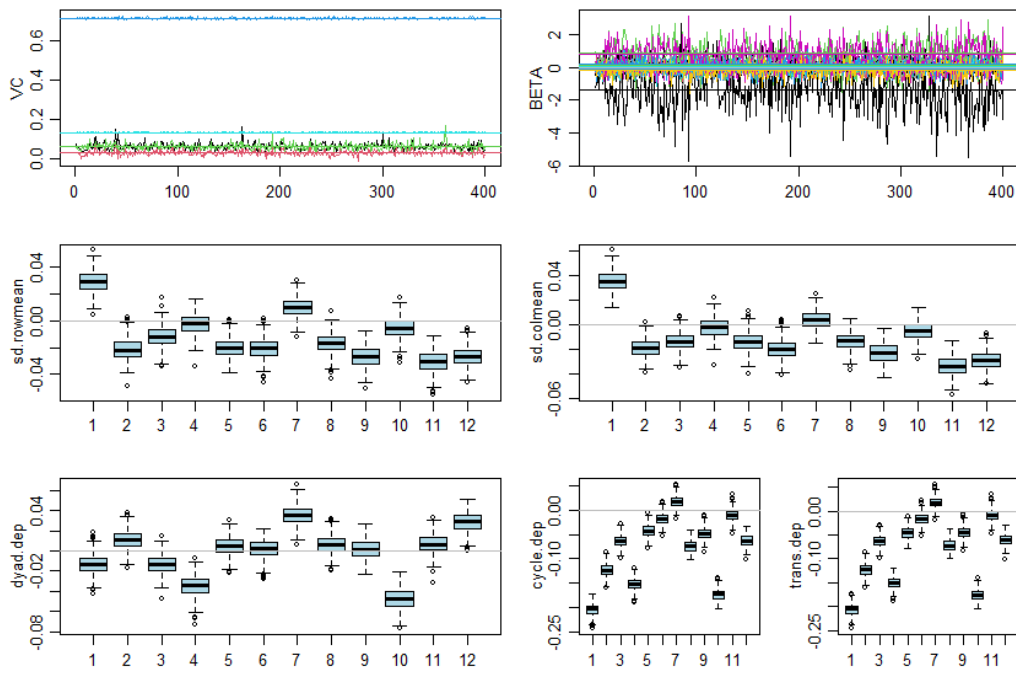


Figura 3.8: Diagnostiche del modello AME sulle cooperazioni

Bibliografia

Caldara, D., & Iacoviello, M. (2022). *Measuring Geopolitical Risk*. American Economic Review.

Enciclopedia Treccani. (2022). *Lebensraum*.

GDELT. (2013). *The GDELT Story, Learn about how GDELT came to be, who made it, and how you can use*.

GDELT. (2013). *The GDELT event database data format codebook v2.0*, 1-7.

Hoff, P. (2021). *Additive and Multiplicative Effects Network Models*. Institute of Mathematical Statistics.

Hoff, P. D., & Ward, M. D. (2003). *Modeling Dependencies in International Relations Networks* (Working Paper no. 35). University of Washington, Seattle.

Kirkley, A. (2024). *Identifying hubs in directed networks*. Institute of Data Science, University of Hong Kong, Hong Kong.

Kolaczyk, E. D. (2009). *Statistical analysis of network data*. Springer.

Krivitsky, P. N., Koehly, L. M., & Marcum, C. S. (2020). *Exponential-family random graph models for multi-layer networks*. *Psychometrika*, 85(3), 630-659.

Leetaru, K., & Schrod, P. A. (2013). *GDELT: Global data on events, location and tone* (Version 1.0).

Pace, L., & Salvan, A. (2001). *Introduzione alla statistica vol.2: Inferenza, verosimiglianza, modelli*. Volume 2, Springer

R Documentation. (2023). *Fit a linear model to the residuals in a gofN object*.

Schrod, P. A. (2012). *CAMEO, conflict and mediation event observations event and actor codebook* (Version 1.1b3).

Statnet. (2023). *Exponential Random Graph Models (ERGMs) using statnet*. The statnet Project.

Statnet. (2023). *Introduction to Exponential-family Random Graph Models with ergm*. The Comprehensive R Archive Network.

UNPD. United Nations Development Programme (2024). *Human Development Index (HDI)*.

Wikipedia. (2024). *Exponential family random graph models*.

Wikipedia. (2024). *List of countries by system of government*.