

CNANT

Complex Network Analysis On Ants Interaction Patterns

Gianluca Dalmaso
Pio Raffaele Fina



1 Introduzione

Lo studio delle formiche - *Mirmecologia* - ha da sempre offerto all'essere umano spunti di riflessione e fonte d'ispirazione nelle scienze e nelle discipline artistiche (Figura 1). Nel corso degli anni, la complessa organizzazione sociale di questi insetti *imenotteri* è stato il principale oggetto di studio dei sociobiologi, contribuendo inoltre alla definizione di metodi formali di ottimizzazione come *ant colony optimization* e concetti dell'Intelligenza Artificiale come la *Swarm Intelligence*.

Le formiche, insieme ad altre specie (es. api, vespe, termiti) sono insetti *eusociali*; animali la cui organizzazione sociale è caratterizzata da: suddivisione del lavoro in caste; funzione riproduttiva riservata ad un'unica o poche femmine; sovrapposizione delle generazioni; cooperazione nella crescita dei piccoli [1]. Questa forma di organizzazione, che raggiunge l'apice nelle colonie di formiche, offre differenti vantaggi rispetto agli insetti solitari. La presenza di molti individui può aumentare la robustezza e affidabilità del "sistema colonia" e le attività possono essere organizzate in modo più efficiente attraverso la suddivisione del lavoro ed il partizionamento dei compiti [2].

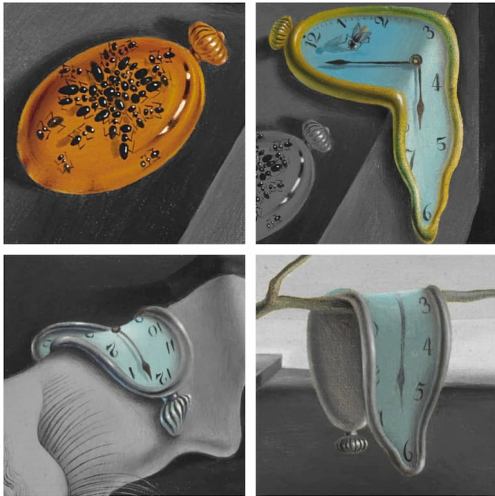
Il repertorio di funzioni cognitive e comportamentali di una singola formica, sono estremamente limitate se comparate a quelle del cervello umano. Considerando però le interazioni complessive della colonia, emergono **complessi comportamenti sociali** e capacità collettive di *decision making* tipiche della società umana.

Uno degli aspetti più interessanti e più studiati è sicuramente quello comunicativo. Le formiche comunicano tra di loro emettendo scie di feromoni recepiti dalle antenne di altre formiche [2], la tipologia del segnale chimico emesso consente di fornire differenti informazioni contestuali: *path planning* per il foraggiamento, allarme per una situazione di pericolo o di attacco da parte di un intruso nella colonia. La comunicazione, dunque, avviene puramente a livello locale ma permette di **sincronizzare** a differenti scale di risoluzione il comportamento di altri individui in base alle necessità della colonia.

Il ruolo ricoperto dalla regina, seppur centrale, è puramente di dominanza piuttosto che di comando e controllo. Le attività della colonia non vengono dunque orchestrate **centralmente** o in modo puramente gerarchico *top-down*, ma piuttosto dalle necessità della colonia stessa (es. foraggiamento di nuove risorse alimentari, difesa e manutenzione del nido, esplorazione, smaltimento rifiuti, navigazione, immunità sociale ai patogeni).

Anche al lettore meno attento, risulterà evidente come le caratteristiche dell'organizzazione sociale di una colonia, coincidono con quelle tipiche dei **sistemi complessi**: self-organization, comportamento emergente, controllo decentralizzato e sincronizzazione. Di conseguenza, la *Network Theory* e la simulazione possono offrire dei framework matematici utili per la caratterizzazione del "sistema complesso colonia" sia dal punto di vista strutturale che dinamico.

In precedenza sono state riportate le principali caratteristiche dell'organizzazione sociale delle formiche, escludendo una serie di aspetti tipici della mirmecologia come l'anatomia, la riproduzione, l'evoluzione e la classificazione tassonomica. Questi ultimi aspetti risultano fondamentali per avere una visione completa di questa specie, ma non aggiungerebbero informazioni rilevanti rispetto agli obiettivi e i temi trattati nel corso. Per una completa trattazione dell'argomento si può fare riferimento a [3].



(a) *Persistenza della Memoria*, 1931, Salvador Dalí.



(b) *La Cicala e la Formica*, Esopo.



(c) *A Bug's Life*, 1998, Pixar Animation Studios.



(d) *Architectural Intersections*, 2017, Gabrielle Duggan.

Figura 1: Alcuni esempi dell'utilizzo delle formiche come soggetto di opere artistiche e letterarie.

2 Dataset

Il *Dataset* utilizzato¹ contiene informazioni sulle **interazioni** tra singoli individui (Figura 2) di 6 differenti colonie di formiche *Camponotus fellah*. Delle 6 colonie presenti nel dataset, abbiamo selezionato in modo totalmente casuale la colonia 2. I dati raccolti nello studio di Mersch *et al.* [4] sono il risultato di un sistema di video tracking automatico basato sull'utilizzo di *fiducial markers*. Questo ha permesso di tracciare il movimento dei singoli individui durante un periodo di osservazione totale di 41 giorni.

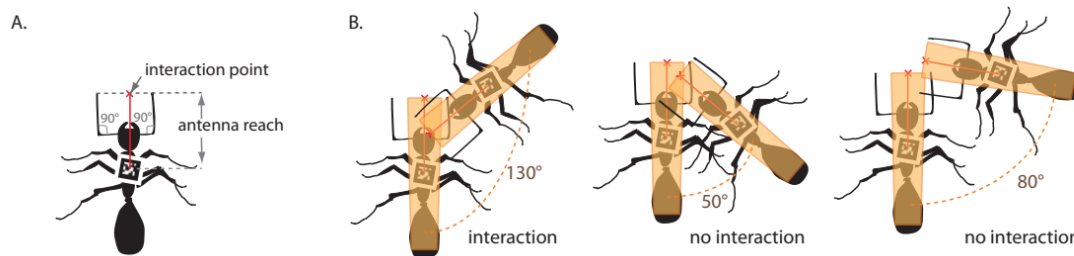


Figura 2: Rilevamento di un'interazione utilizzando un modello trapezoidale. Riprodotto da [4].

Per facilitare il processo di detection e tracking automatico, l'esperimento è stato svolto in un ambiente artificiale controllato formato da due arene collegate tra di loro da un tunnel (Figura 3). La prima arena contiene le risorse per il foraggiamento mentre la seconda permette di ospitare il nido. Il setup completo dell'esperimento è particolarmente articolato e per ulteriori approfondimenti si rimanda al materiale supplementare in [4].

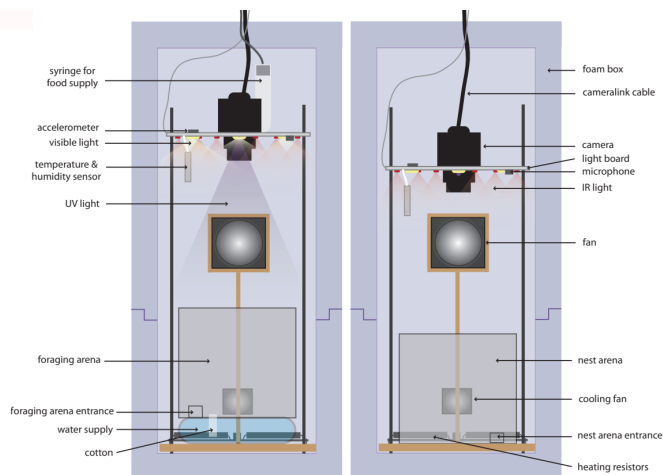


Figura 3: Setup dell'esperimento. Riprodotto da [4].

¹liberamente accessibile al seguente link

Le interazioni all'interno della colonia sono codificate in un **grafo temporale**

$$\mathcal{G} = \{G(t_1), G(t_2), \dots, G(t_{41})\}$$

La rappresentazione adottata è comunemente chiamata in letteratura *snapshots* (Figura 4). Ogni singolo snapshot $G(t)$ è un grafo pesato non diretto, il peso w associato al generico arco (u, v, w) rappresenta il **numero di interazioni** tra due formiche rilevate all'interno della finestra di osservazione di 24h. L'interazione tra due formiche è stabilita in modo puramente quantitativo (Figura 2) e non offre nessun informazione di tipo qualitativo (es. interazione per un evento di *Trofalassi*²).

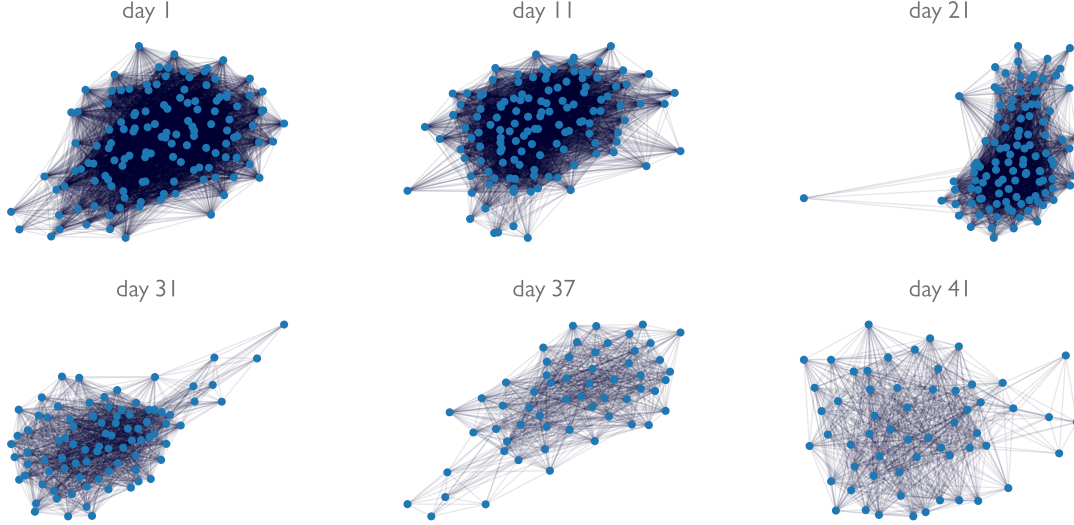


Figura 4: Rappresentazione di alcuni snapshots (force directed layout).

Per la quasi totalità dei nodi della rete vengono forniti i seguenti attributi: ID colonia e ID formica; dimensione ed età; casta ricoperta in 4 periodi temporali; numero di interazioni con regina, foraggiatrici (foragers), "spazzine" (cleaners) e accuditrici (nurses).

Il grafo temporale è suddiviso in 4 periodi, il primo formato da 11 giorni, 10 giorni i restanti tre. In seguito ad una operazione di clustering eseguita dall'algoritmo *INFOMAP* [5], le formiche sono state suddivise in 3 clusters (foragers, cleaners, nurses) in base alla prossimità spaziale rispetto all'entrata del nido, accumulo di rifiuti e posizione delle larve. Per ogni periodo, ad ogni formica viene associato il cluster in cui essa effettuava la suddetta attività in almeno il 75% delle giornate. In Figura 25 (appendice) è mostrata la rete partizionata nelle communities individuate nel primo periodo, per ulteriori dettagli si rimanda al materiale supplementare in [4].

²evento di condivisione di cibo o fluidi tra due insetti

Dalla Figura 5 si può effettuare una prima osservazione: avanzando negli istanti di tempo, il numero di nodi (e di conseguenza anche di archi) negli snapshots si riduce. Questa dinamica, che ritroveremo in seguito, è dovuta alla rimozione di alcuni individui dalla colonia, causata dal loro decesso o dalla perdita del marker associato [4].

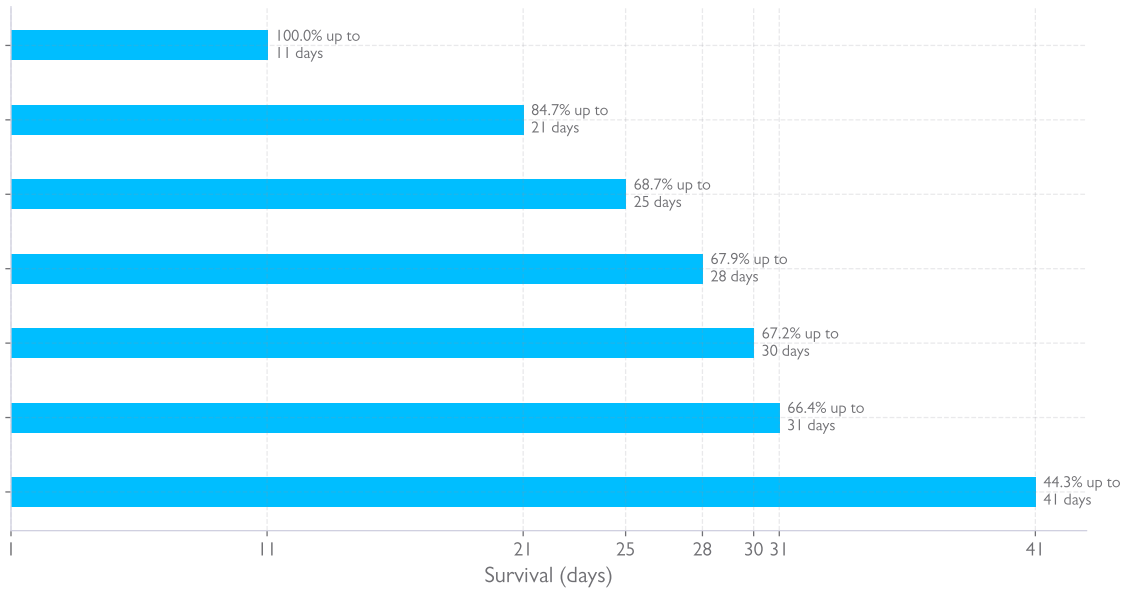


Figura 5: Tempo di sopravvivenza delle formiche nella colonia.

3 Static Network Analysis

Per poter effettuare un'analisi statica sulle caratteristiche strutturali della rete, bisogna affrontare due questioni preliminari:

- Gli snapshots $G(t)$ presentano un'elevata densità sia su scala globale (Figura 4) che locale (Figura 6) comportando problemi in termini computazionali. Bisogna dunque ridurre la connettività della rete attraverso un'operazione di **sparsificazione**.
- Le interazioni degli individui nella colonia sono rappresentate da un grafo temporale $\mathcal{G}(t)$ composto da 41 snapshots. Bisogna dunque **aggregare** l'informazione temporale condensandola in un'unica rete statica.

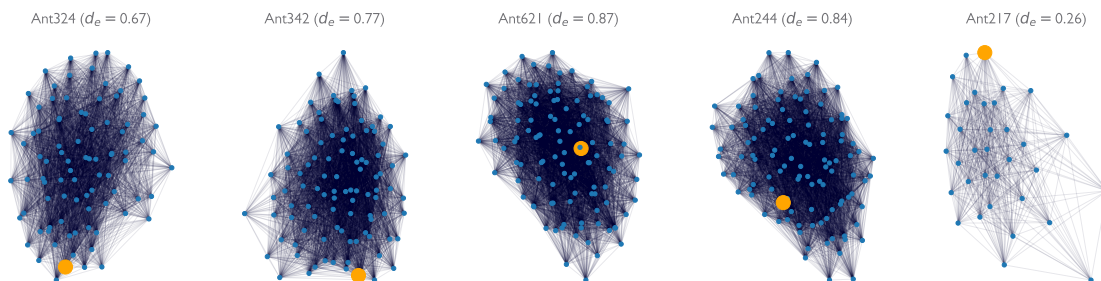


Figura 6: Ego Network di alcuni nodi campionati uniformemente del primo snapshot (giorno 1). I valori d_e rappresentano il rapporto tra i nodi della ego network e i nodi dello snapshot. In colore arancione viene evidenziato il nodo campionato.

Di seguito viene riportata una rappresentazione diagrammatica della metodologia utilizzata per risolvere le due questioni appena descritte (Figura 7).

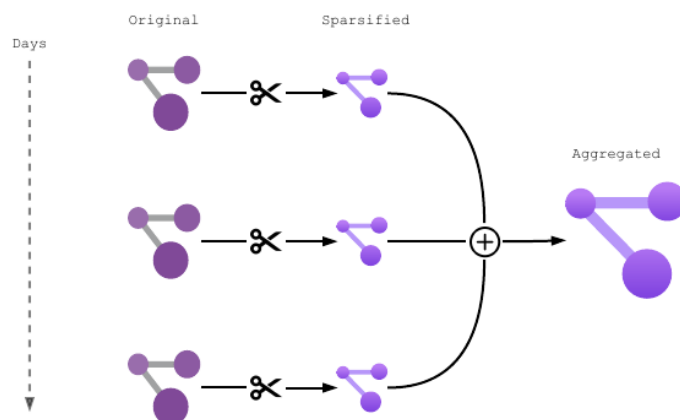


Figura 7: Workflow utilizzato per ottenere la rete aggregata.

Sparsificazione Le motivazioni dietro alla sparsificazione sono molteplici. In primis la possibilità di eseguire algoritmi con elevata complessità computazionale su grafi di elevate dimensioni. In aggiunta, la sparsificazione permette di ridurre la ridondanza e il rumore dell'informazione codificata dalla rete. Quest'ultimo punto è di primaria importanza in quanto il processo di "interaction detection" eseguito in [4] può produrre interazioni false positive. Per aumentare la robustezza sperimentale ci è sembrato dunque naturale, eliminare tutte quelle interazioni (edges) la cui frequenza è minore di una certa soglia, poiché è ragionevole pensare ad esse come a del rumore.

Sebbene in letteratura siano presenti diversi algoritmi di sparsificazione [6], abbiamo scelto di utilizzare il più diretto ed interpretabile: l'*Edges Global Threshold*. Sia $G(N, L)$ un grafo pesato, definiamo $\tilde{G}(\tilde{N}, \tilde{L})$ come il grafo ottenuto da G eliminando i links il cui peso è inferiore ad una soglia globale T definita a priori. Formalmente:

$$\tilde{G}(\tilde{N}, \tilde{L}) \quad \text{con} \quad \tilde{N} = N \quad \text{e} \quad \tilde{L} = \{(u, v, w) \in L \mid w \geq T\}$$

Nel caso specifico, l'algoritmo è stato eseguito su ognuno dei 41 snapshots con $T = 10$ (Figura 8).

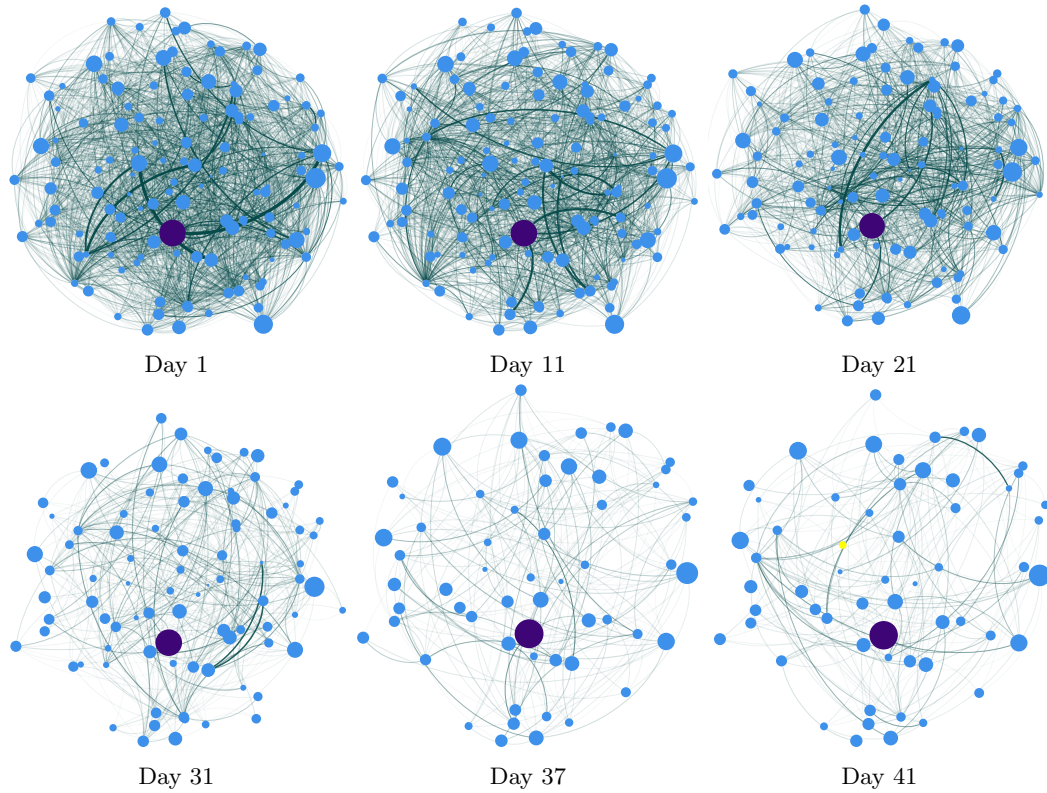


Figura 8: Rappresentazione di alcuni snapshots dopo il processo di sparsificazione. Dimensione dei nodi proporzionale alla dimensione fisica della formica. Trasparenza e spessore degli archi proporzionale al numero di interazioni. Layout utilizzato: *Prefuse Force Directed* con edge betweenness come peso (Cytoscape).

In Figura 9 possiamo osservare l'effetto della sparsificazione sulla densità delle reti. In Figura 9 (A) è possibile osservare un trend temporale decrescente. Questo fenomeno è imputabile principalmente alla rimozione di alcuni individui dalla colonia per decesso o per la perdita dei loro markers [4]. L'operazione di sparsificazione può comportare l'alterazione delle proprietà strutturali e statistiche della rete, per verificarlo è stata effettuata una comparazione della *degree distribution* della rete prima e dopo la sparsificazione utilizzando il test d'ipotesi *Kolmogorov-Smirnoff* a due samples [7]. Il test è stato eseguito sul primo snapshot in quanto è la rete più densamente connessa rispetto ai restanti istanti temporali, dunque quella che permette di valutare al meglio l'impatto della sparsificazione.

Ad un livello di significatività $\alpha = 0.05$ il test permette di rifiutare l'ipotesi nulla (due campioni provengono dalla stessa distribuzione) in favore dell'ipotesi alternativa (due campioni provengono da due differenti distribuzioni). La stessa conclusione si ottiene osservando il *QQ plot* in Figura 9 (B), nella quale si nota chiaramente una deviazione rispetto alla retta bisettrice.

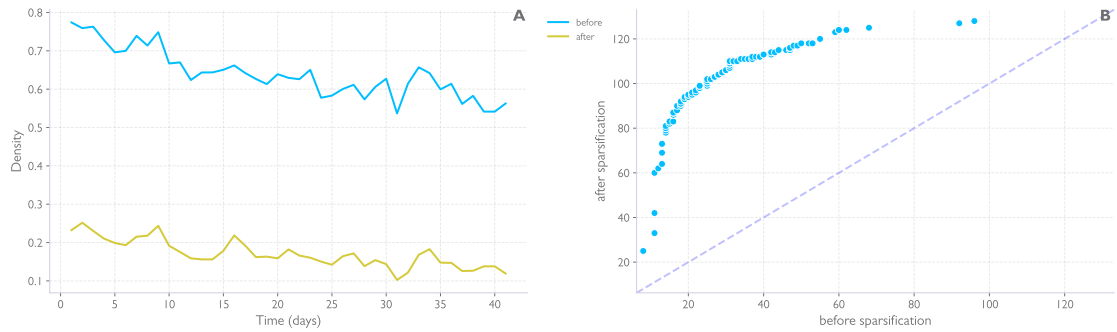


Figura 9: Confronto tra i valori di densità degli snapshots prima e dopo la sparsificazione. Evoluzione temporale (A). QQ plot della degree distribution prima e dopo la sparsificazione (B).

Aggregazione Temporale Dato che i links in uno snapshot rappresentano la frequenza di interazione tra gli individui in un dato giorno di osservazione, è naturale ottenere una rete che descrive le interazioni complessive semplicemente sommando le interazioni avvenute nelle singole giornate.

Sia A il tensore³ di dimensione $N \times N \times t_{max}$ che descrive la rete temporale \mathcal{G} , otteniamo la rete aggregata $\tilde{\mathcal{A}}$ di dimensione $N \times N$ come segue:

$$\tilde{\mathcal{A}}_{ij} = \sum_{t=1}^{t_{max}} A_{ijt}$$

Procedendo in questo modo, il peso w di un generico link (u, v, w) di $\tilde{\mathcal{A}}$ rappresenta le interazioni cumulative tra due formiche nella finestra temporale di 41 giorni. Aggregando gli snapshots, le informazioni associate alla dimensione temporale sono inevitabilmente perse. $\tilde{\mathcal{A}}_{ij}$ risulta dunque un'approssimazione che rende tuttavia l'analisi maggiormente trattabile. Nel seguito ci riferiremo alla rete sparsificata e aggregata come: "rete empirica", "rete aggregata", "rete della colonia" o semplicemente "colonia".

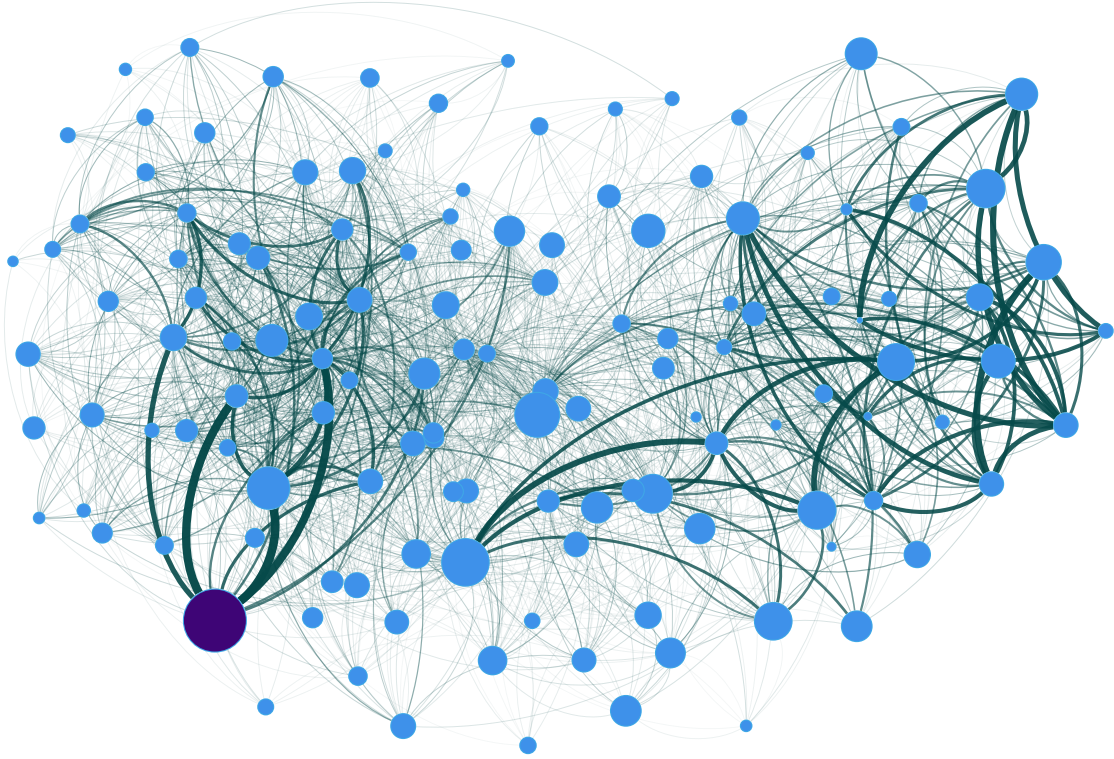


Figura 10: Rete aggregata. Dimensione dei nodi proporzionale alla dimensione fisica della formica. Trasparenza e spessore degli archi proporzionale al numero di interazioni. Layout utilizzato: *Prefuse Force Directed* con edge betweeness come peso (Cytoscape).

³ rappresentazione alternativa ma equivalente a quella a snapshots.

3.1 Degree Distribution

Una delle proprietà che caratterizzano al meglio la struttura complessiva di una rete è sicuramente la *degree distribution*. Dalle considerazioni fatte in precedenza sulla densità della rete, ci aspettiamo che la degree distribution della rete non sia particolarmente eterogenea.

La degree distribution ottenuta in Figura 11, non suggerisce una distribuzione heavy-tail tipica delle reti *scale free*. Per corroborare l'affermazione, procediamo al fitting con alcune distribuzioni teoriche. Nello specifico è stata utilizzata la famiglia delle distribuzioni *power law*. La scelta è guidata dal fatto che la degree distribution di una rete *scale-free* segue una distribuzione power law [8]. Da differenti risultati empirici, i valori tipici del parametro γ sono $2 \leq \gamma \leq 3$ [8].

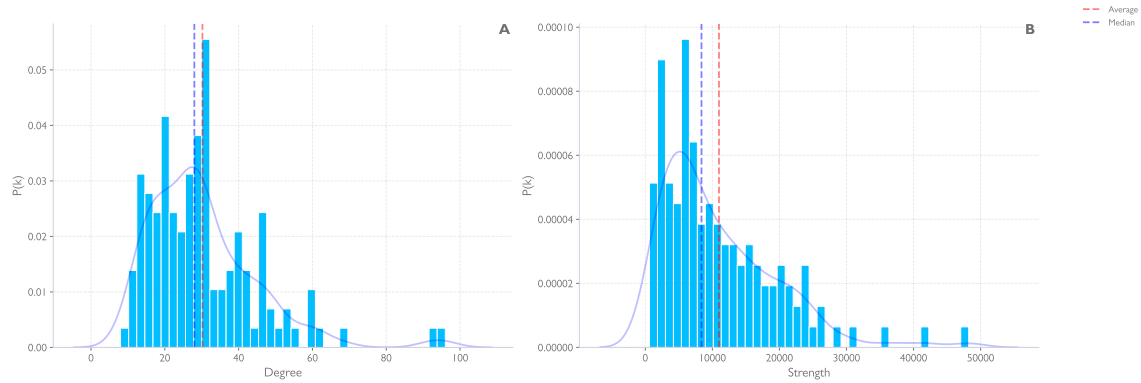


Figura 11: Degree distribution (A), strength distribution (B).

Per valutare la significatività statistica procediamo con il test *one sample Kolmogorov-Smirnoff* tra la distribuzione empirica e una distribuzione power law con $\gamma \in [2.0, 3.0]$ con incrementi $\Delta = \frac{1}{10}$. Nel 100% dei casi, con significatività $\alpha = 0.05$, possiamo rifiutare l'ipotesi nulla in favore dell'alternativa, confermando dunque la mancanza della principale caratteristica delle scale free networks. Nell'analisi temporale di Blonder *et al.* effettuata su differenti colonie di formiche è riportata la medesima conclusione [9].

Da un punto di vista qualitativo, la mancanza della caratteristica distribuzione heavy tail si traduce nell'assenza di nodi hubs. Questa osservazione non sorprende particolarmente e possibili spiegazioni dal punto di vista biologico potrebbero essere:

- La comunicazione nella colonia avviene principalmente a livello locale e manca di una struttura gerarchica top-down. Le interazioni risultano dunque particolarmente eterogenee, non concentrandosi in pochi "hub di comando".
- La dinamica che potrebbe incidere maggiormente sull'assenza di un meccanismo di *preferential attachment* sottostante all'organizzazione della colonia è il *polietismo temporale* o in parole più semplici, la suddivisione del lavoro per età [3]. Questo fenomeno, tipico degli insetti eusociali, permette con l'avanzamento dell'età di cambiare casta, progredendo tipicamente da un ruolo di accudimento (interno al nido) a uno di foraggiamento (esterno al nido). Di conseguenza, il cambiamento di casta di un individuo permetterà l'interazione con nuovi individui di differenti caste, rendendo anche in questo caso le interazioni maggiormente eterogenee.

La validità di tali affermazioni deve essere investigata ulteriormente in quanto l'analisi è stata effettuata su una singola colonia di una specifica specie di formiche.

In modo analogo alla degree distribution è possibile calcolare la *strength distribution* riportata in Figura 11 (B). Il valore di strength di un nodo i è definito ⁴ come:

$$s_i = \sum_{i \neq j} w_{ij}$$

In Figura 11 possiamo notare la differenza tra degree e strength distribution. Qualitativamente, la prima, permette di valutare l'eterogeneità quantificando le interazioni **uniche** tra gli individui della colonia. La seconda, permette di valutare la frequenza delle interazioni e quindi **quanto** gli individui della colonia interagiscono tra di loro. Un esempio è riportato in appendice in Figura 25. Quest'ultimo aspetto verrà utilizzato diverse volte nelle successive analisi, in particolare nella sezione 4 quando verrà analizzata la diffusione di contagi con modelli epidemiologici.

3.2 It's a Ants Ants Ants World

Uno dei pattern strutturali frequentemente riscontrabile nelle reti reali è sicuramente lo "small-worldness": reti con un elevato coefficiente di clustering ed un basso *average path length* (APL) [10]. Dal punto di vista funzionale una tale struttura promuove l'efficienza del trasferimento d'informazione tra gli individui della rete.

In Figura 12 è possibile analizzare qualitativamente la distribuzione delle distanze (A) e il coefficiente di clustering (B). Con i valori ottenuti: APL $L_{emp} = 1.84$ e clustering coefficient medio $C_{emp} = 0.45$ potremmo concludere, per definizione, che la rete presenta le caratteristiche tipiche delle reti small-world. Per poterlo confermare, dobbiamo prima verificare che i valori di L_{emp} e C_{emp} non siano una diretta conseguenza delle ridotte dimensioni ed elevata densità della rete [11].

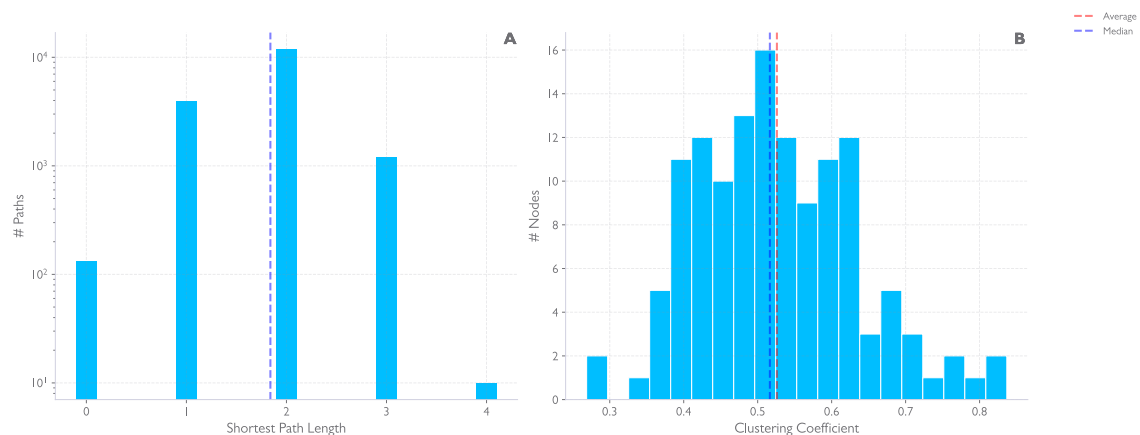


Figura 12: Distribuzione lunghezza shortest path tra tutte le coppie di nodi (A) e Distribuzione del clustering coefficient (B).

⁴utilizzando come rappresentazione del grafo la matrice di adiacenza

La metodologia più diffusa per investigare lo small world prevede di normalizzare e comparare L_{emp} e C_{emp} rispetto ad L_{rand} e C_{rand} calcolate su una qualche famiglia di **null model** [12]. In letteratura sono presenti diverse metriche, tra cui la più nota è il *coefficiente sigma*:

$$\sigma = \frac{\gamma}{\lambda} \quad \text{con} \quad \gamma = \frac{C_{emp}}{C_{rand}} \quad \lambda = \frac{L_{emp}}{L_{rand}}$$

se $\sigma \rightarrow 1$ allora si può concludere che la rete presenta le caratteristiche di una rete small world [12]. Nel corso degli anni sono state messe in luce differenti limitazioni per σ , la principale è il fatto di utilizzare come null model il modello Erdős-Rényi. Quest'ultimo presenta un basso coefficient di clustering che può indurre a errate interpretazioni del risultato [12]. Per ovviare a tale problema, differenti alternative sono state proposte in letteratura [12, 13]. Di queste, abbiamo optato per l'utilizzo del *coefficiente omega* proposto da Telesford *et al.* [12]:

$$\omega = \frac{L_{rand}}{L_{emp}} - \frac{C_{emp}}{C_{latt}}$$

Per generare il null model e calcolare L_{rand} , è stato utilizzato l'algoritmo di *edges-swapping* di Maslov-Sneppen. Tale algoritmo permette di generare reti random che preservano la densità e la degree distribution della rete in input. Per C_{latt} è stata utilizzata una variante dell'algoritmo di Maslov-Sneppen per generare una versione "latticezzata" della rete in input. Dopo 500 simulazioni il coefficiente è pari a $\omega = -0.36$ con $L_{rand} = 1.77 \pm 0.001$ e $C_{latt} = 0.34 \pm 0.003$, le distribuzioni sono riportate in Figura 13. Il valore ottenuto suggerisce una **tendenza** ad una rete small world piuttosto che ad un lattice regolare [12].

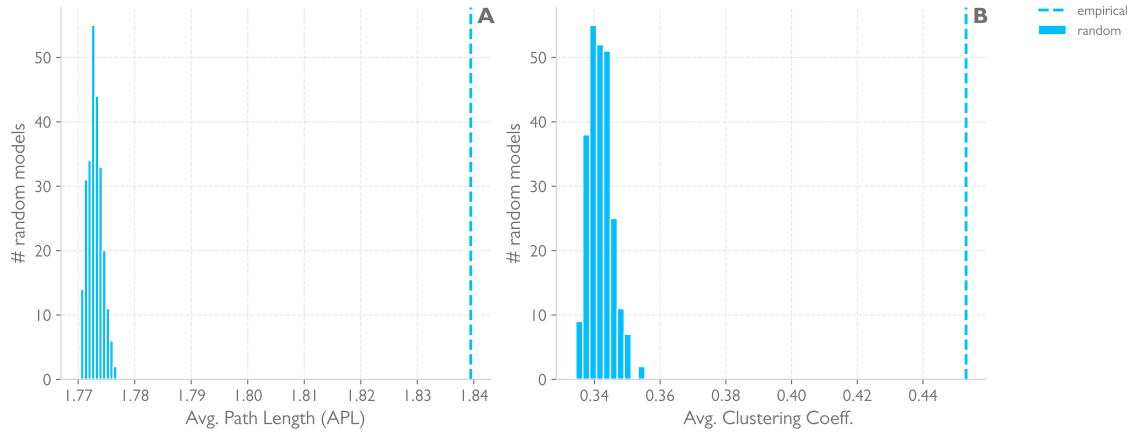


Figura 13: Distribuzione APL (A) e Avg. Clustering Coefficient (B) ottenute da 500 realizzazioni dei rispettivi random models.

3.3 Network Robustness

Sia per i sistemi naturali che per quelli artificiali, la robustezza è una proprietà di fondamentale importanza per la continuità del funzionamento di un sistema sotto molteplici perturbazioni. La robustezza di una rete è infatti spesso definita dalla capacità della rete di mantenere il "normale" funzionamento del sistema in seguito ad un attacco e dal tempo di recupero impiegato per ripristinare le funzionalità.

L'analisi dell'assortatività della rete può fornire differenti informazioni sulla robustezza della stessa, in quanto, spesso, sono quantità correlate fra di loro [14, 15, 16].

- **Reti disassortative:** sono presenti differenti nodi hubs con un numero elevato di componenti connesse molto dense. Data la dipendenza dagli hubs per il flow dell'informazione, risultano particolarmente vulnerabili ad attacchi mirati-
- **Reti assortative:** al contrario, presentano un numero minore di componenti connesse che sono però molto più resistenti ad attacchi mirati.
- **Reti neutre:** risultano un tradeoff tra le due precedenti tipologie, presentando i vantaggi offerti da entrambe.

Il tipo di assortatività della rete, può essere studiato analizzando il pattern di *degree correlation* fra i nodi della rete. Per poter quantificare il valore di correlazione è stato utilizzato il *coefficiente di Pearson*. Un coefficiente positivo ne indica l'assortatività, un valore negativo ne indica la disassortatività, un valore prossimo allo zero ne indica invece la neutralità.

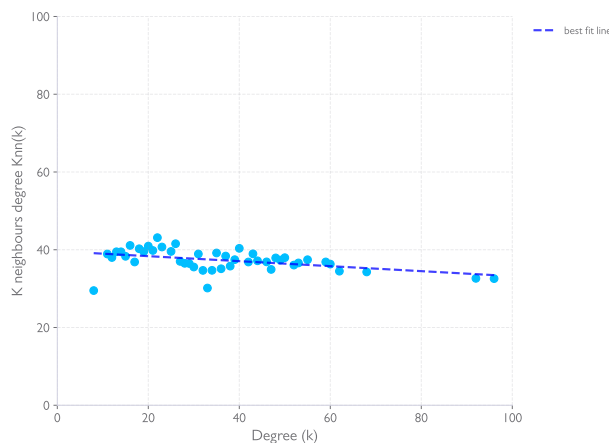


Figura 14: Analisi assortatività della rete

Il coefficiente di pearson calcolato sulla rete risulta $r = -0.086$. Pur essendo un valore negativo, esso è molto vicino allo zero, suggerendo una struttura assortativa **neutrale** della rete, la quale dovrebbe comportare un numero basso di componenti molto densi e teoricamente una buona resistenza ad attacchi mirati e non.

Una caratterizzazione alternativa della robustezza del sistema è possibile investigando alcune proprietà strutturali come la **connettività**. Dalle considerazioni precedenti ci aspettiamo poche componenti connesse, al limite, una singola componente. Per valutare la robustezza della rete sotto diverse condizioni di perturbazione, selezioniamo una percentuale crescente di nodi da rimuovere n_r , quantificando l'effetto sulla rete attraverso:

- n. componenti connesse.
- n. nodi nella *Largest Connected Component* (LCC).
- probabilità di un nodo di appartenere alla LCC.

Con numero di nodi rimossi $n_r = 0$ (rete originale) otteniamo 1 singola componente connessa contenente tutti i 131 nodi, confermando l'ipotesi precedente.

Per valutare la robustezza in differenti scenari, sono state utilizzate tre strategie di rimozione:

- **Random Removal** (RR): ogni nodo ha probabilità uniforme $p_r = \frac{1}{131}$ di essere rimosso.
- **Removal by centrality** (RC): viene definito un'ordinamento dei nodi in base allo score di centralità ed eliminati per primi i nodi più centrali. Lo score può prendere in considerazione anche i pesi associati agli archi (RCW).
- **Removal by attribute** (RA): viene definito un'ordinamento sui valori dell'attributo "age" e rimossi di conseguenza i nodi.

Random Removal In questo caso vogliamo valutare la robustezza della colonia nel caso in cui il decesso degli individui sia dovuto puramente al caso e non ad un particolare fenomeno. Dato che la strategia di rimozione ha una componente stocastica, per ogni percentuale di rimozione $n_r \in [0, 1]$ con incremento $\Delta = \frac{1}{100}$, sono stati eseguiti 1000 campionamenti uniformi sull'insieme dei nodi \tilde{N} . I valori riportati in Figura 15 rappresentano la media aritmetica su differenti campionamenti dei nodi della rete empirica comparati con 1000 realizzazioni dei modelli random teorici.

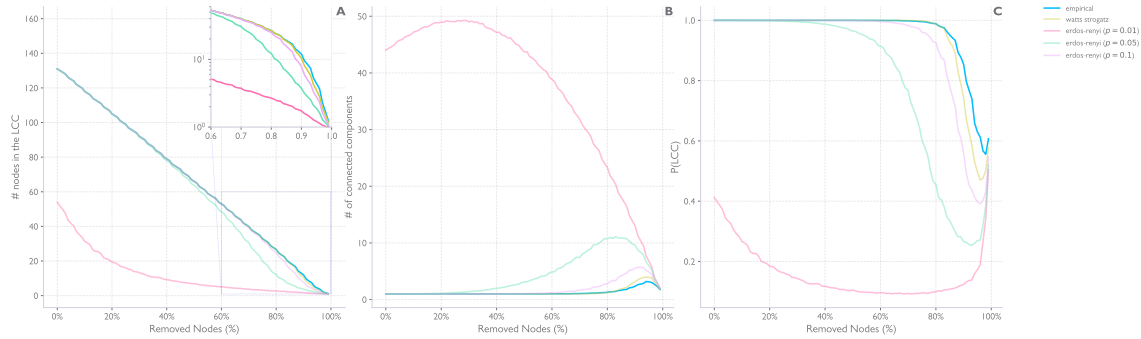


Figura 15: Robustezza della rete al variare della percentuale di rimozione n_r . La robustezza è quantificata da differenti metriche surrogate: n. di nodi nella LCC (A), n. componenti connesse (B), probabilità $P(u \in LCC)$ (C). I modelli teorici utilizzati sono $G_{n,p}$ con $n = 131$ e $p = 0.01$, $p = 0.5$, $p = 0.1$; *Watts-Strogatz* con $n = 131$, $K = \langle k \rangle_{emp}$ e $p = 0.5$.

Dalla Figura 15 emergono 2 osservazioni interessanti. Prima fra tutte, la rete empirica è particolarmente robusta in questo tipo di scenario, in accordo con l'osservazione fatta in precedenza sull'assortatività. Questo lo si può osservare soprattutto nel plot (A) attraverso la relazione lineare inversa tra le due variabili e il plot (B) in cui è richiesta la rimozione di più dell' 80% dei nodi per disconnettere la rete.

Una seconda dinamica si apprezza comparando la curva della rete empirica con quella dei restanti modelli teorici. Con il modello Watts-Strogatz c'è una sovrapposizione quasi totale, lo stesso per Erdős-Rényi con $p = 0.1$, invece il comportamento della rete si differenzia particolarmente per $p \rightarrow 0$.

Removal by centrality Con l'utilizzo di questa strategia vogliamo emulare l'azione di un agente patogeno esterno abbastanza evoluto da individuare ed eliminare gli individui più importanti della colonia. Cerchiamo dunque di riprodurre un **attacco mirato**.

L'importanza degli individui è stata quantificata dalla metrica *degree centrality*. Un confronto con altre metriche note in letteratura è riportato in Figura 16. I risultati sono riportati in Figura 17.

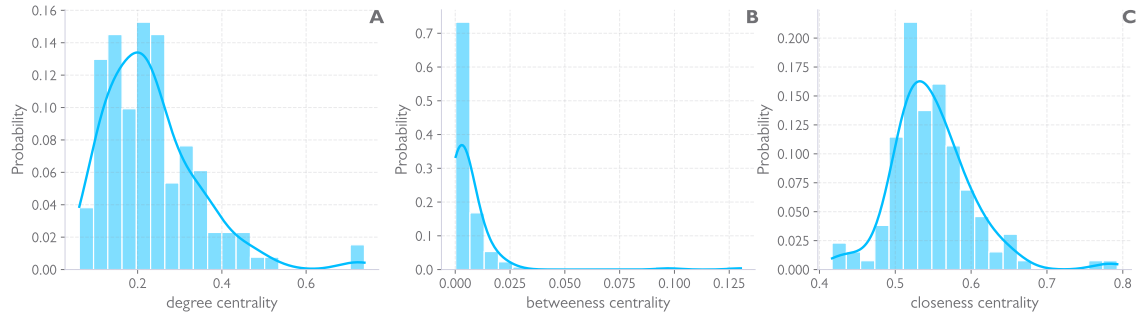


Figura 16: Distribuzioni di diverse misure di centralità della rete: degree centrality (A), betweenness centrality (B), closeness centrality (C).

Removal by attribute Simile alla strategia removal by centrality, in questo caso si vuole analizzare l'influenza di fattori fisiologici degli individui sulla robustezza della rete. Ipotizzando quindi, che gli individui più giovani ed anziani siano più vulnerabili ad un eventuale patogeno, abbiamo investigato la strategia RA rimuovendo per primi i più anziani (RA oldest) e successivamente effettuato l'esatto opposto (RA youngest). I risultati sono riportati in Figura 17.

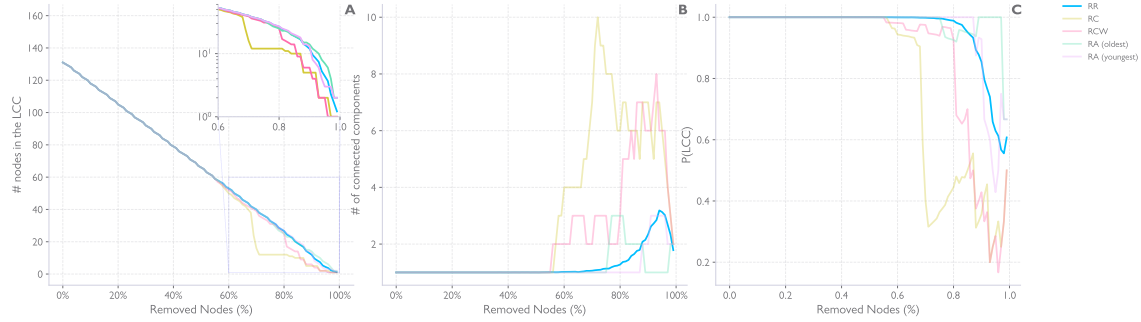


Figura 17: Robustezza della rete utilizzando differenti strategie di rimozione. La robustezza è quantificata da differenti metriche surrogate: n. di nodi nella LCC (A), n. componenti connesse (B), probabilità $P(u \in LCC)$ (C).

Dalla Figura 17 (A) possiamo osservare che la robustezza della rete è pressoché identica fino alla soglia di rimozione $n_r < 0.6$. Superata questa soglia si notano le differenze tra le strategie adottate. Quelle con l'impatto (negativo) peggiore sono sicuramente RC e RCW, ad evidenziare come una buona e fondata strategia di attacco abbia più effetto. L'aspetto interessante è la differenza tra RC e RCW, in questo caso ha più effetto scegliere di attaccare un individuo che viene in contatto con diversi individui, piuttosto che uno che interagisce più frequentemente con alcuni di loro.

L'aspetto che forse lascia più sorpresi è la quasi totale ininfluenza (rispetto la baseline RR) degli attacchi basati su fattori fisiologici (RA oldest e RA younger), ad indicare come l'età di un individuo non influisca particolarmente sul risultato dell'attacco.

4 Temporal Network Analysis

La maggiorparte dei sistemi complessi legati alla realtà che ci circonda si sviluppa su due dimensioni principali, quella temporale e quella spaziale. L'analisi statica di un sistema, seppur informativa sul funzionamento dello stesso, rimane comunque un'approssimazione.

Una delle dinamiche più interessanti degli insetti eusociali (in particolare le formiche) è il **polietismo temporale**, ovvero la capacità degli individui di cambiare casta in base all'età. In questa sezione verrà dunque investigato se questa dinamica è rilevabile o meno attraverso un'analisi strutturale degli snapshots della rete.

Dal punto di vista puramente strutturale una casta può essere ricondotta al concetto di *community*: insieme di individui che interagiscono più frequentemente rispetto agli altri esterni alla community. Per l'obiettivo prefissato, il partizionamento della rete in communities è di relativa importanza, l'aspetto realmente interessante è la transizione nel tempo degli individui tra le communities.

Come primo step, su ognuno dei 41 snapshots viene eseguito l'algoritmo di community detection *Louvain* proposto da Blondel *et al.* basato sull'ottimizzazione della modularità [17]. Il risultato è la *communities membership table* (un estratto in Tabella 1) la quale codifica l'appartenenza di ogni nodo (formica) al rispettivo cluster nei 41 giorni.

La significatività delle communities rilevate dipende principalmente dalla funzione utilizzata per valutare la bontà del partizionamento [18], dunque il partizionamento prodotto potrebbe risultare (parzialmente) non significativo. Per aumentare la robustezza dell'analisi viene effettuato un raggruppamento temporale in 4 periodi (11 giorni il primo, 10 i restanti). In ognuno di questi, ad ogni formica viene assegnata la community più frequente come quella rappresentativa dell'intero periodo. Come ultimo step, viene calcolato "il flow" tra le community come il numero di formiche in comune tra le communities di due periodi successivi. La rappresentazione finale è riportata in Figura 18.

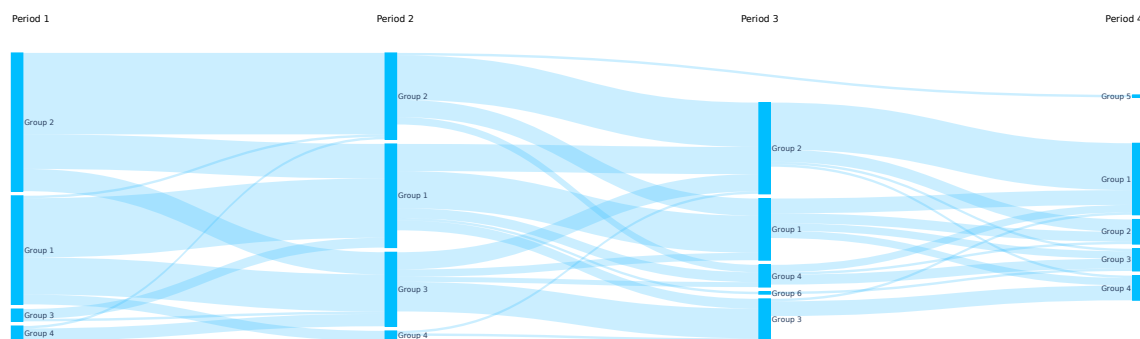


Figura 18: Diagramma Sankey rappresentate il flusso delle formiche tra le communities individuate nei quattro periodi temporali.

Dalla Figura 18 si possono osservare differenti dinamiche. Sicuramente si può notare un trasferimento degli individui tra le communities, che diventa più consistente con l'avanzare del tempo. Quest'ultima dinamica può essere imputabile al fatto che la popolazione della colonia decresce nel tempo e dunque oltre al polietismo temporale interviene anche il cambiamento di casta dovuto ad

	day1	day2	day3	day4	day5	day6	day7	day8	day9	day10
Ant621	0	1	0	1	0	1	0	0	0	0
Ant665	1	2	2	1	1	1	1	1	2	2
Ant356	1	2	2	1	1	1	1	2	2	2
Ant540	2	1	1	0	0	2	0	1	0	3
Ant227	0	1	1	0	3	0	3	2	2	1
Ant114	1	2	2	1	1	1	1	2	2	2
Ant112	0	1	0	0	2	0	3	2	2	0
Ant195	0	0	1	0	0	2	0	0	0	0
Ant191	0	1	0	0	2	2	3	2	2	0
Ant217	2	0	1	0	0	2	2	0	0	3
Ant492	1	2	2	1	1	1	3	1	2	2
Ant324	1	2	2	1	2	0	3	2	1	0
Ant329	2	1	1	0	0	2	2	0	0	1
Ant650	1	2	2	1	1	1	1	2	2	2
Ant556	1	2	2	1	1	1	1	1	2	2
Ant552	1	2	2	1	1	1	1	2	2	2
Ant252	1	1	0	1	3	1	1	2	2	0
Ant256	2	0	1	0	0	2	2	0	0	1
Ant257	1	2	0	1	1	1	3	2	2	2
Ant126	0	0	1	0	0	2	2	0	0	1
Ant26	1	2	2	1	1	1	1	1	2	2
Ant318	1	2	2	1	1	1	1	2	1	2
Ant20	2	0	1	0	0	2	2	0	0	1
Ant137	1	2	2	1	1	1	3	1	2	2
Ant332	1	2	2	1	1	1	1	1	2	2
Ant647	1	2	2	1	3	1	1	2	2	0
Ant560	1	1	2	1	1	1	1	2	2	2
Ant268	2	0	1	0	0	2	2	0	0	3
Ant564	2	0	1	0	0	2	2	0	0	1
Ant243	2	0	1	0	0	2	2	0	0	1
Ant244	1	1	2	1	1	1	3	2	1	0

Tabella 1: Communities membership table dei primi 10 giorni su un campione della colonia.

una maggiore frequenza di ri-organizzazione in base alle necessità della colonia [3]. In Figura 19 viene rappresentato uno scenario ideale in assenza di polietismo temporale in cui le transizioni tra le 3 caste sono dovute principalmente al caso.

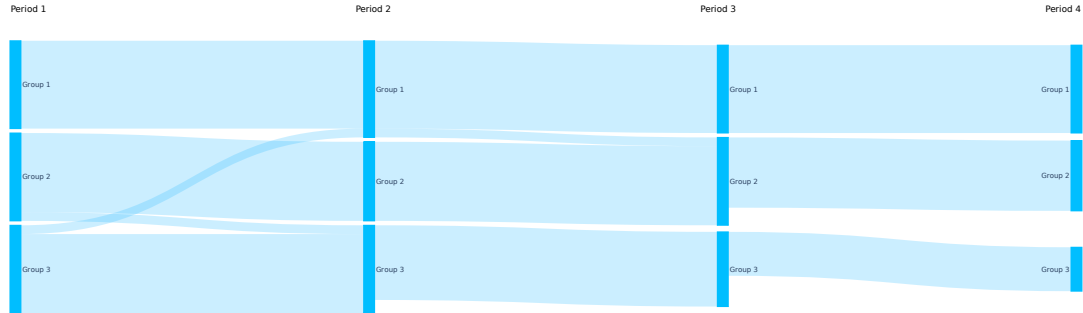


Figura 19: Diagramma Sankey rappresentate il flusso delle formiche tra le communities in uno scenario ideale in assenza di polietismo temporale.

Un aspetto importante da evidenziare è il fatto che non è possibile assegnare una qualche etichetta semantica (foragers, nurses, cleaners) alle communities trovate. Questa limitazione è dovuta principalmente alla mancanza di dati spaziali sulla posizione delle formiche all'interno della colonia, i dati rilasciati dagli autori sono disponibili solo in forma aggregata e dunque non utilizzabili per tale scopo [4].

5 Network Dynamics

In questa seconda parte viene simulato un **processo di diffusione** epidemico all'interno della colonia. Il processo è stato simulato utilizzando i tre modelli epidemiologici compartimentali più diffusi: SIR, SIS, SIRS. I tre modelli non catturano una meccanica di difesa utilizzata dalle colonie per ridurre il contagio durante i periodi di epidemia. Questa meccanica è basata sulla quarantena temporanea dei membri infetti [19], i quali, per non contribuire ad un rapido spread della malattia, cambiano il loro comportamento riducendo le interazioni con il resto della colonia [20].

Per poter modellare questa dinamica è stata implementata una variante del modello SIR chiamata SIR-Q. Ad ogni time step della simulazione viene eseguita una rimozione, disconnettendo con probabilità $p_r = 0.8$, i vicini del nodo infetto. L'utilizzo del comparto Q con i modelli SIS e SIRS necessiterebbe di un'operazione di *attachment* tra i nodi in seguito alla loro re-immissione nel comparto S. Non avendo però conoscenze a priori sul meccanismo naturale sottostante alle interazioni tra le formiche, abbiamo preferito utilizzare come baseline il modello SIR.

Lo studio sulle dinamiche è stato suddiviso in tre parti:

1. Preparazione e normalizzazione dei dati (di alcuni attributi dei nodi)
2. Analisi dei modelli SIR, SIS, SIRS e SIR-Q sulla rete aggregata
3. Analisi dei modelli SIR e SIR-Q sulla rete temporale

5.1 Preparazione Dati

Per eseguire questi modelli dinamici iniziamo definendo i parametri dell'epidemia:

- Probabilità di contagio $\beta = 0.1$
- Tempo di infezione = 3 (o "tempo di recupero")
- Tempo di immunità = 4

I valori dei parametri sono stati definiti dopo differenti simulazioni preliminari: data l'elevata connettività della rete, alzare la probabilità di contagio oltre $\beta > 0.1$ non cambierebbe la dinamica dello spread. Il valore del tempo di infezione è rapportato alla scala temporale della vita media delle formiche, che ricordiamo essere molto minore rispetto a quella umana. Il valore del tempo di immunità, verrà affrontato meglio nella sezione successiva.

Per poter modellare in modo più verosimile la realtà abbiamo eliminato la *fully mixed hypothesis* tipica della formulazione base del SIR. I parametri del modello non vengono assegnati a livello di popolazione ma ai singoli nodi della rete, permettendo dunque di calibrarli in modo eterogeneo sulla base delle seguenti caratteristiche:

- Dimensione del corpo
- Età dell'individuo
- Attività svolta nel periodo corrente

Per semplificare il modello, gli attributi della popolazione sono stati suddivisi in due 'fasce':

- Fascia centrale, composta dai nodi aventi valore dell'attributo più simile al valor medio
- Fascia esterna, composta dai nodi aventi valore dell'attributo distante dal valor medio

I valori di **dimensione corpo** e di **età** sono stati normalizzati in un intervallo $[-1, 1]$. Il range degli intervalli è stato diviso in $2r$ fasce gaussiane attraverso cui possiamo quantizzare i dati normalizzati, negli esperimenti seguenti $r = 2$.

La lunghezza di un intervallo i viene così definita:

$$length_i = 2x \frac{\int_{(i-r-1)/r}^{(i-r)/r} \mathcal{N}(0, 0.6, x) dx}{\int_{-1}^1 \mathcal{N}(0, 0.6, x) dx}$$

In questo modo possiamo partizionare la popolazione, differenziando tra "popolazione media" e "popolazione sensibile", in quanto un elemento troppo giovane o troppo anziano ha una maggiore probabilità di contagiarsi, maggiore tempo di infezione o un minor tempo di immunità rispetto ad un elemento nel fior degli anni (es. difesa immunitaria ridotta).

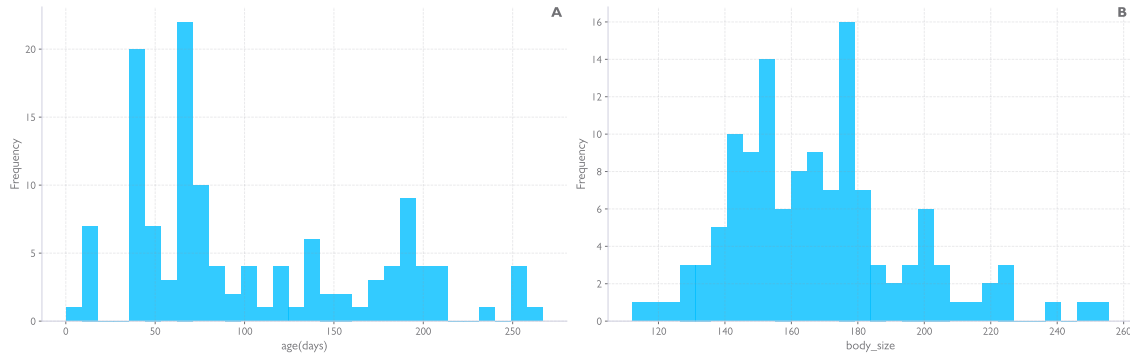


Figura 20: Distribuzioni dell'età (A) e della dimensione del corpo (B) nella colonia

La tipologia di **attività svolta** dalla formica viene utilizzata per decidere da quali individui della colonia far partire l'epidemia. Nel nostro caso le formiche che ricoprono il ruolo di foragers vengono utilizzate come seeds dell'epidemia in quanto, essendo l'interfaccia del nido con il mondo esterno, sono le responsabili dell'introduzione del patogeno.

5.2 Modelli dinamici sulla rete aggregata

In questa sezione vengono riportati i risultati delle simulazioni sulla rete aggregata. Il focus sarà principalmente sulle differenze tra SIR e SIR-Q.

- **SIR**: La sua esecuzione è la più rapida (rispetto a SIR e SIRS), esegue un minor numero di confronti nella rete, concludendo la sua esecuzione in 5-10 istanti temporali.
- **SIS**: Le simulazioni non terminano entro i primi 50 time steps, di conseguenza l'epidemia non si stabilizza ⁵, continuando a creare nuovi infetti. In Figura 21 si può osservare l'andamento dell'epidemia per i primi 30 time steps.
- **SIRS**: Utilizzando un tempo di immunità $TI = 2$, l'andamento dell'epidemia è simile a quello ottenuto con il SIS. Incrementando però questa soglia ad un valore ≥ 3 , l'andamento cambia totalmente diventando molto più simile alle simulazioni ottenute con il modello SIR, concludendo (generalmente) in meno di 10 time steps.
- **SIR-Q**: I risultati di SIR-Q variano totalmente se consideriamo o meno il peso degli archi durante la sua esecuzione. Se consideriamo il peso degli archi (Figura 22), l'andamento dell'epidemia è molto simile al SIR, come ragionevole aspettarsi c'è una riduzione del numero di infetti, ma il tempo per stabilizzarsi aumenta. In modo opposto, se consideriamo tutti gli archi come con peso 1 (Figura 23), l'epidemia si stabilizza in tempi simili a quelli ottenuti con il SIR, ma, come ci aspettavamo, abbiamo una notevole riduzione del numero di infetti che scende anche del 90% in molti casi.

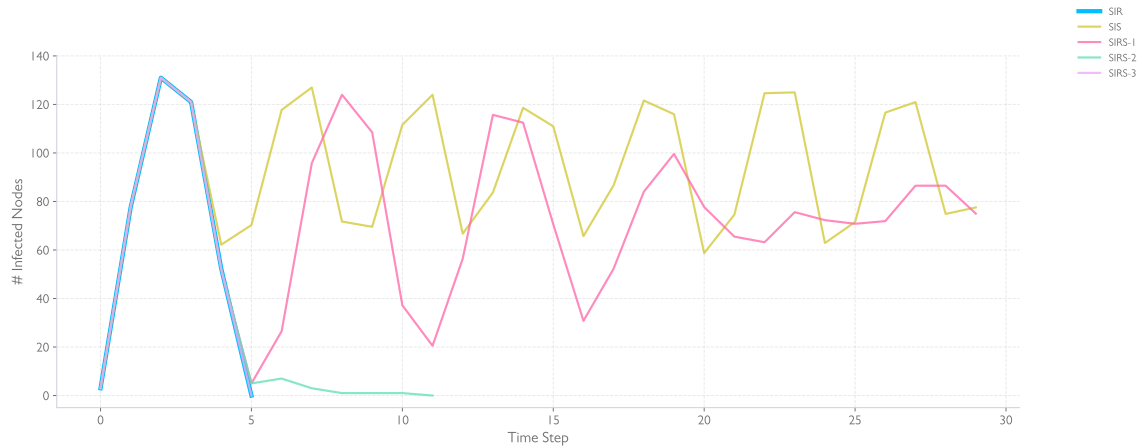


Figura 21: Distribuzioni del numero medio di infetti al variare del tempo di immunità (TI), con: $TI=2$ (SIRS-1), $TI=3$ (SIRS-2), $TI=4$ (SIRS-3). Interessante notare l'andamento ondulatorio di SIS e SIRS-1. Risultati ottenuti dall'esecuzione di 1000 simulazioni.

⁵numero di infetti scende a 0

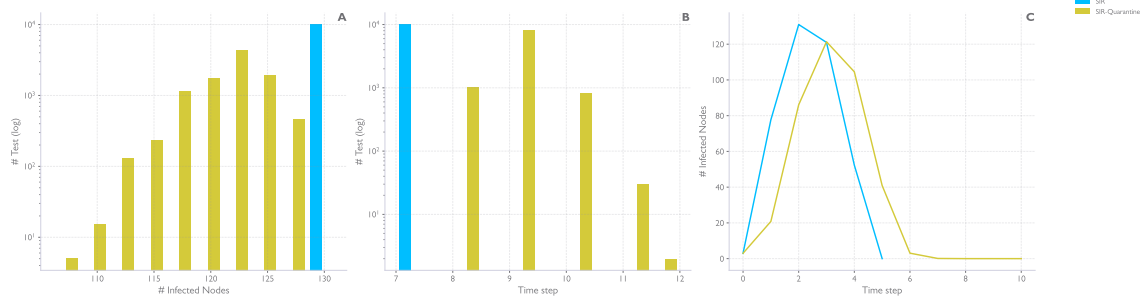


Figura 22: Distribuzioni di: # massimo infetti (A), # istanti di tempo per la stabilizzazione (B), # medio di infetti negli istanti di tempo (C). Risultati ottenuti dall'esecuzione di 10k test effettuati sulla rete aggregata considerando i pesi sugli archi

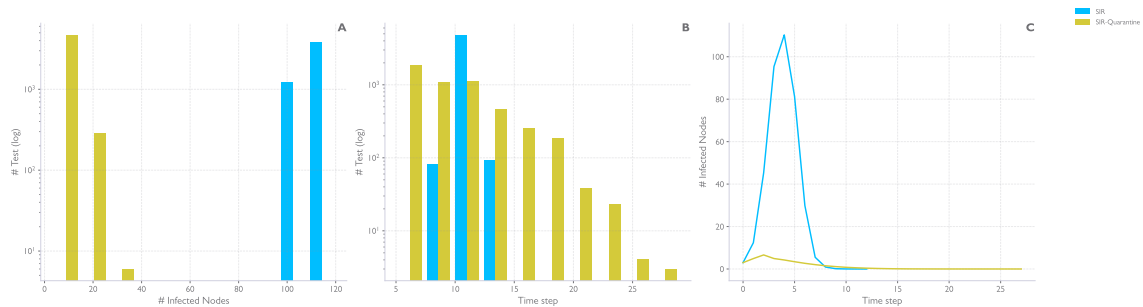


Figura 23: Distribuzioni di: # massimo infetti (A), # istanti di tempo per la stabilizzazione (B), # medio di infetti negli istanti di tempo (C). Risultati ottenuti dall'esecuzione di 10k test effettuati sulla rete aggregata senza considerare i pesi sugli archi

Le principali caratteristiche riscontrate dall'esecuzione di questi algoritmi sulla rete in uso sono:

1. Pur avendo una probabilità molto bassa di infezione, l'alta densità della rete permette comunque l'esplosione dell'epidemia.
2. La modifica del tempo di immunità, seppur di un singolo istante temporale, può cambiare totalmente lo spread dell'epidemia, rendendo l'efficacia degli anticorpi/durata del vaccino degli strumenti indispensabili per combattere l'epidemia in reti così dense.
3. L'utilizzo della quarantena è una tecnica altamente efficace per bloccare l'epidemia o, in generale, ridistribuire nel tempo il carico sul sistema.

Per una visualizzazione dei modelli eseguiti, fare riferimento alle gif nella cartella GIF/Aggregated.

5.3 Modelli dinamici sulla rete temporale

Poichè abbiamo a disposizione una rete temporale, abbiamo modificato gli algoritmi SIR e SIR-Q in modo tale che, ad ogni time step della simulazione, cambiasse lo snapshot $G(t_i)$ utilizzato (con $t_i \in [t_0, \dots, t_{41}]$). Questo fa sì che il numero di nodi e di archi possa cambiare da un step all'altro, permettendoci di modellare un'epidemia su una rete non statica. Il tempo di esecuzione degli algoritmi sulla rete temporale aumenta in confronto al loro utilizzo sulla rete aggregata (passiamo da 25s a 39s circa per ogni 10 test). Questo è causato dai controlli aggiuntivi necessari per la cancellazione degli archi fra istanti temporali (controllo di presenza del nodo al prossimo istante, ecc). Considerando i pesi sugli archi, i risultati ottenuti con i due modelli SIR e SIR-Q sono pressoché simili. Al contrario, eliminando il contributo del peso, l'effetto della quarantena incide notevolmente sull'andamento dell'epidemia, dunque, per la diffusione dell'epidemia l'**eterogeneità** delle interazioni ⁶ risulta di maggiore impatto rispetto alle **quantità** delle interazioni. (Figura 24).

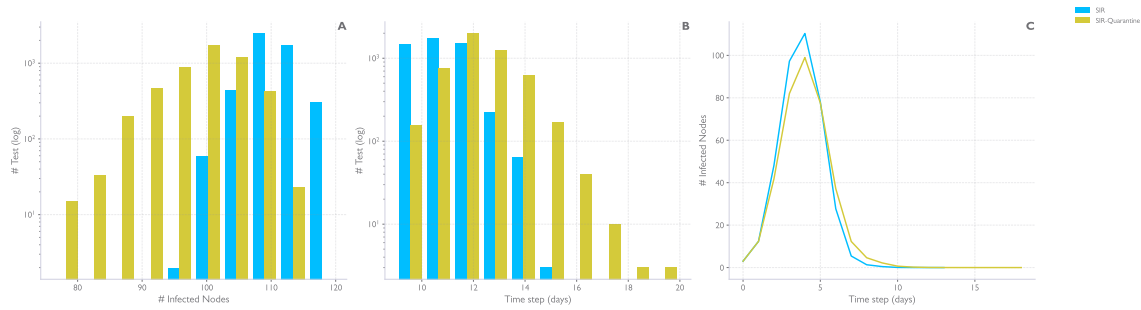


Figura 24: Distribuzioni di: # massimo infetti (A), # istanti di tempo per la stabilizzazione (B), # medio di infetti negli istanti di tempo (C). Risultati ottenuti dall'esecuzione di 10k test effettuati sulla rete temporale senza considerare i pesi sugli archi

Per una visualizzazione dei modelli eseguiti, fare riferimento alle gif nella cartella GIF/Temporal.

Conclusioni come è possibile notare dai risultati ottenuti, con una rete estremamente connessa, il meccanismo di quarantena funziona, permettendo di ridurre il numero medio di individui infetti durante l'epidemia, pur non riducendone per forza la durata.

Più in generale, i risultati ottenuti, potrebbero essere estesi al processo di *Information Spreading*. In una rete con queste caratteristiche strutturali, ipotizziamo che la diffusione di un'informazione possa avvenire in modo esponenziale: una volta che una notizia viene fatta passare non ci sarebbe modo per arrestare la sua diffusione (non avendo una struttura ad hubs, dovremmo rimuovere molti nodi per evitarne lo spread). I pattern d'interazione finora analizzati rispecchiano quello che è il meccanismo alla base della comunicazione nella colonia, ad esempio durante le invasioni, l'allarme deve essere propagato in fretta per poter raggruppare velocemente le formiche soldato.

⁶eterogeneità s'intende come numero di interazioni con differenti individui.

6 Appendice

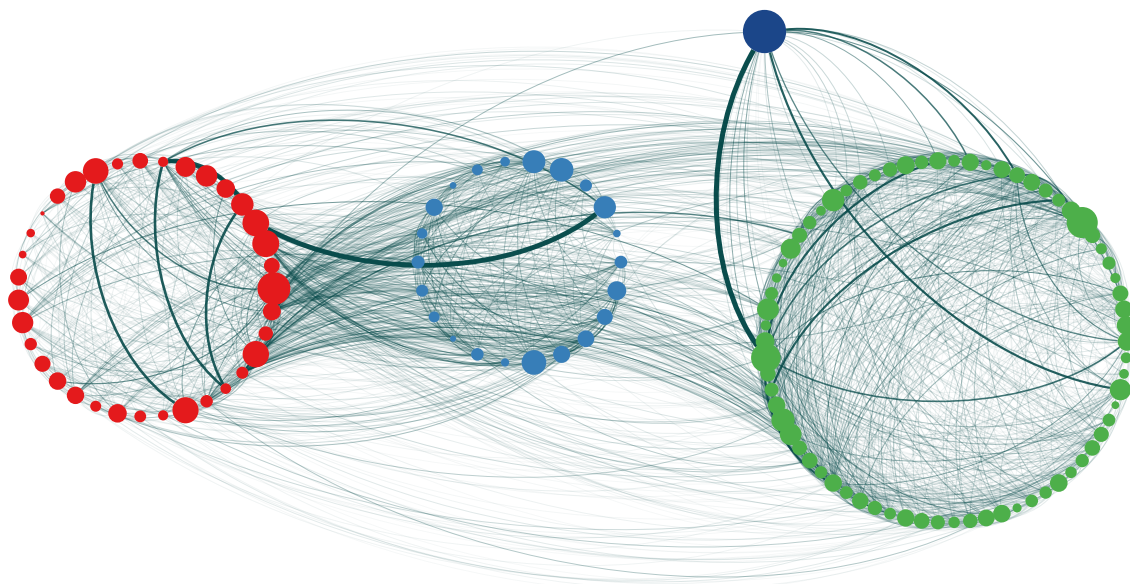


Figura 25: Communities individuate nel primo periodo nello studio di Mersch et al. [4]. Nurses (verde), foragers (blu), cleaners (rosso), regina (viola). Dimensione dei nodi proporzionale alla dimensione fisica della formica. Trasparenza e spessore degli archi proporzionale al numero di interazioni. Layout utilizzato *Group Attribute* (Cytoscape).

Riferimenti bibliografici

- [1] Edward O. Wilson. *The insect societies*. Cambridge, Mass: Belknap Press of Harvard University Press, 1971. ISBN: 9780674454903.
- [2] Duncan E. Jackson e Francis L.W. Ratnieks. “Communication in ants”. en. In: *Current Biology* 16.15 (ago. 2006), R570–R574. ISSN: 09609822. DOI: 10.1016/j.cub.2006.07.015. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0960982206018343> (visitato il 09/09/2021).
- [3] Bert Hölldobler e Edward O. Wilson. *The ants*. eng. Berlin: Springer, 1990. ISBN: 9783540520924.
- [4] Danielle P. Mersch, Alessandro Crespi e Laurent Keller. “Tracking Individuals Shows Spatial Fidelity Is a Key Regulator of Ant Social Organization”. EN. In: *Science* (mag. 2013). URL: <https://www.science.org/doi/abs/10.1126/science.1234316> (visitato il 08/09/2021).
- [5] Martin Rosvall e Carl T. Bergstrom. “Maps of random walks on complex networks reveal community structure”. In: *Proceedings of the National Academy of Sciences* 105.4 (2008), pp. 1118–1123. ISSN: 0027-8424. DOI: 10.1073/pnas.0706851105. eprint: <https://www.pnas.org/content/105/4/1118.full.pdf>. URL: <https://www.pnas.org/content/105/4/1118>.
- [6] Gerd Lindner et al. “Structure-Preserving Sparsification of Social Networks”. en. In: *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. Paris France: ACM, ago. 2015, pp. 448–454. ISBN: 9781450338547. DOI: 10.1145/2808797.2809313. URL: <https://dl.acm.org/doi/10.1145/2808797.2809313> (visitato il 11/09/2021).
- [7] *Kolmogorov-Smirnov Goodness-of-Fit Test*.
- [8] Albert-László Barabási e Márton Pósfai. *Network science*. Cambridge: Cambridge University Press, 2016. ISBN: 9781107076266 1107076269. URL: <http://barabasi.com/networksciencebook/>.
- [9] Benjamin Blonder e Anna Dornhaus. “Time-Ordered Networks Reveal Limitations to Information Flow in Ant Colonies”. en. In: *PLoS ONE* 6.5 (mag. 2011). A cura di Robert Planque, e20298. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0020298. URL: <https://dx.plos.org/10.1371/journal.pone.0020298> (visitato il 12/09/2021).
- [10] Duncan J. Watts e Steven H. Strogatz. “Collective dynamics of ‘small-world’ networks”. en. In: *Nature* 393.6684 (giu. 1998), pp. 440–442. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/30918. URL: <http://www.nature.com/articles/30918> (visitato il 13/09/2021).
- [11] “Chapter 10 - Null Models”. en. In: *Fundamentals of Brain Network Analysis*. A cura di Alex Fornito, Andrew Zalesky e Edward T. Bullmore. San Diego: Academic Press, gen. 2016, pp. 355–381. ISBN: 9780124079083. DOI: 10.1016/B978-0-12-407908-3.00010-8. URL: <https://www.sciencedirect.com/science/article/pii/B9780124079083000108> (visitato il 13/09/2021).
- [12] Qawi K. Telesford et al. “The Ubiquity of Small-World Networks”. en. In: *Brain Connectivity* 1.5 (dic. 2011), pp. 367–375. ISSN: 2158-0014, 2158-0022. DOI: 10.1089/brain.2011.0038. URL: <http://www.liebertpub.com/doi/10.1089/brain.2011.0038> (visitato il 13/09/2021).

- [13] Sarah Feldt Muldoon, Eric W. Bridgeford e Danielle S. Bassett. “Small-World Propensity and Weighted Brain Networks”. en. In: *Scientific Reports* 6.1 (feb. 2016), p. 22057. ISSN: 2045-2322. DOI: 10.1038/srep22057. URL: <http://www.nature.com/articles/srep22057> (visitato il 13/09/2021).
- [14] Masaya Murakami et al. “Robustness and efficiency in interconnected networks with changes in network assortativity”. In: *Applied network science* 2.1 (2017), p. 6.
- [15] Antonio Scala e Gregorio D’Agostino. “The Robustness of Assortativity”. In: *International Workshop on Critical Information Infrastructures Security*. Springer. 2011, pp. 223–226.
- [16] Peter Csermely et al. “Structure and dynamics of core/periphery networks”. In: *Journal of Complex Networks* 1.2 (ott. 2013), pp. 93–123. DOI: 10.1093/comnet/cnt016. eprint: <https://academic.oup.com/comnet/article-pdf/1/2/93/7093753/cnt016.pdf>. URL: <https://doi.org/10.1093/comnet/cnt016>.
- [17] Vincent D Blondel et al. “Fast unfolding of communities in large networks”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10 (ott. 2008), P10008. DOI: 10.1088/1742-5468/2008/10/p10008. URL: <https://doi.org/10.1088/1742-5468/2008/10/p10008>.
- [18] Santo Fortunato. “Community detection in graphs”. In: *Physics Reports* 486.3 (2010), pp. 75–174. ISSN: 0370-1573. DOI: <https://doi.org/10.1016/j.physrep.2009.11.002>. URL: <https://www.sciencedirect.com/science/article/pii/S0370157309002841>.
- [19] Lise Diez et al. “Emergency measures: Adaptive response to pathogen intrusion in the ant nest”. In: *Behavioural Processes* 116 (2015), pp. 80–86. ISSN: 0376-6357. DOI: <https://doi.org/10.1016/j.beproc.2015.04.016>. URL: <https://www.sciencedirect.com/science/article/pii/S0376635715001230>.
- [20] Zhu Zhongming et al. “For ants, unity is strength—and health”. In: (2018).